UNIVERSITÉ DE MONTRÉAL

ÉVALUATION DE L'INCIDENCE DU FINANCEMENT PUBLIC DE LA RECHERCHE UNIVERSITAIRE QUÉBÉCOISE DANS LE DOMAINE DE LA BIOTECHNOLOGIE

MAXIME CLERK-LAMALICE

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)

DÉCEMBRE 2009

© Maxime Clerk-Lamalice, 2009.

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

ÉVALUATION DE L'INCIDENCE DU FINANCEMENT PUBLIC DE LA RECHERCHE UNIVERSITAIRE QUÉBÉCOISE DANS LE DOMAINE DE LA BIOTECHNOLOGIE

présenté par : CLERK-LAMALICE Maxime

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, Doct., présidente

Mme BEAUDRY Catherine, D. Phil., membre et directrice de recherche

Mme SCHIFFAUEROVA Andrea, Ph. D., membre du jury

REMERCIEMENTS

Je voudrais remercier le travail d'équipe de plusieurs personnes qui ont permis la rédaction de ce mémoire, sans elles, la réalisation de ce travail n'aurait pas été possible. Premièrement, ***Catherine Beaudry***, directrice de recherche, pour son enseignement, sa motivation et son extraordinaire disponibilité. ***Carl St-Pierre*** pour son expertise et son enseignement en statistiques. ***Martin Trépanier*** pour son expertise avec les bases de données et la validation d'algorithmes. ***Andrea Schiffauerova*** pour m'avoir permis de reprendre et continuer une partie de son travail effectué avec les bases de données des publications en biotechnologie. ***Ricard-Olivier Moreau*** pour son assistance de recherche au niveau du nettoyage des bases de données. ***Ahmad Barirani*** pour la création du logiciel PatentBot et son assistance à modifier le code source. ***Vincent Larivière*** de l'OST pour l'extraction des données provenant de SIRU et Octroi. ***Carole Guérin*** du BRCDT pour ses explications sur SIRU. ***Mario Pelletier*** d'Extenway Solutions Inc. pour sa patience et pour m'avoir permis de mener une double vie, soit d'alterner entre l'étudiant et le professionnel. Une mention spéciale pour l'équipe de support d'Extenways Solutions Inc. pour la récupération de mes données de recherche.

L'aide financière pour l'ensemble du projet a été fournie par le Conseil de recherches en sciences humaines du Canada (**CRSH**). Les fonds fournis par le **CIRST** et le **Département MAGI** ont permis d'assister à la 7e édition de la conférence Triple Helix à Glasgow.

RÉSUMÉ

Cette recherche traitera de l'impact du financement public et privé de la recherche universitaire québécoise dans le domaine de la biotechnologie. Plusieurs articles analysent déjà en détail les types de collaboration entre universités, firmes et laboratoires gouvernementaux. Cependant, peu d'entre eux s'intéressent aux deux problématiques suivantes, soit l'impact du financement initial des projets et l'évolution de la connaissance générée par la recherche et les collaborations entre chercheurs. En utilisant les données tirées de la base de données d'articles scientifiques Scopus d'Elsevier et du *Système d'information sur la recherche universitaire* (SIRU) détaillant le financement reçu en termes de contrats et subventions, nous construisons une base de données du financement reçu par les chercheurs universitaires et de leurs publications. Le critère de sélection pour l'extraction d'un article est d'avoir au moins un co-auteur canadien. En utilisant les liens entre co-auteurs, nous construisons un réseau basé sur les chercheurs au cours de la période 1985-2005 à l'aide d'un logiciel développé pour automatiser l'intégration complexe de nos différentes sources de données de façon à générer une matrice d'analyse. À la suite de la construction du réseau, nous proposons un modèle temporel de l'impact du financement de la recherche universitaire et des caractéristiques du réseau sur la production scientifique mesurée par la publication d'articles scientifiques. Les résultats suggèrent que le financement individuel a un impact positif sur la production scientifique et que la catégorisation du financement est nécessaire pour comprendre l'impact réel sur la publication d'articles scientifiques. Le nombre croissant d'articles publiés aujourd'hui peut être expliqué de façon plus précise par les contrats reçus depuis les trois dernières années et par les subventions reçues depuis les cinq dernières années. L'environnement universitaire semble avoir un impact spécifique sur le nombre de publications des chercheurs affiliés.

ABSTRACT

This study focuses on the impact of public and private financing of academic research in the Quebec biotechnology sector. Prior research analyzes the economic impact of university-industry-government collaborations but fewer studies are concerned with the initial funding and the resulting evolution of knowledge creation and dispersion by researchers. Using the information contained in Elsevier's Scopus scientific articles database and the Quebec *Système d'information sur la recherche universitaire* (SIRU) which gives a detailed account of grants and contracts obtained by each researcher, we build a database of the funding received by individual academic researchers and their scientific publications. The selection criterion for an article to be extracted is that at least one Canadian scientist is a co-author. Using co-authorship links, we create the network of individual scientists for the period 1985-2005 with the help of custom built software whose purpose is to automate the complex integration of our multiple data sources and generate an analysis matrix. Following the network construction, we propose a time-related model of the impact of academic R&D financing and network structure on research output measured by the number of papers. Results suggest that individual funding has a positive effect on research output and that funding categorization is crucial in understanding the impacts on scientific publications. The increasing number of papers published today can be better explained by the contracts received in the past 3 years horizon and by the grants received in the past 5 years horizon. The university environment seems to have a specific effect on the level of research output for the biotechnology sector.

TABLE DES MATIÈRES

REMERCIEMENTS iii

RÉSUMÉ iv

ABSTRACT v

TABLE DES MATIÈRES vi

LISTE DES TABLEAUX ix

LISTE DES FIGURES xi

LISTE DES ÉQUATIONS xiv

LISTE DES SIGLES ET ABRÉVIATIONS xv

LISTE DES ANNEXES xvi

INTRODUCTION 1

CHAPITRE 1 REVUE DE LITTÉRATURE 5

1.1 La recherche en biotechnologie 5

1.2 Le financement public de la recherche universitaire 9

1.3 Les réseaux sociaux 14

1.3.1 Les mesures et caractéristiques réseaux 17

CHAPITRE 2 HYPOTHÈSES ET MÉTHODOLOGIE 22

2.1 Les hypothèses 22

2.2 Le modèle général 23

2.3 Les données 26

2.4 Méthodologie de recherche 28

2.4.1 Capture, nettoyage et formatage des données 28

2.4.2 Matrice d'analyse 35

2.4.3 Création des réseaux sociaux 39

2.4.4 Génération automatique de la matrice d'analyse 41

2.4.5 Construction des fichiers d'analyse Stata 45

CHAPITRE 3 CARACTÉRISTIQUES ET STATISTIQUES DES DONNÉES 46

3.1 Financement public 46

3.2 Articles scientifiques et chercheurs 48

3.2.1 Articles scientifiques 49

3.2.2 Chercheurs 59

3.3 Système d'information de la recherche universitaire 64

3.3.1 Aperçu 64

3.3.2 Étude des montants des contrats et des subventions 66

3.3.3 Étude du nombre de contrats et de subventions 70

3.3.4 Notes 74

3.4 Statistiques globales des réseaux 74

CHAPITRE 4 ANALYSE ET RÉSULTATS 83

4.1 Analyse des données 83

4.1.1 Variables d'analyse 83

4.1.2 Analyse de régression 86

4.1.3 Analyse de corrélation 95

4.2 Résultats et discussion 96

4.2.1 Résultats des régressions 97

4.2.2 Résultats de corrélation 129

CONCLUSION 131

BIBLIOGRAPHIE 136

ANNEXES 142

LISTE DES TABLEAUX

Tableau 1.1: Secteurs de performance en biotechnologie au Canada tiré de Munn-Venn (2005)... 7

Tableau 2.1: Mots-clés utilisés pour la recherche des articles scientifiques en biotechnologie 29

Tableau 2.2 Grappes liées au Québec 32

Tableau 2.3 Grappes hors Québec 32

Tableau 2.4: Catégories du financement 34

Tableau 2.5: Types de financement retirés de la table de correspondance 35

Tableau 3.1 Différence au niveau de la comptabilisation des articles publiés annuellement 56

Tableau 4.1 Types de variables d'analyse 84

Tableau 4.2: Sections des modèles d'analyse 85

Tableau 4.3 Modèles basés sur des réseaux d'un an 88

Tableau 4.4 Modèles basés sur des réseaux de trois ans 89

Tableau 4.5 Modèles basés sur des réseaux de cinq ans 89

Tableau 4.6 Statistiques AIC et BIC de sélection de modèles pour les réseaux d'un an 97

Tableau 4.7 Statistiques AIC et BIC de sélection de modèles pour les réseaux de trois ans 98

Tableau 4.8 Statistiques AIC et BIC de sélection de modèles pour les réseaux de cinq ans 98

Tableau 4.9 Comparaison des analyses #4 et #7 (aucun décalage) - 3 ans 101

Tableau 4.10 Comparaison des analyses #5 et #8 (décalage un an) - 3 ans 102

Tableau 4.11 Comparaison des analyses #6 et #9 (décalage deux ans) - 3 ans 103

Tableau 4.12 Comparaison des analyses #4 et #7 (aucun décalage) - 5 ans 104

Tableau 4.13 Comparaison des analyses #5 et #8 (décalage un an) - 5 ans 105

Tableau 4.14 Comparaison des analyses #6 et #9 (décalage deux ans) - 5 ans 106

Tableau 4.15 Analyses #4, #5 et #6 pour les réseaux sur 3 ans 108

Tableau 4.16 Analyses #4, #5 et #6 pour les réseaux sur 5 ans 109

Tableau 4.17 Résultats de régression pour l'analyse #5 - 3 ans - L1 111

Tableau 4.18 Résultats de régression pour l'analyse #5 - 5 ans - L1 112

Tableau 4.19 Variations du modèle B11 113

Tableau 4.20 Mesures AIC/BIC pour les variations du modèle B11 113

Tableau 4.21 Résultats de régressions pour le modèle B11_1 114

Tableau 4.22 Résultats de régressions pour le modèle B11_2 115

Tableau 4.23 Résultats de régressions pour le modèle B11_3 116

Tableau 4.24 Résultats de régressions pour le modèle B11_4 117

Tableau 4.25 Résultats de régressions pour le modèle B11_5 118

Tableau 4.26 Résultats de régressions pour le modèle B11_6 119

Tableau 4.27 Résultats de régressions pour le modèle B11_7 120

Tableau 4.28 Résultats de régressions pour le modèle B11_8 121

Tableau 4.29 Catégorisation des subventions pour le modèle B11_6 123

Tableau 4.30 Catégorisation des contrats pour le modèle B11_6 124

Tableau 4.31 Mesures AIC/BIC pour le modèle B11_6 124

Tableau 4.32 Régressions pour le modèle B11_6 125

Tableau 4.33 Régressions pour le modèle B11_6AIF 126

Tableau 4.34 Matrice de corrélation du modèle B11_6 129

Tableau 4.35 Matrice de corrélation du modèle B11_6AIF 129

LISTE DES FIGURES

Figure 1.1 Cadre d'analyse adapté de (Link & Scott, 2004) 5

Figure 1.2 Réseau tous canaux (graphe complet) ("all-channel") et réseau étoile 20

Figure 2.1 Automatisation de l'extraction des données Scopus 30

Figure 2.2 Structure de la table ARTICLES de la base de données Scopus 30

Figure 2.3 Structure de la table SCIENTISTS de la base de données Scopus 31

Figure 2.4 Génération de la table de mappage des grappes 33

Figure 2.5 Utilisation de la table de mappage des grappes pour la génération de la matrice 33

Figure 2.6 Aperçu des opérations 36

Figure 2.7: Capture et nettoyage de la base de données Scopus. 37

Figure 2.8: Création de la matrice d'analyse 37

Figure 2.9: Étapes du principe général d'analyse des données 38

Figure 2.10 Format du fichier d'entrée pour Pajek 40

Figure 2.11 Mécanisme d'automatisation de la création du fichier d'entrée pour Pajek 41

Figure 2.12: Vue globale des tables 42

Figure 2.13 Méthode contrôlant l'ajout de mesures réseaux dans la matrice finale 43

Figure 2.14 Matrice d'analyse: identificateurs, grappes et articles 43

Figure 2.15 Matrice d'analyse: contrats, subventions, universités et réseaux 44

Figure 2.16 Matrice d'analyse: mesures réseaux 44

Figure 2.17: Transformations de la matrice pour l'obtention du fichier de données Stata 45

Figure 3.1 Financement provincial annuel R&D (*Science Statistics*, 2008) 47

Figure 3.2 Financement provincial [sciences naturelles et ingénierie] (*Science Statistics*, 2008). 48

Figure 3.3: Articles scientifiques publiés annuellement en biotechnologie 49

Figure 3.4 Articles scientifiques publiés annuellement 1985-2005 50

Figure 3.5 Nombre moyen d'articles publiés annuellement par chercheur 51

Figure 3.6 Nombre moyen d'articles publiés annuellement par chercheur 1985-2005 51

Figure 3.7 Nombre moyen annuel de co-auteurs par article 52

Figure 3.8 Nombre moyen annuel de co-auteurs par article 1985-2005 53

Figure 3.9 Articles publiés annuellement par grappe 55

Figure 3.10 Articles publiés annuellement par grappe - Montréal 57

Figure 3.11 Articles publiés annuellement par grappe - Québec 58

Figure 3.12 Articles publiés annuellement par grappe - Sherbrooke 58

Figure 3.13 Chercheurs actifs annuellement en biotechnologie 59

Figure 3.14 Chercheurs actifs annuellement en biotechnologie 1985-2005 60

Figure 3.15 Nombre de collaborateurs annuels par chercheur 61

Figure 3.16 Nombre de collaborateurs annuels par chercheur 1985-2005 62

Figure 3.17 Nombre de collaborations annuelles par chercheur. 63

Figure 3.18 Nombre de collaborations annuelles par chercheur 1985-2005 63

Figure 3.19 Financement annuel des chercheurs affiliés 1985-2005 - SIRU 65

Figure 3.20 Proportion des catégories de financement selon les montants - SIRU 66

Figure 3.21 Montants des contrats ventilés annuellement - SIRU 67

Figure 3.22 Montants des subventions ventilées annuellement - SIRU 68

Figure 3.23 Montants des contrats et des subventions par université québécoise - SIRU 69

Figure 3.24 Proportion du nombre de contrats et de subventions - SIRU 70

Figure 3.25 Nombre de contrats ventilés annuellement - SIRU 71

Figure 3.26 Nombre de subventions ventilées annuellement - SIRU 72

Figure 3.27 Nombre de contrats et subventions pour les établissements québécois - SIRU 73

Figure 3.28 Diamètre maximal des réseaux 75

Figure 3.29 Centralisation des réseaux 75

Figure 3.30 Nombre de composantes 76

Figure 3.31 Distribution de fréquences des singletons pour les réseaux 1y 77

Figure 3.32 Taille de la super composante pour les réseaux 1y 78

Figure 3.33 Distribution de fréquences des composantes pour les réseaux 1y 78

Figure 3.34 Distribution de fréquences des singletons pour les réseaux 3y 79

Figure 3.35 Taille de la super composante pour les réseaux 3y 79

Figure 3.36 Distribution de fréquences des composantes pour les réseaux 3y 80

Figure 3.37 Distribution de fréquences des singletons pour les réseaux 5y 80

Figure 3.38 Taille de la super composante pour les réseaux 5y 81

Figure 3.39 Distribution de fréquences des composantes pour les réseaux 5y 81

Figure 4.1 Processus de sélection de l'année de référence 87

Figure 4.2 Analyse # 5 90

Figure 4.3 Construction des modèles 91

Figure 4.4 Premier groupe d’analyses des modèles 91

Figure 4.5 Deuxième et troisième groupes d’analyses des modèles 92

Figure 4.6 Exemple de régression d’un modèle binomial négatif en format Stata 10 94

Figure 4.7 Distribution du ratio IRR pour les années 128

LISTE DES ÉQUATIONS

Équation 1.1 Degré de centralisation d'une composante 19

Équation 2.1 Modèle général 23

Équation 4.1 Mesures AIC et BIC..... 93

Équation 4.2 Test statistique z 94

LISTE DES SIGLES ET ABRÉVIATIONS

La liste des sigles et abréviations, dans l'ordre alphabétique, utilisés dans le mémoire ainsi que leur signification.

IPC	Indice des prix à la consommation.
MELS	Ministère de l'Éducation, du Loisir et du Sport.
R&D	Recherche et développement.
Scopus	Base de données d'Elsevier contenant des résumés d'articles et des citations pour la littérature de la recherche scientifique.
SIRU	Système d'information sur la recherche universitaire.
SRI	Système régional d'innovation.
OST	Observatoire des sciences et des technologies.

LISTE DES ANNEXES

ANNEXE A - AUTOMATISATION 142

ANNEXE B - TYPE DE FINANCEMENT SIRU 146

ANNEXE C - MODÈLES D'ANALYSE DE RÉGRESSION 151

ANNEXE D - RÉSULTATS DES MESURES AIC/BIC 160

INTRODUCTION

Ce mémoire est une étude de l'impact du *financement public* des chercheurs universitaires québécois dans le domaine de la *biotechnologie*.

Le financement public de la recherche universitaire est un sujet qui sera toujours d'actualité, quel que soit le contexte économique. Par contre, on note une recrudescence de l'intérêt public pour le sujet depuis l'intégration formelle de la recherche universitaire aux plateformes gouvernementales visant à créer l'innovation. Le discours du trône de 2004 identifiait la biotechnologie comme une des trois plateformes technologiques essentielles à la prospérité du Canada et pouvant transformer positivement la société (Munn-Venn, 2005). Cette situation amène une pression supplémentaire sur les organismes subventionnaires quant à la justification des fonds publics alloués aux différents projets mais aussi quant à la communication des résultats obtenus. La justification de l'importance des sommes allouées nécessaires à la recherche universitaire mais aussi la vulgarisation des résultats de recherche à la population, notamment dans le cas de la recherche fondamentale, restent des objectifs non triviaux pour les organismes subventionnaires.

La biotechnologie, de son côté, est un jeune mais vaste domaine de recherche regroupant un nombre important de spécialités liées à la biologie et à la technologie. Les projets, autant que les résultats, de par leurs caractéristiques très techniques et spécialisées sont difficilement accessibles par la population si bien que les images véhiculées par les médias varient souvent de « l'invention du siècle pouvant révolutionner le monde » à la pure science fiction. Cette situation aura évidemment un impact sur les politiques gouvernementales du financement de la recherche.

Jusqu'à présent, peu d'analyses ont fait le lien entre le niveau de financement et la production de connaissances pouvant mener à des innovations. Peu font le lien entre le réseau de co-auteurs, le financement et les publications qui en découlent. Celles qui l'ont fait utilisaient des mesures de base comme le nombre de publications ou citations par dollar investi mais à notre connaissance aucunes ne font de liens entre les mesures et les conséquences résultantes surtout dans un contexte de financement sur plusieurs années.

L'objectif est donc de palier cette situation avec la construction et l'analyse des réseaux de collaboration de co-auteurs canadiens dans le domaine de la biotechnologie de façon à valider le modèle développé dans le contexte québécois. Ce modèle veut codifier les facteurs clés expliquant la production d'articles scientifiques et faire le lien avec le financement.

Ce travail de recherche contribue à l'élaboration d'une *méthodologie* permettant de mesurer l'impact du financement de la recherche scientifique universitaire et de son application à fournir des données empiriques pour illustrer le modèle.

Le modèle développé pour l'analyse de l'impact du financement de la recherche universitaire en biotechnologie permettra de valider les hypothèses posées. Celles-ci font le lien avec les acteurs (chercheurs, gouvernements et industries), le domaine de recherche, le financement et la connaissance générée dans les réseaux.

La première hypothèse, est la suivante:

Les chercheurs financés par les organismes subventionnaires ont une plus grande propension à collaborer par l'intermédiaire de publications scientifiques et ils ont, par le fait même, un réseau d'innovation plus étendu, et contribuent de façon plus active aux flux de connaissance du réseau;

La seconde hypothèse, est la suivante :

Les chercheurs financés par les organismes subventionnaires produisent plus d'articles en collaboration avec d'autres chercheurs universitaires, et de par leur position centrale dans les réseaux d'innovation contribuent à transmettre la connaissance à travers le milieu universitaire.

Le réseau canadien des collaborations entre les chercheurs en biotechnologie a été construit mais l'analyse du financement s'effectuera sur la partie québécoise du réseau, les données pour les autres provinces n'étant pas disponibles au moment de la rédaction.

Le domaine d'application est la biotechnologie. Nous examinerons en particulier les aspects du financement des réseaux de collaboration et d'innovation par l'intermédiaire d'une analyse des réseaux sociaux issue de la bibliométrie et de la scientométrie.

La collaboration avec plusieurs organismes tels le Ministère de l'Éducation, du Loisir et du Sport (MELS) à travers le système d'information sur la recherche universitaire (SIRU), l'Observatoire des sciences et des technologies (OST), la base de données Scopus d'Elsevier et Statistique Canada nous a permis d'accéder à de nombreuses bases de données. L'utilisation d'outils d'automatisation tels PatentBot[1], MatrixGenerator[2] pour la récupération et le formatage des données, ont permis d'obtenir un ensemble de données cohérent pour nos analyses.

L'objectif de ce mémoire est l'étude de l'incidence du financement public de la recherche sur la structure des réseaux et les caractéristiques des chercheurs qui les composent. L'approche envisagée pour l'analyse nécessitera l'intégration de données sur le financement de la recherche des universitaires aux données et mesures réseaux. Les réseaux, d'où les mesures seront tirées, se basent sur les liens existants entre les co-auteurs des articles scientifiques. Nous déterminerons si les chercheurs subventionnés ont un plus grand nombre de collaborateurs (degré de centralité), s'ils sont plus près des autres chercheurs (centralité de proximité), s'ils sont des joueurs importants de la transmission de connaissance (centralité de la relation d'intermédiaire). Nous construirons aussi des indicateurs simples comme le nombre d'articles scientifiques par dollar investi. À l'aide de modèles statistiques simples, nous étudierons les *corrélations* qui existent entre ces indicateurs et l'obtention du financement public de la part des chercheurs universitaires. Ces mesures des réseaux d'innovation seront obtenues à l'aide du logiciel Pajek[3] et leur analyse nécessitera l'utilisation du logiciel Stata[4]. L'analyse de l'évolution temporelle de ces réseaux, au fil des collaborations entre chercheurs, permettra de comprendre l'incidence du financement sur la structure de ceux-ci. Nous recalculerons ainsi les indicateurs de la position des chercheurs au sein des réseaux pour des périodes d'un, trois et cinq ans. L'utilisation de plusieurs fenêtres temporelles nous permettra de trouver la fenêtre temporelle ou la combinaison de fenêtres qui, à l'aide des modèles statistiques présentés, expliqueront le mieux les données. L'ajout d'une dimension temporelle aux modèles statistiques permettra d'évaluer si le fait de recevoir du financement public pour ses recherches rend un chercheur plus central ou plus important dans un réseau d'innovation.

---

[1] Logiciel d'automatisation de capture de données de Scopus. Développé par Ahmad Barirani.
[2] Logiciel générant une matrice d'analyse de réseaux. Développé par Maxime Clerk-Lamalice
[3] Logiciel d'analyse et de visualisation de grands réseaux. Développé par Vladimir Batagelj et Andrej Mrvar.
[4] Logiciel d'analyse statistique. Développé par la firme StataCorp.

Ce projet de recherche est caractérisé par plusieurs phases clés réalisées sur une période de deux années, soit de l'été 2007 à l'automne 2009. L'organisation de ces phases clés débute par le ***recensement de la littérature*** existante touchant directement le sujet du financement public de la recherche universitaire. Cette phase permet de situer l'état actuel de la recherche et de voir, à priori, les secteurs d'intérêts pour les futurs résultats de la recherche. La seconde phase consiste en ***l'extraction de données*** de multiples sources puis, dans un second temps, du nettoyage méthodique et de la normalisation de ces bases de données. La troisième phase représente ***l'intégration des données*** dans un but d'analyse statistique. Cette phase critique a donc comme résultat une matrice de données. On notera qu'une une partie importante de l'effort pré-analyse a été mis à l'automatisation de la création de cette matrice d'analyse. La quatrième phase, soit la ***caractérisation des données*** nous amène naturellement à la cinquième et sixième phases soit respectivement ***l'analyse statistique*** et la ***discussion des résultats*** concluront cette aventure intellectuelle.

Ce mémoire pose les bases d'un ambitieux projet de recherche visant à analyser l'impact du financement de la recherche universitaire. Le CHAPITRE 1 présente les concepts importants à cette étude. Le CHAPITRE 2 pose les hypothèses en plus de définir les différentes sources des données et la méthodologie. Le CHAPITRE 3 permet d'illustrer les données avant de passer à l'analyse de celles-ci au CHAPITRE 4. Bien que la recherche universitaire peut mener à du développement de produits s'inscrivant ainsi dans un cycle complet de recherche et développement (R&D), l'analyse portera spécifiquement sur la partie recherche universitaire de la R&D. Une des contributions de cette recherche est donc la combinaison de la scientométrie et des données financières détaillées du financement en portant une attention particulière à l'historique à court et moyen termes des variables clés. L'objectif final est donc de créer une méthodologie permettant l'évaluation de *l'impact réel* du financement sur la création de l'innovation émanant du secteur universitaire.

CHAPITRE 1 REVUE DE LITTÉRATURE

L'objectif de ce chapitre est d'introduire le sujet du mémoire en identifiant et présentant la littérature en lien avec les thèmes abordés dans ce travail de recherche. Ce chapitre est divisé en trois sections. La première section introduit brièvement le contexte et les caractéristiques propres à la biotechnologie. La deuxième section recense la littérature traitant du financement public de la recherche universitaire. La troisième section traite de l'utilisation des réseaux sociaux pour modéliser les liens entres les chercheurs universitaires.

1.1 La recherche en biotechnologie

Le cadre fondamental à la base de notre analyse du financement de la recherche et du développement peut être généralisé selon le flux suivant :

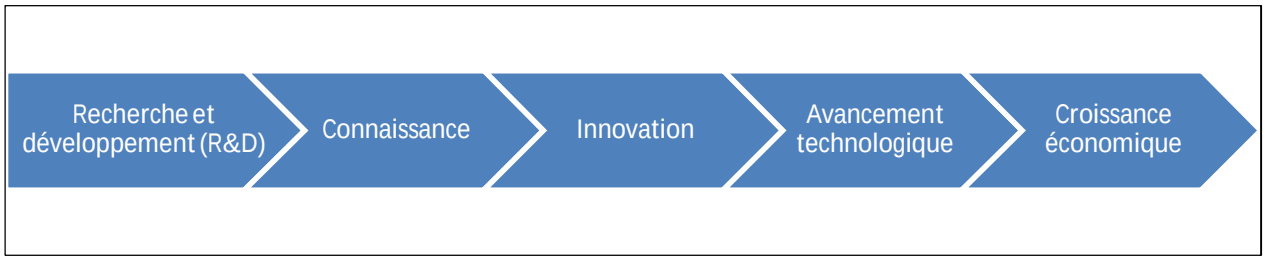


Figure 1.1 Cadre d'analyse adapté de (Link & Scott, 2004)

Tel qu'illustré à la Figure 1.1, l'étape initiale du cadre représente l'ensemble des activités de recherche et développement qui mène éventuellement à la croissance économique. C'est donc à travers ce cadre que nous allons concentrer nos efforts d'analyse sur l'impact du financement de la *recherche* universitaire. Les universités et centres de recherche affiliés constituent les éléments clés (Zucker, Darby & Armstrong, 2002) du cadre présenté à la figure précédente. C'est dans les établissements universitaires que se réalisent une majorité de projets de recherche menant à la production d'articles scientifiques. Il est important de souligner que la connaissance utilisée par les chercheurs est justement souvent appropriée à partir de ces publications de la communauté scientifique (Dasgupta & David, 2002). Ces publications, en plus de servir de sources de connaissances, deviennent même des pistes d'entrée pour des futurs projets de recherche. Créant ainsi un cycle entre l'appropriation et la production de connaissances. Voyons maintenant les spécificités du domaine de la biotechnologie dans le cadre de notre étude.

Malgré qu'aucune définition standardisée n'existe encore pour le domaine de la biotechnologie, celui-ci existe formellement depuis 1973 et est associé à Stanley Cohen et Herbert Boyer, deux chercheurs universitaires américains (Zucker, Darby & Brewer, 1998). La découverte de ces deux chercheurs, soit le développement de la technique de base pour l'ADN recombinée, a permis de poser les bases commerciales de la biotechnologie mais aussi de générer un nombre important d'innovations à travers divers laboratoires dans le monde. L'évolution internationale du domaine est le résultat naturel des expérimentations, des collaborations et des publications propulsant la croissance de ce domaine autant dans le secteur universitaire, que dans le secteur public et privé. L'aspect « développement » de la biotechnologie, soit les projets de recherche menant à des applications plus concrètes ont été exploités par une multitude de firmes, dont une grande partie en démarrage, pour créer une industrie valant maintenant plusieurs milliards de dollars. Comme pour la majorité des domaines scientifiques, la biotechnologie n'y faisant pas exception, la génération constante de nouvelles connaissances est essentielle à la continuité des recherches mais aura évidemment un coût non négligeable. Le financement, qui est au cœur de cette étude, sera donc détaillé pour le secteur universitaire. Bien que l'étude portera sur le secteur universitaire, il est important de mentionner la symbiose existante et nécessaire entre le secteur universitaire générant la connaissance (et formant les chercheurs) et le secteur privé fournissant les capitaux nécessaires dans leur domaine d'activités.

La biotechnologie est un domaine de recherche dynamique dont les moindres gestes sont rapportés par les médias autant au niveau des investissements massifs nécessaires à propulser la recherche à un niveau supérieur que des problèmes étiques liés aux expérimentations. Ce dynamisme se reflète aussi au niveau des publications scientifiques tel que discuté par Arora, David & Gambardella (1998) ayant effectué une excellente analyse des publications scientifiques en biotechnologie au sein de divers groupes de recherche. Ceux-ci concluent que l'historique des publications a un impact important (est fonction de) sur les niveaux de subventions à venir et donc sur la productivité en termes de publications scientifiques. Le financement public de la recherche en biotechnologie représente une proportion importante des revenus de recherche dans le cadre de la plateforme canadienne pour l'innovation et sera discuté dans la section 1.2.

La biotechnologie est un jeune domaine scientifique regroupant un ensemble de sous-domaines spécialisés caractérisés par la diffusion rapide des nouvelles (dernières) connaissances générées (Gittelman & Kogut, 2003). On retrouve donc globalement pour ce domaine un nombre

important d'articles scientifiques publiés depuis les deux dernières décennies tel qu'observé avec nos propres données de la biotechnologie présentées à la Figure 3.3. Bonaccorsi & Thoma (2007) font le même constat en nano science et technologie (nanotechnologie) où il y a aussi mobilisation des chercheurs dans ce domaine entrainant un taux croissant de publications depuis 1990. La nanotechnologie et la biotechnologie suivent une courbe de croissance équivalente malgré que deux décennies séparent leur naissance (Darby & Zucker, 2003). On peut donc parler de la biotechnologie en tant que domaine multidisciplinaire créant ainsi un historique de publications scientifiques très riche malgré qu'une grande partie de la base de connaissance est tacite et donc difficilement codifiable (Aharonson, Baum & Feldman, 2004). La collaboration est essentielle surtout dans un contexte canadien (Schiffauerova & Beaudry). La stratégie développée pour capturer cette richesse et diversité des publications sera présentée à la section 2.4. Au niveau canadien, les secteurs industriels de performance sont fortement liés aux secteurs de recherche universitaire et sont présentés au Tableau 1.1. On remarque la grande diversité des secteurs associés au domaine de la biotechnologie ce qui sera évidemment un enjeu lors du choix des mots-clés présents dans les requêtes de recherche des publications scientifiques.

Tableau 1.1: Secteurs de performance en biotechnologie au Canada tiré de Munn-Venn (2005)

Secteurs de performance en biotechnologie au Canada					
Santé	Agriculture et transformation des aliments	Environnement	Bio-informatique	Ressources naturelles	Aquaculture

La notion de *convergence* des projets de recherche en biotechnologie est expliquée par deux approches différentes, l'une basée sur la juxtaposition de projets liés, l'autre sur la juxtaposition de projets différents (Avenel, Favier, Ma, Mangematin & Rieu, 2007). Cette notion deviendra intéressante lors de l'étude de l'impact du financement. La convergence représente concrètement des collaborations, entre les chercheurs, basées sur l'échange d'expertises, sur le financement et sur les infrastructures dans l'optique de maintenir un taux satisfaisant de résultats demandé par les instances administratives et subventionnaires.

Toujours au niveau de la convergence des projets de recherche, la complexité de ces projets mais aussi le faible nombre de chercheurs dans ces domaines très spécialisés (Stephan, Black & Chang, 2007) forcent les spécialistes de différentes universités et centres de recherche à collaborer. Lorsque l'on fait référence à la *complexité*, il s'agit non seulement des connaissances requises pour réaliser les projets ou manipuler les instruments nécessaires aux expérimentations mais aussi pour l'évaluation et l'analyse des résultats de façon à pouvoir, dans un premier temps, s'approprier cette nouvelle connaissance, puis dans un second temps, communiquer celle-ci. La communication et la sensibilisation est un des mandats des universités autant avec ses partenaires de recherche qu'au sein du grand public (Hall, Link & Scott, 2003). Cette internationalisation des efforts de recherche présente au niveau de l'industrie sous la forme des alliances de R&D (Narula & Duysters, 2004) se retrouve aussi au niveau universitaire. On assiste donc à un ensemble de collaborations forcées visant à diminuer le *risque* des projets plus ambitieux tout en favorisant une augmentation des transferts de connaissances, du partage des équipements de recherches et aussi une redistribution plus complexe du financement public et privé. Cette mondialisation des équipes se reflète aussi au niveau du nombre de co-auteurs d'articles scientifiques qui proviennent non seulement d'universités différentes mais de pays différents (Guan & Ma, 2007).

Lorsque l'on parle de collaborations, il est important de bien définir qui sont les acteurs présents dans celles-ci. Pour le domaine de la biotechnologie, ces acteurs sont ceux qui sont typiquement présents en recherche universitaire soit :

- **Universités/Chercheurs** : Le chercheur est l'unité d'analyse de base utilisée pour l'étude du financement. Ce raffinement par rapport à travailler avec le département ou l'unité de recherche nous permet de tenir compte de l'hétérogénéité des chercheurs et de capturer sur une base individuelle les liens avec les organismes subventionnaires ou contractuels.

- Les **gouvernements** fournissent du financement aux firmes selon des critères bien établis, souvent sous forme de crédits, mais aussi aux universités, sous forme de subventions, qui nous intéressent pour notre étude. Ils suivent donc l'évolution de leurs « investissements ».

- Les **firmes** en biotechnologie, principalement des PME, trouvent des applications aux résultats théoriques. Elles établissent des alliances de R&D et financent sous forme de contrats des projets de recherche appliquée. Dans le cas de la biotechnologie en santé

humaine, les différentes phases de la recherche sont gérées par des grandes firmes de pharmaceutiques mais collaborent à divers degrés avec les universités pour les phases de recherche, les phases précliniques (tests sur les animaux), les phases cliniques (normalement dans les hôpitaux) et la commercialisation.

1.2 Le financement public de la recherche universitaire

Dans le monde de la recherche universitaire, le financement et toutes les activités connexes sont essentiels à la mise sur pied et à la continuité des projets de recherche. L'entrée de fonds au sein d'unités de recherche prend une place prépondérante, que se soit en tant que catalyseur de projets, pour l'ajout de personnel spécialisé et d'équipement plus performant ou pour assurer le bon fonctionnement du matériel de recherche. De plus, le financement est à l'origine de plusieurs comportements collaboratifs entre chercheurs universitaires. L'objectif de cette analyse est de quantifier les influences et résultats du financement de façon à pouvoir en saisir les *impacts réels* sur les activités collaboratives de la recherche. L'intérêt d'étudier l'impact du financement provient de l'importance de la recherche universitaire sur la génération de la connaissance et de l'innovation qui contribuent au *changement technologique*.

Il est important de mentionner que cette étude porte sur l'impact du financement <u>public</u> de la recherche universitaire et que plusieurs autres sources de financement existent. C'est donc une partie du financement total des chercheurs qui sera étudiée et dont la proportion est difficilement estimable à cause justement des différentes autres sources privées comme les entreprises ou même les centres interuniversitaires dont la traçabilité des transactions de financement est très complexe.

L'étude porte sur deux types de financement soit les subventions et les contrats. Bien que ceux-ci peuvent avoir des objectifs communs, rien, à priori, ne le prouve comme l'illustre la citation suivante « Research is a sponsored activity, and researcher effort is contracted » (Goldfarb, 2008, p. 44). Les contrats sont habituellement associés à des objectifs très précis, souvent à court terme, et visent la production de connaissances utilisables rapidement (Goldfarb, 2008). Les sources des contrats viennent en majeure partie des entreprises privées. De leur côté, les subventions sont

plus souvent associées à des projets de plus grande envergure et les sources de financement viennent majoritairement des organismes gouvernementaux.

L'intégration du financement de la recherche universitaire aux programmes gouvernementaux de stimulation de l'innovation est un phénomène récent (Link & Scott, 2004) mais est une stratégie de plus en plus en plus utilisée. Cette approche de financement, rendue possible grâce aux finances publiques, demande évidemment des comptes à rendre tant au niveau des sommes dépensées que des résultats. De plus, l'intervention des programmes gouvernementaux dans le processus d'innovation en fournissant à la fois le capital et la direction des recherches, crée une relation qui est, et doit être, surveillée. Il est important de noter que le financement des projets de recherche universitaire bénéficie directement au secteur académique mais aussi indirectement au secteur industriel et les résultats atteignent éventuellement la population générale.

Il y a donc émergence d'une pression au niveau des universités, à travers les chercheurs subventionnés, qui est liée à la performance des projets de recherche. Toute la question de réussite et de performance au niveau de la recherche universitaire est évidemment très subjective. Il est très difficile de comparer sur une base commune des résultats de recherche appliquée (souvent associée aux brevets qui ne seront pas traités dans le cadre de ce mémoire) avec des résultats de recherche fondamentale (souvent associée aux publications). L'objectif de la présente analyse étant de modéliser la création même de ces résultats à travers les publications et non pas d'évaluer la qualité des résultats, laissant cette tâche à un futur projet de recherche.

Nous savons, à prime abord, que dans la plupart des cas les projets de recherche universitaire sont financés par le secteur public (Hagedoorn, Link & Vonortas, 2000). C'est donc de reconnaitre l'importance du financement public où celui-ci est évidemment dirigé selon la politique de développement de l'innovation (Hart, 2001). L'objectif général de ce type de politique étant la création et le maintient de centres de compétences à travers le réseau universitaire. Ces derniers, étant constitués des universités et des centres de recherche affiliés, est une partie intégrante du système d'innovation national (Hall et al., 2003).

Le financement de la recherche est aussi utilisé en tant qu'outil d'influence et comme incitatif pour promouvoir un secteur particulier ou forcer la collaboration entre chercheurs. On peut alors faire référence à la *théorie des incitations* qui décrit l'influence possible d'un agent économique, qui pourrait prendre la forme d'une subvention ou un contrat, sur le type de comportement d'un

autre agent qui serait, dans notre cas, un chercheur universitaire. On peut donc lier l'incitation possible avec l'information disponible, soit les deux thèmes principaux abordés par cette théorie. Dans cette optique, le financement peut être vu comme un mécanisme pour obtenir un résultat voulu qui souvent s'inscrit dans une politique nationale. On note que Zucker, Darby, Furner, Liu & Ma (2007) abondent dans le même sens en insistant sur la nécessité d'avoir des collaborations inter-institutions pour favoriser l'échange d'informations. Les centres de recherche multi-projets ont comme effets de forcer les chercheurs et leurs universités d'attache à collaborer plus efficacement à l'aide de la diversité des ressources disponibles. Plusieurs auteurs dont Daniel, Hempel & Srinivasan (2003) s'intéressent aux avantages des collaborations et à la gestion des projets de recherche dans le but d'obtenir du financement.

Une des caractéristiques majeures du domaine de la biotechnologie est l'importance des infrastructures nécessaires pour réaliser les projets de recherche (Robinson, Rip & Mangematin, 2007). Lorsqu'on parle d'infrastructure, il ne s'agit pas seulement de l'achat des machines complexes ou d'instruments plus précis, mais de leur fonctionnement et de la maintenance qui nécessitent du personnel spécialisé (compétences) et de la logistique pour optimiser le temps d'accès des différentes équipes de recherche.

On parle donc de financement important sur de longues périodes visant la création d'une *plateforme technologique* qui n'est pas seulement un ensemble d'équipements spécialisés mais un lieu où les opportunités peuvent émerger. Les universités et centres de recherche affiliés sont des lieux naturels pour favoriser la création des plateformes technologiques. L'objectif ultime d'une plateforme technologique étant la création de connaissances pouvant mener à des innovations technologiques.

Le financement peut intervenir à plusieurs étapes du cycle de vie d'un projet de recherche présenté à la première étape du cadre d'analyse de la Figure 1.1 et dans bien des cas, les projets se basent sur les résultats de projets passés. Le cycle de vie d'un projet de recherche représente l'ensemble des étapes suivies par un chercheur ou par une équipe de recherche à partir de l'idée initiale jusqu'au suivi du projet une fois celui-ci complété. Un projet universitaire génère habituellement des publications scientifiques. Dans le cadre des projets universitaires, on peut faire le lien avec le concept de capacité d'absorption (« absortive capacity ») (Cohen & Levinthal, 1990) représentant la capacité de reconnaitre l'importance et d'exploiter des

connaissances venues de l'extérieur de la firme. Ce concept de capacité d'absorption, étudié chez les firmes, est développé grâce au financement reçu au niveau des unités de recherche. Le financement public permet de la même façon d'accroître les capacités d'une équipe de recherche et donc d'augmenter les chances d'obtenir des résultats publiables.

Plusieurs auteurs font le lien entre financement et succès de la recherche. Par exemple, Aharonson, Baum & Plunket (2008) font référence aux régions inventives (« inventive locations ») qui ont plus d'investissements en R&D autant du public que du privé. Par contre, plusieurs émettent des réserves quand à l'impact de la source du financement. La littérature propose plusieurs études sur les liens entre l'industrie et la productivité scientifique. Par exemple, Goldfarb (2008) examine la question des limitations du système de subventions à travers les liens entre le financement par les contrats gouvernementaux. Il constate même une décroissance des publications des chercheurs universitaires qui gardent un lien avec l'organisme contracteur une fois le contrat réalisé.

Il faut aussi tenir compte que la préparation et la rédaction des demandes de financement est une activité coûteuse en temps et en personnel de recherche. Autant pour les contrats que pour les subventions, ces demandes de financement nécessitent des ressources temporelles importantes, la différence se situant aux niveaux des étapes requises. Ainsi pour les contrats, les délais se situent au niveau de l'approbation par les spécialistes des bureaux de la recherche et par les avocats des entreprises sur les clauses des contrats. Pour les subventions, les délais se situent au niveau de la construction des dossiers nécessaires à la présentation des projets.

De leur côté, Payne & Siow (2003) et Blume-Kohout, Kumar & Sood (2009) parlent d'une influence positive de l'augmentation depuis la dernière décennie du financement public de la recherche universitaire dans les secteurs touchant la science. De plus, le financement public devrait être vu comme complémentaire au financement privé, augmentant la valeur totale des investissements en R&D. Le choix du financement public des projets de recherche peut aussi être perçu comme un signal de qualité au niveau des chercheurs récipiendaires mais aussi pour leur université d'attache. Cette situation favorise ainsi le financement futur autant de la part du public que du privé.

Certains auteurs comme Zucker et al. (2007) parlent d'un impact majeur du financement autant au niveau de la publication d'articles scientifiques que de la création de brevets. Quelques

résultats chiffrés vont dans le même sens et nous permettent de voir une tendance positive du financement. L'étude présentée par Payne & Siow (2003) trouvent avec un décalage (« lag ») d'une année pour le financement public et de trois années pour les instruments qu'un financement de 1 million de dollars permet de publier entre 8 et 11 nouveaux articles annuellement (la variation provient de l'estimateur utilisé soit la méthode des moindres carrés (OLS) ou variables instrumentales (IV). De leur côté, Jacob & Lefgren (2007) trouvent qu'une subvention de recherche de type F32 (« Postdoctoral Individual National Research Service Award ») ou R01 (« Research Project (Traditional) ») du « National Institutes of Health » (NIH) mène à une publication additionnelle lors des cinq années suivant la réception de la subvention.

La recherche et le développement effectués en biotechnologie et dans les sous-domaines spécialisés sont liés aux centres de recherche où se trouvent le matériel et les compétences pour l'opération des instruments de recherche. Il est intéressant de constater que plusieurs des projets de recherche visent justement le développement d'instrumentation et de processus plus performants. Il y a donc création d'un cycle où le financement crée une demande accrue de financement dans le but de pousser encore plus loin les limites de la recherche mais aussi pour faire face à une l'utilisation croissante et au maintient des équipements.

Tel que discuté précédemment, des investissements majeurs sont nécessaires à la mise en place d'infrastructures et de personnel possédant les compétences techniques pour l'opération des systèmes. Dans un contexte de mobilité accrue des chercheurs, le concept d'*exportabilité* prend son importance. Le financement public engendre la création de deux catégories de résultats : fixes et exportables. Par exemple :

- Mise en place d'outils spécialisés et d'instrumentation [fixe];
- Formation de personnel ayant les connaissances et compétences [exportable];
- Utilisation d'infrastructures fixes peuvent être utilisées par plusieurs chercheurs de la même grappe[5] mais aussi de grappes différentes [fixe et exportable];

[5] Le concept de grappe utilisé dans ce travail est un ensemble d'une ou plusieurs universités et des centres de recherche affiliés. La définition formelle est présentée à la section 1.3.1.1.

- Utilisation de résultats pouvant mener à des articles (connaissance) et à des innovations [fixe et exportable].

Ainsi la majorité des investissements en infrastructure et en formation visant la mise en place des compétences locales sont difficilement exportables ce qui justifie l'augmentation de la mobilité nationale voire internationale. Les systèmes régionaux d'innovation (SRI) englobent toutes les institutions et organisations dans une région spécifique et permettent de répondre aux besoins spécialisés des chercheurs. Chacun de ces SRI peut donc devenir une référence dans des domaines ou sous-domaines spécifiques et l'outil stratégique qu'est le financement public vient solidifier cette assise.

En terminant cette section, on note que le chercheur financé, que ce soit par des contrats ou des subventions, devra travailler en *collaboration* avec d'autres chercheurs ou bien avec des étudiants ou des techniciens de façon à prendre avantage du financement reçu mais aussi pour résoudre des problèmes liés à la complexité du domaine.

1.3 Les réseaux sociaux

Nous avons vu, grâce à la section précédente, l'importance qu'occupe le financement alloué aux *chercheurs* dans le cadre de leurs différents projets de recherche. Il est bien rare que les sommes reçues soient intégralement utilisées par le récipiendaire mais l'est plutôt par son *équipe* de recherche. Lorsqu'on fait référence aux équipes de recherche, cela implique nécessairement des *liens* entre les individus.

L'analyse des caractéristiques des réseaux, tirée de la théorie des graphes, est utilisée dans plusieurs domaines de recherche pour approfondir la compréhension des phénomènes observés dans ces réseaux. Ce type d'analyse est donc de mise pour saisir l'impact du financement dans les projets de recherche à travers les liens entre les chercheurs.

Il est important de préciser que dans le cadre de notre étude un *lien* représente la collaboration menant à une publication scientifique conjointe entre deux ou plusieurs chercheurs. Pour une même publication, chacun des co-auteurs est lié aux autres co-auteurs par un lien de collaboration. Les liens servent à canaliser l'information et chacun des chercheurs agit comme un émetteur et un récepteur d'informations et de connaissances (Ahuja, 2000). La structure des liens

aura un impact direct sur la dynamique du réseau, qu'on parle de liens directs ou indirects entre les chercheurs. Différentes mesures des réseaux présentées à la section 1.3.1 nous permettent d'évaluer la diffusion de l'information et d'établir l'impact du financement sur celle-ci.

Les chercheurs sont associés à une institution principale donc indirectement les liens entre les institutions peuvent être créés. La force des liens entre les institutions présentes dans le réseau permet une dynamique dans le transfert de la connaissance (Bozeman, Laredo & Mangematin, 2007) qui rapproche la science et la technologie. Cette connaissance qui transige à travers le réseau est fonction de ce qui a été développé dans le passé d'où l'importance de la notion d'historique pour notre étude.

Le concept d'*historique* des réseaux d'individus fait référence à la structure de la collaboration passée. Ce concept étant très important pour expliquer les événements courants surtout lorsqu'on tient compte du fait que les subventions durent habituellement plusieurs années. Par exemple, les chaires du Canada de niveau I sont renouvelables aux 7 ans tandis que les chaires de niveau 2 le sont aux 5 ans. Ainsi certaines subventions ne durent qu'un an alors que d'autres s'étendent sur 3 à 5 ans. Les individus appartenant aux réseaux reçoivent donc ponctuellement du financement. Plusieurs auteurs utilisent donc des fenêtres d'analyses d'une durée supérieure à une année pour l'étude des réseaux. Par exemple, Schilling & Phelps (2007) utilisent des fenêtres d'analyse de 3 ans (1990-1992, 1991-1993, etc.), tandis que Gulati & Gargiulo (1999) utilisent des fenêtres d'analyse de 5 ans et Stuart (2000) utilise aussi des fenêtres de cinq ans avec un délai d'un an. L'utilisation des fenêtres d'analyse de 3 ou 5 ans permettent donc de conserver la richesse des évènements passés dans le réseau comme par exemple le nombre de subventions et contrats reçus.

L'analyse des réseaux à l'aide des fenêtres d'analyse se fait souvent en combinaison des retards de phase qui correspondent à un délai d'une ou plusieurs périodes d'analyse (une ou plusieurs années dans notre cas). Plusieurs exemples se trouvent dans la littérature, par exemple Schilling & Phelps (2007) utilisent des retards d'un (« L1 »), deux (« L2 ») et trois (« L3 ») ans pour tester la robustesse des résultats tandis qu'un délai d'un an est utilisé par Aharonson et al. (2004) pour éviter des problèmes de simultanéité avec les variables indépendantes.

L'approche typique d'analyse des réseaux tel que résumée par Cantner & Graf (2006) est la construction de réseaux à l'aide de relations de collaboration entre co-auteurs (Zucker et al.,

2007) et c'est cette approche qui sera utilisée dans le cadre de ce mémoire. Plusieurs auteurs utilisent cette approche pour la construction des réseaux, par exemple Zucker et al. (2002) font le lien entre les publications des chercheurs des universités et celles des firmes.

Bien que des réserves sont émises (Goldfarb, 2008) quant à la difficulté de lier directement la *quantité* des citations ou des publications comme mesure de la valeur sociale ou commerciale, ces données nous permettent de quantifier le lien entre les moyens qui ont permis leur création (financement) et le résultat produit (les publications). Nous pouvons aussi émettre des réserves sur la *qualité* des publications.

Les réseaux de co-auteurs nous permettent de suivre la fréquence des *collaborations* entre chercheurs et leur *mobilité* à partir de leur première publication présente dans le réseau jusqu'à leur dernière. La mobilité des chercheurs et leur importance au sein d'une grappe ont fait l'objet de plusieurs études notamment par Robinson et al. (2007).

Newman (2001) mentionne que la probabilité des co-auteurs à travailler ensemble augmente avec le nombre de collaborateurs communs et que la probabilité d'un chercheur à trouver des collaborateurs est fonction du nombre de collaborateurs passés dans un contexte de clique. La notion d'historique en tant que structure des événements passés est cruciale pour expliquer le nombre d'articles publiés à un moment précis.

Les systèmes régionaux d'innovation sont des regroupements géographiques d'institutions et de firmes dans un secteur particulier. Il est à noter que cette analyse ne tentera pas de justifier la création des SRI existantes dans les réseaux étudiés mais portera son attention sur l'évolution du réseau en fonction du financement reçu par les chercheurs qui les composent. Par contre, le financement pourra expliquer indirectement (ou directement) l'évolution des SRI. Tel que discuté dans Porter (1998), les grappes représentent à la fois la compétition et la coopération entre les membres d'une grappe. Il faut donc s'attendre à ce que les montants disponibles pour les projets de recherche soient attribués selon un processus transparent aux différentes universités d'un même SRI mais aussi au sein d'une région ou d'une province. Le processus « transparent » d'attribution des subventions nécessite l'évaluation par les pairs alors que l'attribution des contrats implique des affinités entre l'organisation et l'université. C'est donc encore une fois une forme de collaboration qui aura un impact sur le financement reçu par les chercheurs.

1.3.1 Les mesures et caractéristiques réseaux

Cette section traite des différentes mesures[6] utilisées dans le cadre de cette analyse. Le choix des mesures utilisées a été lié au développement du logiciel d'automatisation détaillé à la section 2.4.3.

L'analyse des réseaux permet de regrouper les chercheurs entre eux et d'étudier les caractéristiques selon différentes perspectives. L'importance des réseaux autant du point de vue économique que pour leur rôle de diffusion des connaissances codifiées n'est plus à prouver (Aharonson et al., 2004). Plusieurs auteurs utilisent cette approche notamment pour étudier les caractéristiques des groupes de chercheurs (Bonaccorsi & Thoma, 2007). C'est la combinaison judicieuse de l'ensemble de ces informations qui permet de comprendre la structure des réseaux de collaboration et des influences possibles sur la création de connaissance.

Le réseau sur lequel les mesures sont prises est basé sur les publications scientifiques. Il faut noter que celles-ci représentent un des moyens pour avoir une visibilité accrue au sein de la communauté scientifique et par le fait même accroître sa réputation en tant que chercheur.

La notion d'*information* est essentielle dans les réseaux sociaux. Celle-ci permet d'évaluer l'importance des sommets selon leur capacité à gérer un flux d'informations qui peut autant être entrant que sortant dans le cas de réseaux non-dirigés.

Une autre notion importante est la *diffusion* des connaissances dont on présume être présente entre les chercheurs qui collaborent mais dont on ne peut être certain.

Lorsque deux ou plusieurs chercheurs publient annuellement plus d'un article, les liens entre ces individus du réseau sont plus forts. Cette collaboration qu'on peut lier à la force des liens du réseau (Burt, 2001) permet une augmentation du flux de connaissances générées par l'expérience acquise entre ces collaborateurs pouvant mener à une collaboration plus efficace.

---

[6] L'ensemble des mesures discutées dans cette section s'appliquent à un réseau non-dirigé, car pour les liens entre les co-auteurs, soit des articles, la notion de direction de l'information n'existe pas.

1.3.1.1 Concept de grappe (« cluster »)

Le concept de grappe utilisé dans ce travail est un ensemble d'une ou plusieurs universités et des centres de recherche affiliés.

1.3.1.2 Concept de composante (« component »)

Dans un réseau non dirigé, un sous-groupe de sommets (les chercheurs pour notre étude) connectés ensembles forment une *composante*. Il n'existe pas de liens entre deux composantes d'un réseau, chacune existe comme une entité isolée. C'est donc une mesure de cohésion du réseau. Les singletons sont un type de composante de degré un.

1.3.1.3 Concept de centralité (« degree centrality »)

La centralité est une caractéristique individuelle liée à un chercheur (un sommet précis) et fait référence sa position dans le réseau (Wouter, Andrej & Vladimir, 2004). On mesure la centralité par le degré d'un sommet, soit le nombre de voisins directs. Plus le scientifique a de voisins, plus il est en contact avec des sources d'informations et plus l'information a des chances de se rendre rapidement, il est donc dit plus central.

1.3.1.4 Concept de centralisation (« centralization »)

La centralisation est une caractéristique globale du réseau. L'étendu d'un réseau aura évidemment une influence sur ses propriétés de diffusion (Schilling & Phelps, 2007) ainsi un réseau très centralisé en est un où la majorité des liens passent par un nombre restreints de sommets diminuant la distance entre les sommets et permettant un transfert d'information plus efficace. Ce concept permet d'évaluer la relation entre le centre du réseau et sa périphérie en termes de transfert d'information (Wouter et al., 2004). Il permet aussi de différencier les réseaux entre eux (Gulati & Gargiulo, 1999).

1.3.1.5 Degré de centralisation d'une composante (« component degree centralization »)

Le concept de degré de centralisation est adapté de la formule générale de Freeman (1979) pour la centralisation des graphes. Cette mesure nous permet de mesurer l'étendue ou la variabilité des mesures des individus (sommets) pour une composante en comparant chacune des mesures avec la valeur maximale atteinte. Le degré de centralisation calculé pour chacune des composantes est

donc la centralisation des composantes. Nous utilisons donc l'Équation 1.1 tirée de Wasserman & Faust (1994) :

$$C_D = \frac{\sum_{i=1}^{g}\left[C_D(n^*) - C_D(n_i)\right]}{\left[(g-1)(g-2)\right]}$$

Équation 1.1 Degré de centralisation d'une composante

Où g représente l'ensemble des sommes de la composante;

$C_D(n_i)$ représente le degré de centralité du sommet i;

$C_D(n^*)$ représente la plus grande valeur de la centralité de la composante;

La valeur maximale de 1 représente la situation du graphe étoile où un sommet est connecté avec l'ensemble des autres. La valeur minimale de 0 signifie que tous les degrés des sommets sont égaux, soit celui du graphe circulaire.

1.3.1.6 Concept de centralité de relation d'intermédiaire (« betweenness centrality »)

La relation d'intermédiaire (« betweenness ») est une propriété structurelle d'un réseau et provient de la théorie des graphes. Le concept, défini par Freeman (1977), est applicable aux réseaux connectés ainsi qu'aux réseaux non connectés. Il est lié au contrôle du flux d'information dans le réseau et vient palier le problème observé par Moxley & Moxley (1974) quant au besoin d'avoir un réseau connecté pour obtenir des mesures de centralité significatives.

Cette mesure est basée sur la capacité d'un chercheur à « contrôler » le flux d'informations en fonction de sa position dans le réseau. C'est donc l'*importance d'une personne pour la transmission des informations dans un réseau*. Un chercheur qui fait parti du plus grand nombre de lignes géodésiques[7] entre les paires de sommets a donc une position plus centrale. Comme suggéré par Bavelas (1947), cette position centrale permet d'influencer les communications qui

[7] Ligne la plus courte qu'on puisse mener entre deux points donnés d'une surface.

passent par cet individus. On peut ainsi parler du « potentiel » (Cohn & Marriott, 1958) de ces sommets centraux. De son côté, Goldfarb (2008) parle de chercheurs de catégorie « mid-level » et « high-ability » pour catégoriser l'importance des chercheurs dans le réseau. Schilling & Phelps (2007) utilisent cette mesure pour contrôler la possibilité qu'une firme occupant une position centrale générera plus d'innovations que celle qui est située en périphérie. On applique ce concept à la position des chercheurs où on s'attend à ce que les plus centraux publient plus.

1.3.1.7 Concept de relation d'intermédiaire de centralisation («betweenness centralization»)

Pour un réseau, c'est la variation de la relation d'intermédiaire de centralité des sommets divisé par les variations maximales des relations d'intermédiaire de centralité possibles d'un réseau de la même taille (Wouter et al., 2004). Un réseau considéré structurellement centralisé s'approche du modèle « étoile » de réseau tandis qu'un réseau décentralisé est un ensemble de canaux (« all-channel ») (Mackenzie, 1966). La Figure 1.2 présente les deux modèles de réseaux mentionnés précédemment.

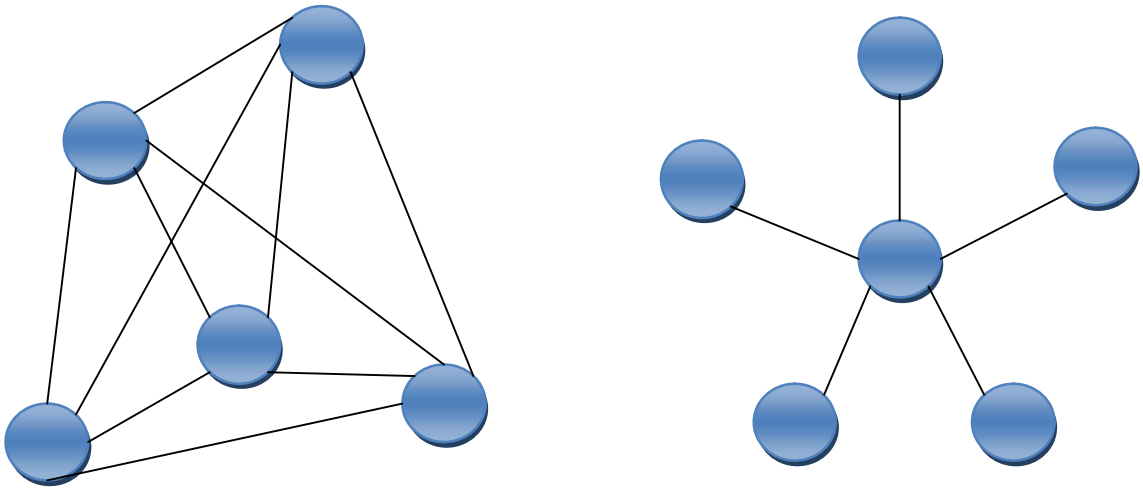


Figure 1.2 Réseau tous canaux (graphe complet) ("all-channel") et réseau étoile

1.3.1.8 Coefficient de groupement local (« local clustering coefficient »)

La mesure de coefficient de groupement local a été introduite par (Watts & Strogatz, 1998) sous la forme d'un coefficient qui quantifie à quel point les voisins d'un sommet peuvent constituer une *clique* (graphe complet). C'est une mesure de *sommet* représentant la fraction des liens possibles entre ses sommets voisins. Elle représente l'interconnexion du réseau et on peut parler de « cliquishness » ou « egocentric density ».

Les valeurs possibles de cette mesure varient entre 0 et 1. Une valeur de 0 représente un sommet où chacun de ses voisins ne sont pas liés entre eux (aucun lien entre les voisins). À l'autre extrême, une valeur de 1 représente un graphe complet, c'est-à-dire où tous ses voisins sont liés entre eux.

1.3.1.9 Coefficient de groupement global (« global clustering coefficient »)

Le coefficient de groupement global pour le réseau complet a été défini par (Watts & Strogatz, 1998) comme étant la moyenne des coefficients de groupement local pour l'ensemble des sommets.

1.3.1.10 Concept de « small world »

Cowan & Jonard (2004) étudient la relation entre la structure du réseau et la performance de la diffusion de la connaissance pour trouver que les systèmes dits « small world » sont les plus performants. Un système est dit « small world » lorsque la plupart des liens entre les sommets sont locaux. Une étude des « small world » permet de déterminer à quel point les relations entre les individus sont basées sur les individus communs mais aussi sur les individus intermédiaires (les individus intermédiaires font partis de la chaîne liant deux individus du réseau). Les échanges de connaissances sont plus élevés entre les sommets situés dans la région « small world ». Un lien pourra évidemment être établi entre les collaborations entre chercheurs et leur « performance » de publication.

CHAPITRE 2 HYPOTHÈSES ET MÉTHODOLOGIE

Dans ce chapitre, nous présenterons les hypothèses qui sont au cœur du processus d'analyse des données. Nous documenterons par la suite les étapes nécessaires à la vérification de ces hypothèses. Ce chapitre est divisé en quatre sections. Premièrement, les hypothèses guidant notre démarche seront posées. Deuxièmement, le modèle général sera présenté. Troisièmement, nous détaillerons les sources de données utilisées. La dernière section décrira la méthodologie suivie pour obtenir les données puis construire automatiquement différents réseaux ayant des caractéristiques propres. Les caractéristiques des réseaux étudiés seront aussi détaillées dans cette quatrième section.

2.1 Les hypothèses

Les hypothèses posées permettent de faire le lien entre le sujet de notre recherche et la littérature consultée et présentée au CHAPITRE 1. On tente donc d'approfondir le sujet de l'impact financement des chercheurs se situant dans un certain réseau de collaboration (très certainement plusieurs) et la production scientifique calculée ici par les articles scientifiques.

La première hypothèse, est la suivante:

Les chercheurs financés par les organismes subventionnaires ont une plus grande propension à collaborer par l'intermédiaire de publications scientifiques et ils ont, par le fait même, un réseau d'innovation plus étendu, et contribuent de façon plus active aux flux de connaissance du réseau;

La seconde hypothèse, est la suivante :

Les chercheurs financés par les organismes subventionnaires produisent plus d'articles en collaboration avec d'autres chercheurs universitaires, et de par leur position centrale dans les réseaux d'innovation contribuent à transmettre la connaissance à travers le milieu universitaire;

Maintenant que ces hypothèses sont formulées nous devons nous donner les outils pour pouvoir confirmer ou infirmer ces hypothèses. Pour ce faire, nous allons construire un modèle statistique qui nous permettra d'analyser les données et en comprendre les tendances.

2.2 Le modèle général

Ces hypothèses nous amènent à construire un modèle général paramétré ayant pour objectif de tester l'ajustement (« fit ») des données avec celui-ci. Ce modèle, présenté par l'Équation 2.1, veut expliquer le nombre d'articles scientifiques publiés par un chercheur universitaire au cours d'une année en fonction du financement reçu dans le passé et de la structure des collaborations passées.

$$\boxed{\begin{aligned} A_{it} &= \beta_G G_{it-d} + \beta_{TG} TG_{it-d} + \beta_{TMG} TMG_{it-d} + \beta_{TMAG} TMAG_{it-d} + \\ &\beta_C C_{it-e} + \beta_{TC} TC_{it-e} + \beta_{TMC} TMC_{it-e} + \beta_{TMAC} TMAC_{it-e} + \\ &\beta_{PC} PC_{it-e} + \beta_{NDI} NDI_{it-f} + \beta_{NBI} NBI_{it-f} + \beta_{NCI} NCI_{it-f} + \\ &\beta_{DUniv} DUniv_{it-g} + \beta_{DYear} DYear_{it-h} + \alpha + \varepsilon_{it} \end{aligned}}$$

Équation 2.1 Modèle général

Le modèle général peut être divisé en 5 sections, soit :

- Mesures de subventions
- Mesures de contrats
- Mesures de réseau
- Universités
- Années

Voici la description des termes présents dans le modèle :

- $\mathbf{A}_{it}$ représente le nombre d'articles du scientifique *i* à l'année *t*;
- β_G $\boldsymbol{G}_{it-d}$ est le nombre de *subventions* obtenues par le scientifique *i* au cours de la période *t-d*, β_G est son coefficient ;
- $\beta_{TG}\boldsymbol{TG}_{it-d}$ est le montant total de *subventions* en dollars constants obtenues par le scientifique *i* au cours de la période *t-d*, β_{TG} est son coefficient ;
- $\beta_{TMG}\boldsymbol{TMG}_{it-d}$ est le montant total moyen de *subventions* en dollars constants obtenues par le scientifique *i* au cours de la période *t-d*, β_{TMG} est son coefficient ;
- $\beta_{TMAG}\boldsymbol{TMAG}_{it-d}$ est le montant total moyen de *subventions* annuelles en dollars constants obtenues par le scientifique *i* au cours de la période *t-d*, β_{TMAG} est son coefficient ;
- β_C $\boldsymbol{C}_{it-e}$ est le nombre de *contrats* obtenus par le scientifique *i* au cours de la période *t-e*, β_C est son coefficient ;
- β_{TC} $\boldsymbol{TC}_{it-e}$ est le montant total de *contrats* en dollars constants obtenus par le scientifique *i* au cours de la période *t-e*, β_{TC} est son coefficient ;
- β_{TMC} $\boldsymbol{TMC}_{it-e}$ est le montant total moyen de *contrats* en dollars constants obtenus par le scientifique *i* au cours de la période *t-e*, β_{TMC} est son coefficient ;
- β_{TMAC} $\boldsymbol{TMAC}_{it-e}$ est le montant total moyen de *contrats* annuels en dollars constants obtenus par le scientifique *i* au cours de la période *t-e*, β_{TMAC} est son coefficient ;
- β_{PC} $\boldsymbol{PC}_{it-e}$ est la proportion de *contrats* par rapport aux *subventions* obtenus par le scientifique *i* au cours de la période *t-e*, β_{PC} est son coefficient ;
- β_{NDI} $\boldsymbol{NDI}_{it-f}$ est la la mesures du degré centralité du scientifique *i* au cours de la période *t-f*, β_{NDI} est son coefficient ;
- β_{NBI} $\boldsymbol{NBI}_{it-f}$ est la mesure de centralité de relation d'intermédiaire « betweenness centrality » du scientifique *i* au cours de la période *t-f*, β_{NBI} est son coefficient ;
- β_{NCI} $\boldsymbol{NCI}_{it-f}$ est la mesure de centralité de regroupement local « cliquishness » du scientifique *i* au cours de la période *t-f*, β_{NCI} est son coefficient ;

- β_{DUniv} $\boldsymbol{DUniv}_{it-g}$ est l'université où le scientifique *i* est affilié à l'année *t-g*, β_{DUniv} est son coefficient ;
- β_{DYear} $\boldsymbol{DYear}_{it-h}$ est l'année *t-h* par laquelle on veut contrôler le scientifique *i*, β_{DYear} est son coefficient ;
- α représente une constante;
- ε_{it} est un terme d'erreur.

Pour cette étude, la variable dépendante, qui est le nombre de publications scientifiques annuelles pour un scientifique, est une variable de dénombrement (« count variable ») et peut uniquement prendre des valeurs entières et non négatives. Le choix de cette variable dépendante s'explique par son utilisation comme mesure de production scientifique. On note aussi l'importance des publications scientifiques et de la génération de connaissance pour le domaine de la biotechnologie. L'accessibilité (non pas la facilité d'utilisation ou de traitement) des données pour l'ensemble des publications scientifiques canadiennes est aussi un facteur de choix. Le modèle tente d'expliquer le nombre d'articles générés à l'aide de plusieurs variables indépendantes qui peuvent avoir des durées variables. Un nombre précis d'articles publiés pendant une année spécifique est le résultat de l'historique des variables impliquées. Cette notion d'historique discutée par (Newman, 2001) devient important pour permettre de prévoir le nombre d'articles publiés.

Le critère de présence d'un chercheur au niveau du modèle est qu'il soit affilié à une université québécoise et ait reçu du financement.

Le choix des mesures de réseau provient de la nécessité de quantifier les liens existants entre les chercheurs pour modéliser les interactions au niveau des flux d'informations et de financement.

L'étape d'analyse consiste à modéliser les variables du modèle de façon à valider la pertinence de l'ensemble des paramètres du modèle. Une approche fréquemment utilisée dans le cas des données de dénombrement (« count data ») est la régression de Poisson (Hausman, Hall & Griliches, 1984). Par contre, les distributions de données de type Poisson se basent sur l'hypothèse que la moyenne de la distribution est égale à la variance, ce qui n'est pas le cas de nos données, on peut donc parler d'*hyperdispersion*. Ce concept quantifie la distribution des

données au niveau de la variabilité de celles-ci et il implique la présence d'une plus grande variabilité des données que celle qui devrait être attendue et expliquée par un modèle statistique simple.

2.3 Les données

Cette section décrit les données nécessaires à la validation des hypothèses à l'aide du modèle présenté. L'approche générale adoptée face aux données était de démarrer avec des ensembles de grande taille, avec un maximum de données, et puis de nettoyer de façon incrémentale les sous-ensembles nécessaires aux analyses.

Les données utilisées dans le cadre de ce projet sont liées autant aux articles, aux chercheurs, qu'au financement de la recherche universitaire. Ces données proviennent de différentes sources et organismes. Le résultat de cette situation est donc un ensemble de *différences* autant au niveau de la technologie de stockage des données, de la structure relationnelle entre les tables d'une même base de données, des standards de noms des champs que de l'intégrité des données eux-mêmes.

Les deux prochaines sections décrivent les sources de données qui nous ont permis de construire les réseaux et de calculer les différentes mesures nécessaires à l'analyse. Le choix de l'année 1985 comme année initiale provient d'un choix stratégique basé sur le nombre de données de publications et de financement disponibles des différentes sources de données.

2.3.1.1 Scopus

Les données détaillant les publications et leurs auteurs proviennent de la base de données d'articles scientifiques Scopus d'Elsevier qui est reconnue comme étant une des plus complètes pour les articles scientifiques dans un grand nombre de domaines de recherche. La section 2.4.1.1 détaillera le processus d'extraction et de formatage des données. Dès le départ, il existe deux problèmes liés aux données de Scopus. Premièrement, plusieurs articles scientifiques ne sont pas présents dans la base de données extraite de Scopus à cause du choix initial des mots-clés. L'inexistence d'une standardisation des mots-clés liés à la biotechnologie en plus de la difficulté de sélectionner un ensemble des mots-clés complet, font en sorte que peu importe la précision de

la requête de recherche, certains articles scientifiques nous échapperons. C’est une limite connue. Deuxièmement, l’utilisation de la base de données Scopus fait qu'il y aura possiblement une augmentation soudaine à l’année 1996 du nombre d’articles scientifiques causée par un changement interne chez Elsevier. Ce changement n’est pas limité au domaine de la biotechnologie mais a comme impact l’augmentation du nombre d’articles scientifiques. On note de plus qu’avant l’année 1996, la base de données Scopus n'était pas réputée pour être aussi complète.

2.3.1.2 SIRU

Le « *Système d'information sur la recherche universitaire* » (SIRU) permet de dresser annuellement un inventaire des subventions et des contrats de recherche accordés aux professeurs-chercheurs des universités du Québec. Les universités et leurs établissements affiliés ont l’obligation de transmettre, sur une base annuelle, toutes les données concernant les projets de recherche. L’exactitude et l’intégralité des données sont assurées par les universités et la mise à jour des données est la responsabilité du ministère de l’Éducation, du Loisir et du Sport (MELS). Le détail des montants du financement par chercheur est ventilé annuellement. La section 2.4.1.2 détaillera certaines des manipulations effectuées sur les données. Un problème initialement connu avec les données est le manque de standardisation au niveau de la catégorisation annuelle du financement reçu par les chercheurs.

2.4 Méthodologie de recherche

Pour tester nos hypothèses à travers le modèle proposé à la section 2.2 un ensemble de données de panel (« panel data ») sur la période 1985-2005 sera colligé. Nous analyserons ensuite ce modèle à l'aide du logiciel statistique Stata. Cette section décrira le cheminement pour passer des données initiales aux statistiques finales.

2.4.1 Capture, nettoyage et formatage des données

Voyons maintenant en détails la capture et les transformations effectuées sur les sources de données détaillées précédemment aux sections 2.3.1.1et 2.3.1.2.

2.4.1.1 Scopus

Le domaine de la biotechnologie est marqué par un manque de consensus pour en définir les frontières. Cette situation rend l'étude bibliométrique plus difficile, car non seulement les experts ne s'entendent pas sur une définition universelle mais les mots-clés évoluent avec le temps. L'inexistence d'une définition commune et reconnue pour la biotechnologie fait en sorte que la collecte des données entre plusieurs organismes et bases de données est encore rudimentaire. Ce qui est catégorisé comme lié à la biotechnologie peut varier selon la source des données. La qualité des données extraites est liée à la fiabilité des systèmes de catégorisations et aux mots-clés utilisés par les différentes sources utilisées. La nature multidisciplinaire de ce domaine apporte un défi additionnel en nécessitant l'utilisation d'une méthode incrémentale pour la sélection des mots clés à utiliser lors de la recherche de l'ensemble des articles scientifiques pour éviter les doublons. Plusieurs approches et méthodes ont été étudiées dont celle de Mogoutov & Kahane (2007) et la méthode retenue pour ce mémoire se base sur celle de Porter, A. L., Youtie, Shapira & Schoeneck (2008). Celle-ci se divise en trois étapes :

- Premièrement, le recensement de mots-clés déjà utilisés dans des études antérieures.
- Deuxièmement, la revue des mots-clés retenus par des experts du domaine de façon à améliorer la sélection.
- Troisièmement, l'ajout incrémental des nouveaux mots clés qui apparaissent à travers le temps, de façon à suivre l'évolution du domaine.

Voyons plus en détails dans le Tableau 2.1 les mots-clés inclus dans la requête utilisée pour extraire les détails des articles scientifiques liés à la biotechnologie.

Tableau 2.1: Mots-clés utilisés pour la recherche des articles scientifiques en biotechnologie

Mots-clés utilisés pour la recherche des articles scientifiques en biotechnologie
gene OR antigen OR antibody OR genom* OR hybridoma OR dna OR rna OR proteom* OR protein OR peptide OR nucleotide OR embryogenesis OR enzyme OR clon* OR monoclon* OR polyclon* OR biomarker
gene OR antigen OR antibody OR genom* OR hybridoma OR dna OR rna OR proteom* OR protein OR peptide OR nucleotide OR enzyme OR clon* OR monoclon* OR polyclon* OR probiotic* OR biofertiliz* OR bioherbicid* OR bioinsecticid* OR biocontrol OR biomarker
gene OR antigen OR antibody OR genom* OR hybridoma OR dna OR rna OR proteom* OR protein OR peptide OR nucleotide OR enzyme OR clon* OR monoclon* OR polyclon* OR biofuel OR bioremediat* OR biodegrad* OR bioleach* OR bioseparat* OR biosorpt* OR biocatal* OR biofilter* OR bioreact* OR bioprocess* OR biomarker
gene OR antigen OR antibody OR genom* OR hybridoma OR dna OR rna OR proteom* OR protein OR peptide OR nucleotide OR embryogenesis OR enzyme OR clon* OR monoclon* OR polyclon* OR biomarker OR probiotic* OR biofertiliz* OR bioherbicid* OR bioinsecticid* OR biocontrol OR biofuel OR bioremediat* OR biodegrad* OR bioleach* OR bioseparat* OR biosorpt* OR biocatal* OR biofilter* OR bioreact* OR bioprocess*

Une fois qu'une base de mots-clés a été définie, les informations sur les chercheurs et les articles sont récupérées à l'aide du logiciel PatentBot. PatentBot est un logiciel d'automatisation de capture de données de Scopus développé par Ahmad Barirani. Le critère de sélection des articles est qu'au moins un des auteurs doit avoir une affiliation canadienne. Les champs récupérés détaillaient d'une part les articles et de l'autre, les scientifiques. Parmi les données liées aux chercheurs une attention particulière a été portée aux données permettant la localisation de ces chercheurs. Les adresses des chercheurs, qui comportent le nom de la ville et le code postal, sont des données essentielles pour la construction des grappes géographiques comme (Aharonson et

al., 2008) le mentionnent avec la division des grappes par codes postaux. La Figure 2.1 illustre l'outil d'extraction PatentBot.

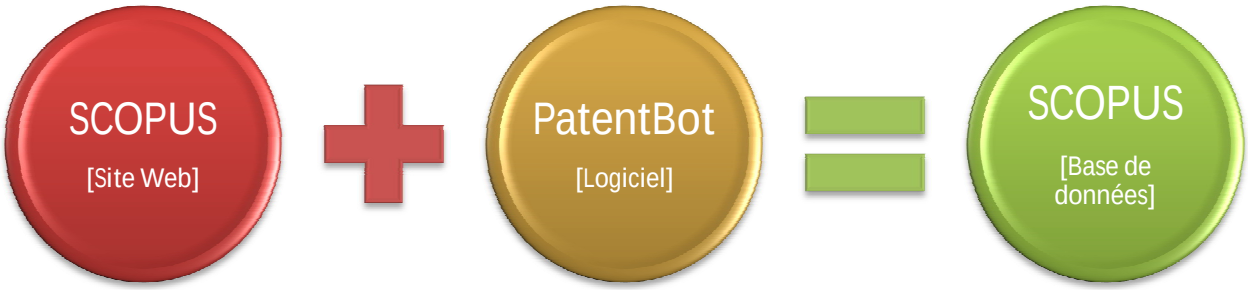


Figure 2.1 Automatisation de l'extraction des données Scopus

Une fois les données extraites du site Scopus, un effort majeur de normalisation et de nettoyage des bases de données a été entrepris pour permettre une première analyse des données. L'objectif était d'assurer une compatibilité entre les requêtes intra-base de données et éviter des problèmes d'intégrité entre celles-ci. La base de données Scopus telle qu'extraite par le logiciel PatentBot et modifiée à la suite d'une première phase de nettoyage et est divisée en deux tables illustrées à la Figure 2.2 et à la Figure 2.3.

ARTICLES

Field Name	Data Type
ID	AutoNumber
Name of the Article	Text
Name of the article (Original)	Text
Journal name	Text
Other info (volume, issue, pages)	Text
Note	Text
Year	Number
First author's surname	Text
Abstract	Memo
Keywords	Text
# of References	Number
Flag	Text

Figure 2.2 Structure de la table ARTICLES de la base de données Scopus

SCIENTISTS

Field Name	Data Type
ARTICLE_ID	Number
ind_numero	Number
Aut_last_name	Text
Aut_Original last_name	Text
Aut_First_name	Text
Aut_First_name (WITH DOT)	Text
Aut_Original first_name	Text
Aut_First_N_FULL	Text
Affiliation	Text
Affiliation_Comment	Text
Country	Text
Organisation_Type	Text
Cluster	Text
Province	Text
Nom_Prenom	Text
Bourse	Text
Note	Text
ID	AutoNumber
Aut_postal_code	Text
Aut_postal_code_2e problable	Text
codep_AJOUT	Text
postal_code_plus_fréquent	Text
Label	Text

Figure 2.3 Structure de la table SCIENTISTS de la base de données Scopus

Bien que la méthode utilisée soit jugée valide pour cette recherche, il faut quand même mentionner ses limitations. Les mots-clés liés à la biotechnologie évoluent à travers le temps et cette situation peut mener à exclure involontairement des articles. Les mots-clés qualifiant le domaine d'aujourd'hui n'étaient certainement pas les mêmes qu'il y a quelques décennies. Il faut donc toujours enrichir la banque de mots-clés avec ceux qui s'ajouteront dans le futur, mais aussi revoir l'historique des publications passées de façon à ajouter celles qui avaient été exclues des recherches.

Pour notre étude, la capacité de regrouper les auteurs par grappe était importante. Nous avons donc créé des regroupements géographiques logiques selon deux catégories, soit liée à une ville

de la province de Québec ou soit liée à une région géographique hors Québec. Le Tableau 2.2 et le Tableau 2.3 illustrent les deux catégories de grappes utilisées pour l'analyse.

Tableau 2.2 Grappes liées au Québec

Grappes liées au Québec
Gatineau
Montreal
Sherbrooke
Chicoutimi
Quebec
Rimouski
Saguenay/ Lac St-Jean

Tableau 2.3 Grappes hors Québec

Grappes hors Québec
Ontario
Prairies
Atlantic
British Columbia
United States
United States - Northwest
United States - Southwest
United States - Midwest
United States - Northeast
United States - South
South America
Europe
Africa
Asia
Australia

Tel qu'illustrée à la Figure 2.3, la base de données Scopus contient dans sa table SCIENTISTS l'information au niveau de la grappe (le champ « cluster ») qui correspond en fait à la ville de l'affiliation de l'auteur (note: si pour un article plusieurs affiliations sont présentes pour le chercheur, plusieurs entrées seront présentes dans la table SCIENTISTS et donc plusieurs grappes seront associées à un chercheur pour un article). Dans la matrice finale, pour chacun des

auteurs ayant publié pour l'année en cours, la valeur utilisée pour la grappe est celle dont la fréquence est la plus élevée. Par exemple, un auteur ayant publié 3 articles et qui possèderait 3 affiliations pour chacun de ses articles (Québec [Université Laval], Montréal [UQAM] et Montréal [Polytechnique]) se verrait attribuer Montréal comme grappe principale.

De façon à rendre son utilisation dynamique, un utilitaire nommé « **Generate_Cluster_Mapping** » a été créé pour générer automatiquement une table de mappage entre les grappes définies au niveau de Scopus et celles qui nous intéressent pour cette étude définies au Tableau 2.2 et au Tableau 2.3. La Figure 2.4 illustre le processus de l'utilisation de l'outil tandis que la Figure 2.5 montre son intégration dans la matrice finale d'analyse.

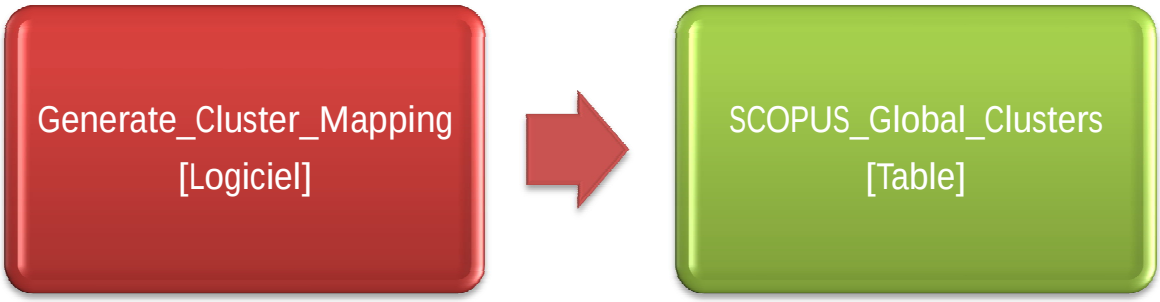


Figure 2.4 Génération de la table de mappage des grappes

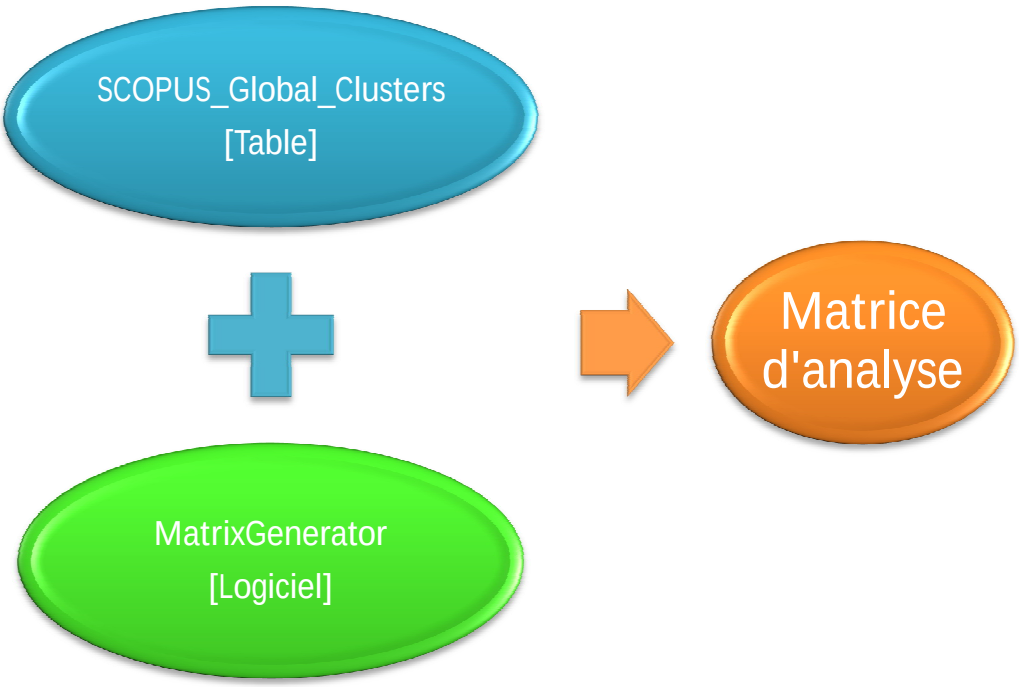


Figure 2.5 Utilisation de la table de mappage des grappes pour la génération de la matrice

2.4.1.2 SIRU

Un des choix méthodologique important pour notre étude a été posé pour permettre un raffinement additionnel des données:

Un modèle utilisant des variables ventilées par type de financement permet d'obtenir un modèle plus précis qui décrit mieux l'effet des variables individuelles liées au financement sur la production d'articles scientifiques.

Nous verrons à travers les prochaines sections l'importance de ce choix méthodologique sur l'analyse de nos données.

Les données reçues de l'organisme SIRU sont divisées sommairement selon le type de financement. Les experts consultés nous recommandent cette approche de catégorisation modifiée pour tenir compte de certains types et sous-types qui ne seraient pas standardisés à travers les années. Les données de financement provenant de l'organisme SIRU ont donc été divisées en plusieurs catégories présentées au Tableau 2.4.

Tableau 2.4: Catégories du financement

Catégories du financement
contrat-appareillage
contrat-fonctionnement
contrat-infrastructure
subvention-appareillage
subvention-fonctionnement
subvention-infrastructure

Plusieurs types de financement liés à la rémunération des chercheurs ou à l'aspect de la commercialisation étaient présents dans les données reçues du SIRU. Comme ce n'est pas l'ensemble des types de financement qui s'appliquent au financement des projets de recherche, certains ont été retirés. Le tableau Tableau 2.5 contient donc les types de financement retirés.

Tableau 2.5: Types de financement retirés de la table de correspondance

Type financement	*Description du type financement*
Contrat	Bourses salariales nominatives aux chercheurs principaux
Contrat	Rémunération des chercheurs principaux
Contrat	Valorisation et commercialisation
Fonds internes de recherche	Bourses salariales nominatives aux chercheurs principaux
Fonds internes de recherche	Rémunération des chercheurs principaux
Subvention	Bourses salariales nominatives aux chercheurs principaux
Subvention	Rémunération des chercheurs principaux
Subvention	Subventions sous forme de bourses salariales nominatives
Subvention de partenariat	Bourses salariales nominatives aux chercheurs principaux
Subvention de partenariat	Rémunération des chercheurs principaux
Subventions	08-Bourses salariales

Voyons maintenant comment ces sources de données ont été intégrées dans une matrice d'analyse.

2.4.2 Matrice d'analyse

Plusieurs étapes ont été nécessaires pour mettre en commun et manipuler les données dans le but de les analyser. Voici un aperçu des opérations qui doivent être effectuées pour obtenir une matrice d'analyse finale :

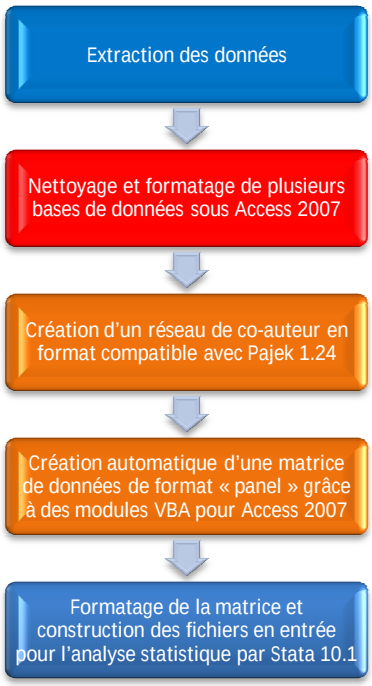


Figure 2.6 Aperçu des opérations

La Figure 2.7 est un aperçu des opérations tandis que la Figure 2.8 et la Figure 2.8 illustrent les liens entre les bases de données, les processus et les tables générées menant à la création de la matrice d'analyse finale. Chacune des figures représentent des étapes clés dans l'évolution du projet de recherche.

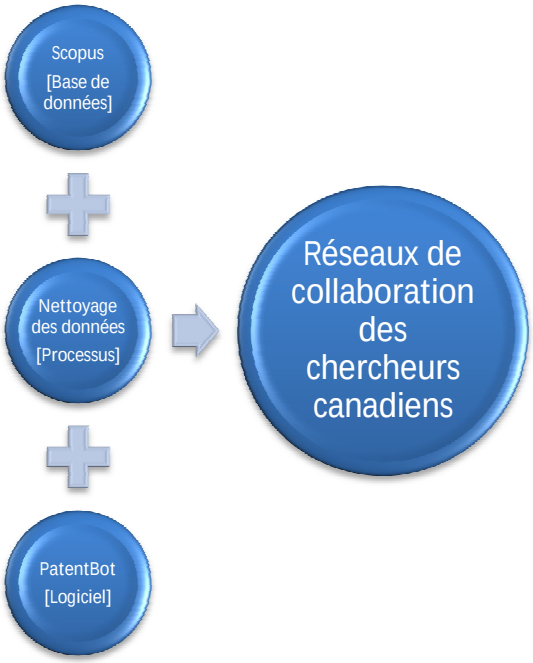


Figure 2.7: Capture et nettoyage de la base de données Scopus.

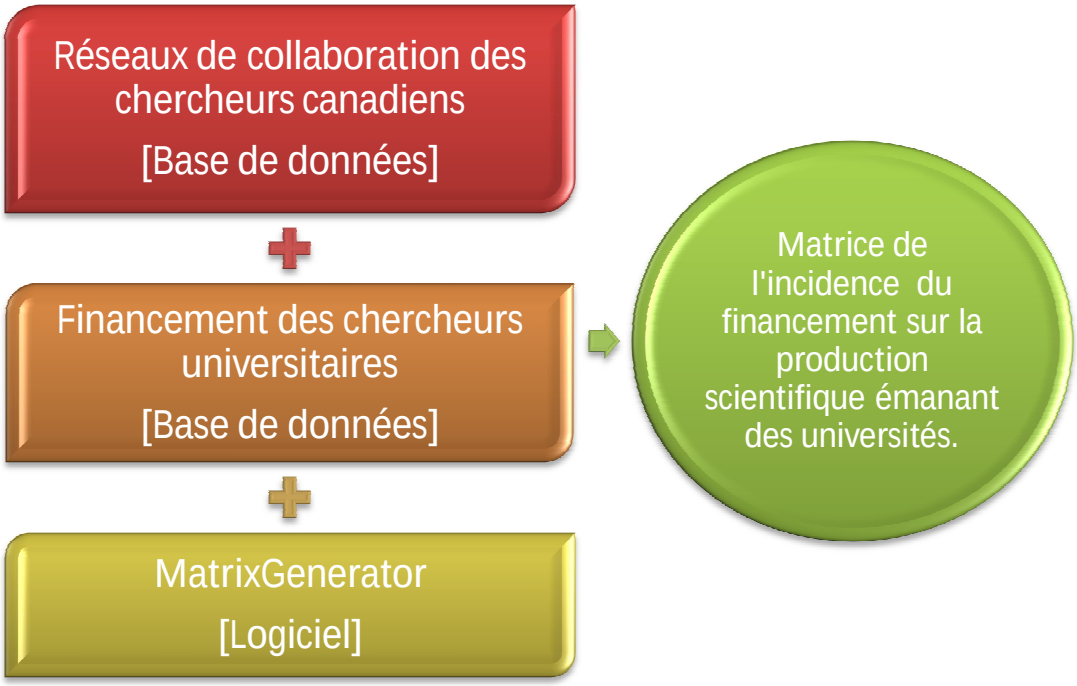


Figure 2.8: Création de la matrice d'analyse

L'objectif du projet de recherche n'est pas l'étude des sources possibles de financement ou les conditions et critères de financement des projets de recherche étudiés notamment par (Link & Scott, 2004) pour établir les cas où le financement public doit être utilisé mais bien leur utilisation et leur impact sur la *génération de l'innovation*. Nous utilisons des données existantes et analysons la distribution et l'utilisation du financement pour l'ensemble des projets de recherche pour le domaine de la biotechnologie.

Le principe général d'analyse des données suivi pour ce mémoire se décompose en quatre étapes :

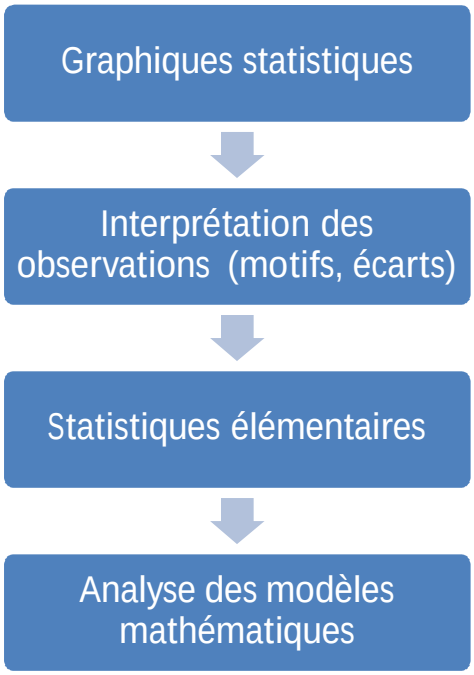


Figure 2.9: Étapes du principe général d'analyse des données

L'ensemble de ces étapes seront présentées dans les deux prochains chapitres :

- **Graphiques statistiques** : représentations graphiques des données pour donner un sens de l'état initial de l'ensemble des informations liées au projet.
- **Interprétation des observations** : une analyse sommaire des graphiques statistiques pour trouver les motifs et les écarts globaux présents dans les données.

- **Statistiques élémentaires** : utilisation de mesures statistiques précises pour documenter les observations.
- **Modèles mathématiques** : plusieurs variations du modèle général seront étudiées pour déterminer le modèle statistique correspondant le mieux aux données. Les résultats seront discutés.

2.4.3 Création des réseaux sociaux

Le logiciel Pajek permettra de calculer les mesures de réseau détaillées à la section 1.3.1. Nous posons maintenant les bases de notre analyse. Un réseau est un graphe contenant de l'information sur ses lignes ou arêtes. Pour notre étude, les réseaux construits sont constitués des arêtes (« edges ») représentant des articles unissant des sommets (« vertices »). Ces sommets représentent les co-auteurs des articles scientifiques. Les réseaux ainsi créés sur un nombre fixe d'années sont *non-dirigés*, ce qui signifie qu'ils ne contiennent pas d'arc (où la notion de direction est présente) et que toutes les lignes entre les sommets sont des arêtes (Wouter et al., 2004).

- Chacun des sommets ont un identifiant unique représentant les noms des chercheurs;
- Une arête représente toujours un article, dans le cas de plusieurs articles entre deux mêmes co-auteurs, plusieurs liens existent entre les deux arêtes.

Des hypothèses doivent être faites quant aux durées d'analyses des réseaux et c'est donc la durée des subventions (et les données liées à celles-ci) qui dicte l'intervalle minimum de temps requis à l'analyse de ces réseaux. Pour cette étude, une approche conservatrice est utilisée avec des périodes d'une, trois et cinq années de façon à couvrir un ensemble de type de financement périodique.

Une approche conservatrice est aussi utilisée au niveau des retards de phase (« lag ») en utilisant des retards d'une (L1) et deux (L2) années. L'utilisation d'un retard de phase de trois ans (L3) n'est pas jugée nécessaire pour cette analyse car un retard de phase de 3 ans semble très long pour un domaine aussi dynamique que la biotechnologie.

Une des caractéristiques des réseaux de co-auteurs est la présence d'un grand nombre de composantes de taille unitaire, soit des singletons. Ceux-ci représentant des articles n'étant pas le résultat d'une collaboration mais ayant un seul chercheur comme auteur. Lorsque ces composantes représentent une majorité du réseau, on peut donc parler de réseau très fragmenté («unconnected »).

À la suite d'une publication, le ou les chercheurs intègrent le réseau, car de notre point de vue la « naissance » du chercheur est lors de sa première publication alors que sa « mort » est associée à sa dernière publication.

Le logiciel Pajek 1.24 a été sélectionné pour le calcul de l'ensemble des mesures réseaux. Ce logiciel a besoin d'avoir en entrée un format spécifique de données. La Figure 2.10 représente le format requis par Pajek pour construire un réseau.

```
*Vertices 254543 153796
1 "Aakre M" [2005]
2 "Aalbaek B" [1998]
3 "Aalberse RC" [1986]
...
153793 "Zysman M" [2002]
153794 "Zysman MA" [2002]
153795 "Zytner RG" [2005]
153796 "Zywulko M" [1990-1992]
153797 1 [2006]
153798 2 [2006]
153799 3 [2006]
153800 4 [2006]
...
153796 178457 1 [1991]
153796 180095 1 [1990]
153796 176167 1 [1992]
153796 229460 1 [1990]
153796 197108 1 [1992]
```

Figure 2.10 Format du fichier d'entrée pour Pajek

Un logiciel a été créé en VB6 sous Microsoft Office Access 2007 pour automatiser les deux principales étapes liées au logiciel Pajek soit :

1. La génération des fichiers d'entrée de données;

2. Le calcul et de la récupération des mesures réseaux effectuées par le logiciel Pajek.

Au niveau de l'automatisation de la génération du fichier d'entrée de Pajek, la Figure 2.11 représente un aperçu des opérations effectuées. Les données utilisées pour cette étape proviennent de Scopus et ont été préalablement nettoyées et formatées pour être utilisées directement.

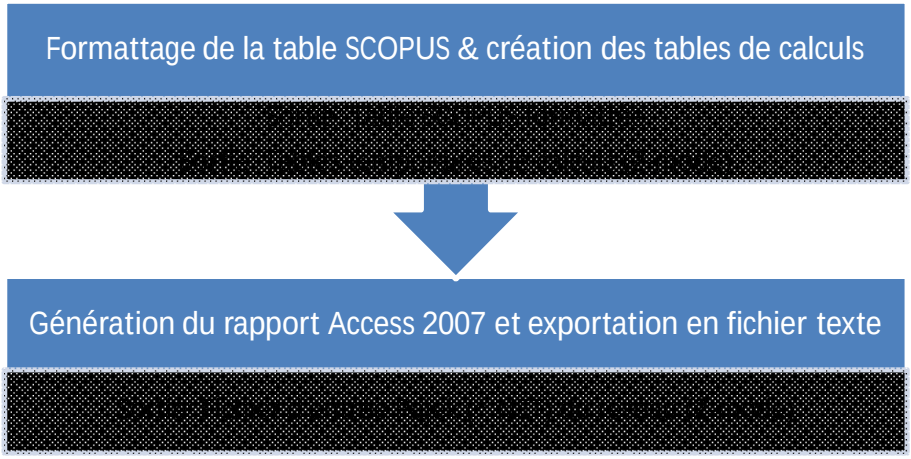


Figure 2.11 Mécanisme d'automatisation de la création du fichier d'entrée pour Pajek

La deuxième étape consiste à automatiser l'exécution de Pajek pour :

1. Charger un nombre élevé et variable de réseau sur 1, 3 et 5 ans, puis effectuer différentes commandes pour les calculs des mesures réseaux. Ces étapes sont possibles grâce à la génération dynamique d'un fichier d'automatisation (macro) de Pajek, soit le fichier Pajek.log documenté à l'Annexe 1 qui décrit séquentiellement l'ensemble des commandes à être exécutées mais aussi la localisation des fichiers en sortie.

2. Récupérer les fichiers en sortie, effectuer une analyse syntaxique des données pour récupérer uniquement les sections voulues et les insérer dans des tables Access 2007.

2.4.4 Génération automatique de la matrice d'analyse

Nous avons détaillé à la section 2.4.2 l'utilité de la matrice d'analyse, voici une vue de la logique d'intégration automatique des sources de données de façon à obtenir une matrice d'analyse.

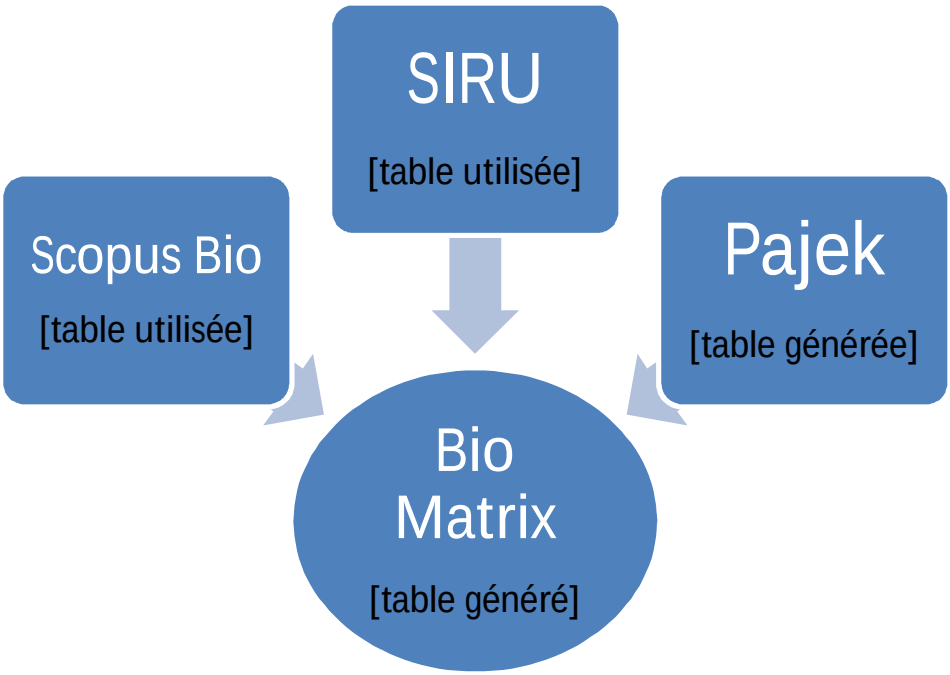


Figure 2.12: Vue globale des tables

La Figure 2.12 présente donc comment sont intégrées les différentes tables par un module VB6. L'outil est conçu pour être configurable et permettre l'ajout de nouvelles mesures suivant l'évolution de la recherche. La présence de chacune des mesures de la matrice finale est contrôlée par une méthode précise. La Figure 2.13 est un exemple d'une méthode qui contrôle l'ajout des mesures réseaux. On note qu'elle s'occupe de faire des appels aux méthodes spécialisées donc l'ajout ou le retrait de méthodes spécialisées est simplifié.

```
'    ******************************************************************************
'    Function updating only line matching Pajek data
'    ******************************************************************************
Public Function PopulateTable_with_Pajek(ByVal strTableName As String, ByVal intYearRangeStart As Integer, ByVal intYearRangeEnd As Integer,

    If bolDebug Then
        Debug.Print Now() & " [PopulateTable_with_Pajek] - START DEBUG"
    End If

    ' Variables declaration
    Dim intPosition As Integer
    Dim varArray As Variant
    Dim strRegexPattern As String

    ' Setting the regex pattern to be found
    strRegexPattern = "([\d]+)"
    ' Getting the values present in the first line
    Call DynamicRegexLineParsing(Trim(strYearStep), strRegexPattern, varArray, False)

    If bolDebug Then
        ' Displaying size of array
        Debug.Print Now() & " [PopulateTable_with_Pajek] - Size of array containing year step: " & UBound(varArray)
        ' Looping through all element of array
        For intPosition = 0 To UBound(varArray) - 1
            Debug.Print Now() & " [PopulateTable_with_Pajek] - Item # : " & intPosition & " with value: " & varArray(intPosition)
        Next intPosition
    End If

    ' Looping through all year step found and calling statistics functions
    For intPosition = 0 To UBound(varArray) - 1

        ' 1- Update table with CENTRALITY statistics
        Call PopulateTable_with_Pajek_centrality(strTableName, intYearRangeStart, intYearRangeEnd, varArray(intPosition), False)

        ' 2- Update table with BETWEENNESS statistics
        Call PopulateTable_with_Pajek_betweenness(strTableName, intYearRangeStart, intYearRangeEnd, varArray(intPosition), False)

        ' 3- Update table with CLIQUISHNESS statistics
        Call PopulateTable_with_Pajek_cliquishness(strTableName, intYearRangeStart, intYearRangeEnd, varArray(intPosition), False)

        ' 4- Update table with COMPONENT statistics
        Call PopulateTable_with_Pajek_component(strTableName, intYearRangeStart, intYearRangeEnd, varArray(intPosition), False)

        ' 5- Update table with NETWORKS statistics
        Call PopulateTable_with_Pajek_networkinfo(strTableName, intYearRangeStart, intYearRangeEnd, varArray(intPosition), True)

    Next intPosition

    If bolDebug Then
        Debug.Print Now() & " [PopulateTable_with_Pajek] - END DEBUG"
    End If

End Function
```

Figure 2.13 Méthode contrôlant l'ajout de mesures réseaux dans la matrice finale

La structure de la matrice d'analyse générée est présentée à la Figure 2.14, la Figure 2.15 et la Figure 2.16.

BIO_Matrix

Field Name	Data Type
ID	AutoNumber
ID_Scientist	Number
ID_SCOPUS	Number
ID_SIRU	Number
Year	Number
Last_Name_Full	Text
First_Name_Initials	Text
Label	Text
SIRU_Matricule	Number
SIRU_Regroupement	Text
Postal_Code	Text
Postal_Code_Count_Yearly	Number
Cluster	Text
Cluster_Count_Yearly	Number
Cluster_Primary_Yearly	Text
Cluster_Global	Text
Article_Count_Yearly	Number

Figure 2.14 Matrice d'analyse: identificateurs, grappes et articles

BIO_Matrix

Field Name	Data Type
Contract_Count_Yearly_appareillage	Number
Contract_Total_Value_appareillage	Number
Contract_Count_Yearly_fonctionnement	Number
Contract_Total_Value_fonctionnement	Number
Contract_Count_Yearly_infrastructure	Number
Contract_Total_Value_infrastructure	Number
Grant_Count_Yearly_appareillage	Number
Grant_Total_Value_appareillage	Number
Grant_Count_Yearly_fonctionnement	Number
Grant_Total_Value_fonctionnement	Number
Grant_Count_Yearly_infrastructure	Number
Grant_Total_Value_infrastructure	Number
University_Yearly	Text
University_Count_Yearly	Number
University_Primary_Yearly	Text
1_Year_Unit_Centrality	Number
1_Year_Normalized_Unit_Centrality	Number
1_Year_Network_Centralization	Number
3_Year_Unit_Centrality	Number
3_Year_Normalized_Unit_Centrality	Number
3_Year_Network_Centralization	Number
5_Year_Unit_Centrality	Number
5_Year_Normalized_Unit_Centrality	Number
5_Year_Network_Centralization	Number

Figure 2.15 Matrice d'analyse: contrats, subventions, universités et réseaux

BIO_Matrix

Field Name	Data Type
1_Year_Betweenness_Centrality	Number
1_Year_Betweenness_Centralization	Number
3_Year_Betweenness_Centrality	Number
3_Year_Betweenness_Centralization	Number
5_Year_Betweenness_Centrality	Number
5_Year_Betweenness_Centralization	Number
1_Year_Number_of_Lines_Among_Vertices_in_1-Neighbourhood	Number
1_Year_Clustering_Coefficient_only_1-Neighbourhood	Number
1_Year_Clustering_Coefficient_only_1-Neighbourhood_prime	Number
3_Year_Number_of_Lines_Among_Vertices_in_1-Neighbourhood	Number
3_Year_Clustering_Coefficient_only_1-Neighbourhood	Number
3_Year_Clustering_Coefficient_only_1-Neighbourhood_prime	Number
5_Year_Number_of_Lines_Among_Vertices_in_1-Neighbourhood	Number
5_Year_Clustering_Coefficient_only_1-Neighbourhood	Number
5_Year_Clustering_Coefficient_only_1-Neighbourhood_prime	Number
1_Year_ID_Weak_Component_Min_Size_1	Number
3_Year_ID_Weak_Component_Min_Size_1	Number
5_Year_ID_Weak_Component_Min_Size_1	Number
1_Year_ID_Strong_Component_Min_Size_1	Number
3_Year_ID_Strong_Component_Min_Size_1	Number
5_Year_ID_Strong_Component_Min_Size_1	Number
1_Year_ID_Strong_Periodic_Component_Min_Size_1	Number
3_Year_ID_Strong_Periodic_Component_Min_Size_1	Number
5_Year_ID_Strong_Periodic_Component_Min_Size_1	Number
1_Year_Diameter	Number
3_Year_Diameter	Number
5_Year_Diameter	Number

Figure 2.16 Matrice d'analyse: mesures réseaux

La dernière étape du processus est donc d'exporter la matrice d'analyse pour qu'elle soit utilisable par le logiciel Stata.

2.4.5 Construction des fichiers d'analyse Stata

La figure suivante résume les transformations qui doivent être appliquées à la matrice générée pour pouvoir en étudier ses propriétés à l'aide du logiciel statistique Stata.

Figure 2.17: Transformations de la matrice pour l'obtention du fichier de données Stata

La matrice contient un ensemble de variables dont les relations doivent être analysées. Pour étudier efficacement ces relations un ensemble d'opérations doivent être effectuées, en voici le résumé :

- La première opération à effectuer est de refaire le mappage des variables de la matrice pour les standardiser dans Stata.
- La seconde opération est l'actualisation des données de financement à l'aide de l'indice des prix à la consommation (IPC). Nous ramènerons les montants des contrats et des subventions en dollars constants de 2002.

Nous concluons ainsi la section décrivant les différentes opérations effectuées pour obtenir des données analysables à l'aide du logiciel statistique Stata.

CHAPITRE 3 CARACTÉRISTIQUES ET STATISTIQUES DES DONNÉES

Ce chapitre présente un ensemble de graphiques et de statistiques élémentaires des données utilisées dans le cadre de ce projet de recherche. L'objectif est de donner au lecteur une perspective, un état global des données, avant de passer à une analyse plus formelle à l'aide d'analyses de régressions et de corrélations présentées au CHAPITRE 4.

Le chapitre est divisé en 4 sections. Premièrement, des statistiques sur le financement public seront présentées. Deuxièmement, des statistiques sur les publications scientifiques tirées de Scopus seront illustrées à l'aide de graphes. Troisièmement, une description sommaire des données tirées de SIRU permettra de situer les récipiendaires du financement et de ventiler le financement en catégories tel que présentées à la section 2.3.1.2. Dernièrement, des statistiques élémentaires tirées du réseau, généré à l'aide des données de Scopus, nous permettront de mieux situer les réseaux de co-auteurs étudiés.

3.1 Financement public

Pour cette étude, les données liées au financement proviennent de la base de données du SIRU qui permet de dresser annuellement un inventaire des subventions et contrats de recherche accordés aux professeurs-chercheurs des universités du Québec. Il est donc intéressant de comparer ces montants avec ceux calculés au niveau provincial pour en évaluer les tendances.

La Figure 3.1 dresse un bref aperçu de l'importance des montants dépensés dans les activités scientifiques de recherche et développement au Québec, tous domaines confondus.

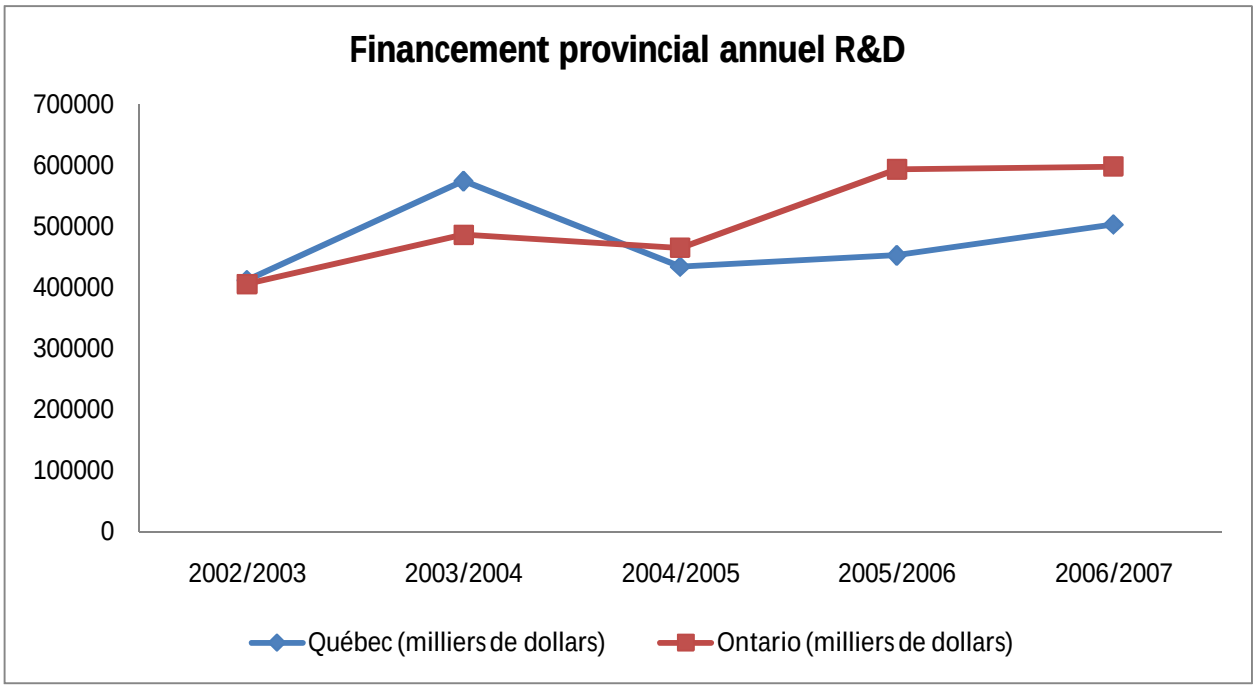


Figure 3.1 Financement provincial annuel R&D (*Science Statistics*, 2008)

La Figure 3.1 présente les montants alloués en dollars constants (2002) pour l'*ensemble* de l'activité scientifique de recherche et développement pour les deux provinces recevant le plus de financement. Au niveau québécois, on observe une pointe pour 2003/2004 (139%) puis une reprise par la suite pour 2004/2005 (76%), puis une croissance constante en 2005/2006 (104%) et en 2006/2007 (111%). La province de l'Ontario est présentée à titre comparatif mais on note que l'Ontario devance le Québec pour les sommes totales investies. Pour les deux provinces, bien qu'on présente des données globales, on observe un certain ralentissement des investissements. Il sera intéressant de comparer cette statistique avec les données spécifiques des chercheurs en biotechnologie et évidemment l'impact sur la génération de publications scientifiques. Pour démontrer l'importance des montants en dollars constants (2002) dépensés en sciences naturelles et en l'ingénierie, donc la biotechnologie fait partie, la Figure 3.2 présente des données spécifiques aux activités de sciences naturelles et d'ingénierie pour le Québec.

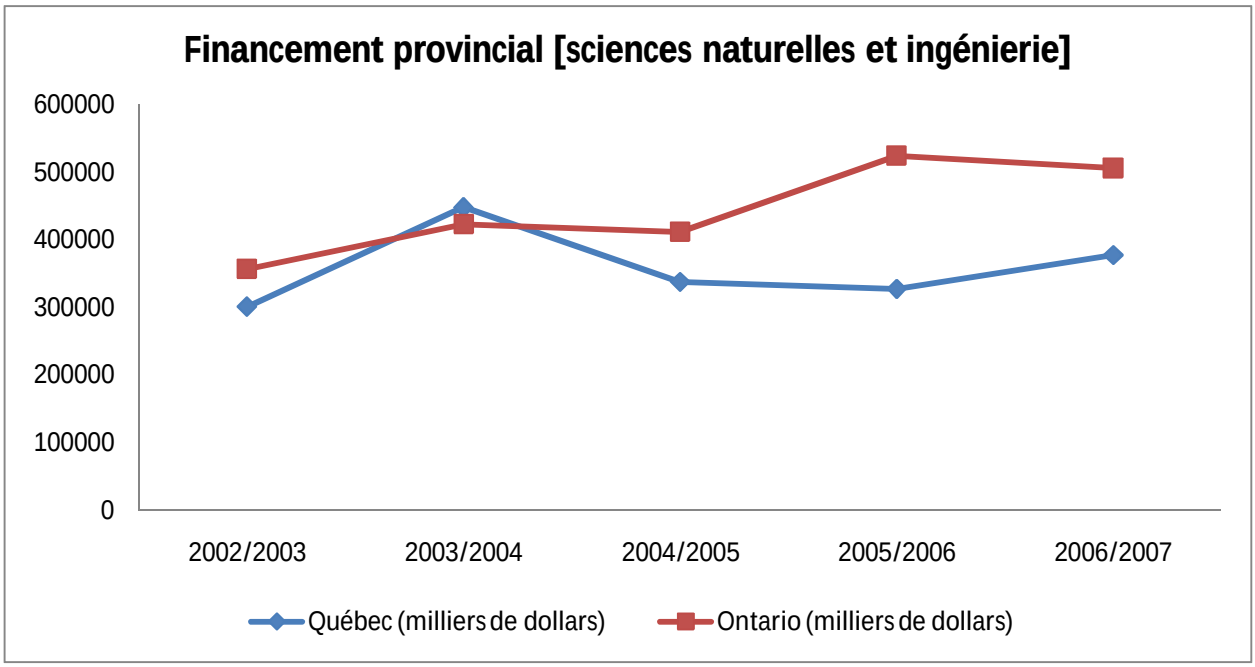


Figure 3.2 Financement provincial [sciences naturelles et ingénierie] (*Science Statistics*, 2008)

Au niveau des données québécoises, on retrouve encore une pointe en 2003/2004 (149%) puis une reprise amorcée en 2004/2005 (75%), 2005/2006 (97%) et finalement une croissance en 2006/2007 (115%). Les données pour la province de l'Ontario sont toujours présentées à titre comparatif mais on note au passage l'écart existant entre le financement des activités ontariennes et les activités québécoises. Il sera intéressant de voir si la décroissance des investissements entre 2003/2004 et 2005/2006 aura un impact sur la publication des articles scientifiques dans le domaine de la biotechnologie.

Nous avons vu la courbe globale des investissements au niveau de la recherche au Québec, la section suivante présentera les données liées au domaine de la biotechnologie dans le secteur universitaire québécois.

3.2 Articles scientifiques et chercheurs

Cette section traite des publications scientifiques et de leurs auteurs obtenus à partir de la base de données Scopus. Le critère d'inclusion des articles scientifiques étaient qu'il y ait au moins un auteur ayant une affiliation au Canada. À l'aide de cette base de données, nous avons extrait **82061** articles scientifiques rédigés par **3678** chercheurs. La méthodologie utilisée pour l'extraction de ces données a été présentée en détails à la section 2.4. Les sections suivantes

traitent d'abord des articles scientifiques puis des chercheurs. Les aspects géographiques des publications seront aussi traités dans la section suivante.

3.2.1 Articles scientifiques

Pour bien situer le lecteur par rapport aux données extraites, nous débutons cette section en présentant un aperçu global avec la Figure 3.3 représentant l'ensemble des articles scientifiques recensés annuellement dont au moins un des auteurs possède une affiliation canadienne. Bien que notre étude porte sur la période 1985-2005, le graphe complet est présenté pour illustrer la croissance du nombre de publications scientifiques depuis les six dernières décennies. La somme annuelle des articles est basée sur des identificateurs uniques de publications.

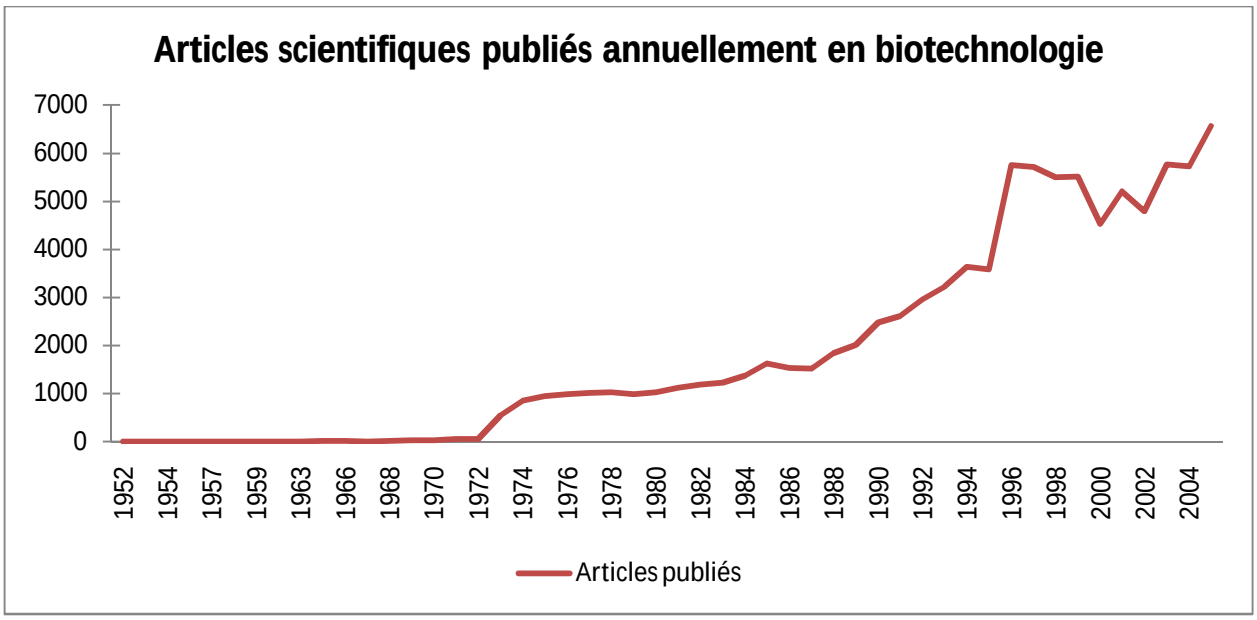


Figure 3.3: Articles scientifiques publiés annuellement en biotechnologie

Globalement la Figure 3.3 présente une croissance de la publication d'articles scientifiques depuis les années soixante-dix avec une pointe de croissance pour la période 1994-2000. La Figure 3.4 présente l'évolution des publications de façon plus précise pour la période étudiée, soit 1985-2005.

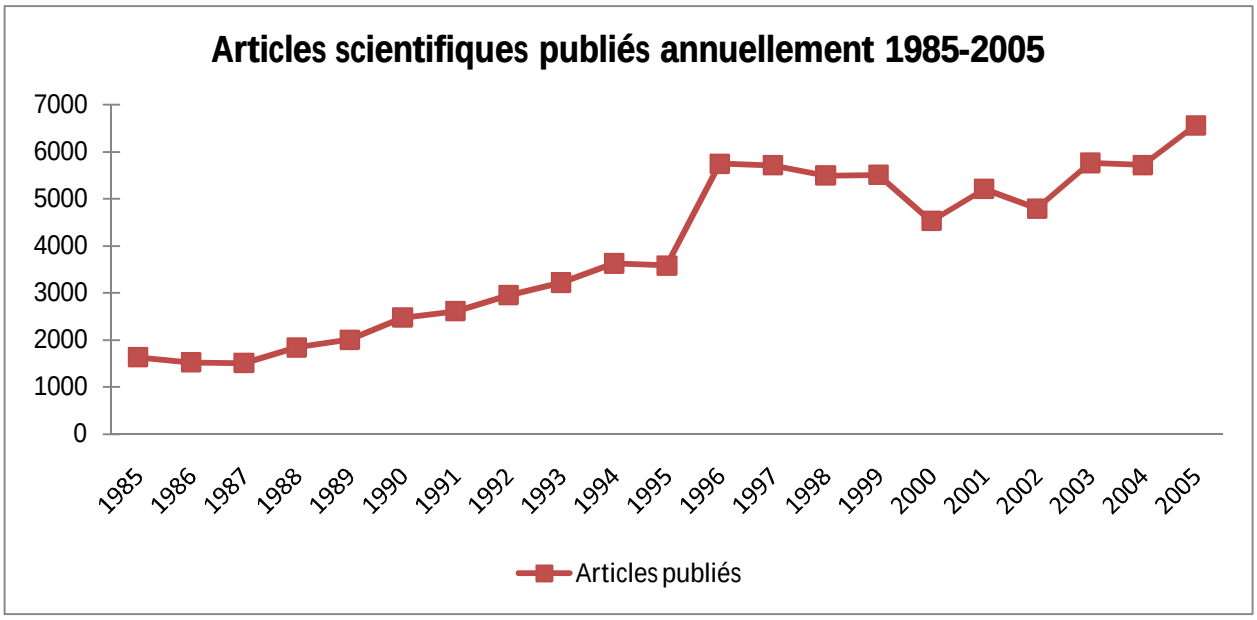


Figure 3.4 Articles scientifiques publiés annuellement 1985-2005

La courbe de publications de la Figure 3.4 présente plusieurs pointes de croissance importantes au cours de la période étudiée. À la suite d'une croissance plutôt constante du nombre de publications pour la période 1987-1994, on note une variation importante des pointes suivantes où les pourcentages représentent la croissance (ou décroissance) du nombre d'articles entre les années spécifiées: 1995/1996 (160%), 2000/2001 (115%), 2002/2003 (120%) et 2004/2005 (115%). L'explosion des publications pour 1995/1996 pourrait s'expliquer par une combinaison de la popularité soudaine de l'utilisation des mots-clés liés à la biotechnologie dans les articles scientifiques et des modifications internes effectuées à la base de données Scopus. La problématique liée à l'utilisation des mots-clés avait été abordée à la section 2.4.1 pour documenter l'effet inverse, soit de ne pas extraire des articles à cause des mots-clés utilisés. La Figure 3.4 illustre bien une autre problématique liée aux mots-clés soit l'inclusion massive de mots-clés spécifiques. Il serait intéressant de faire l'étude des mêmes mots-clés au niveau des *demandes* de financement pour les projets de recherche au cours de la même période pour voir si on y retrouverait la même « popularité » soudaine de certains mots-clés.

Analysons maintenant à l'aide de la Figure 3.5 la contribution annuelle par chercheur aux données de publications annuelles totales présentées précédemment à la Figure 3.4.

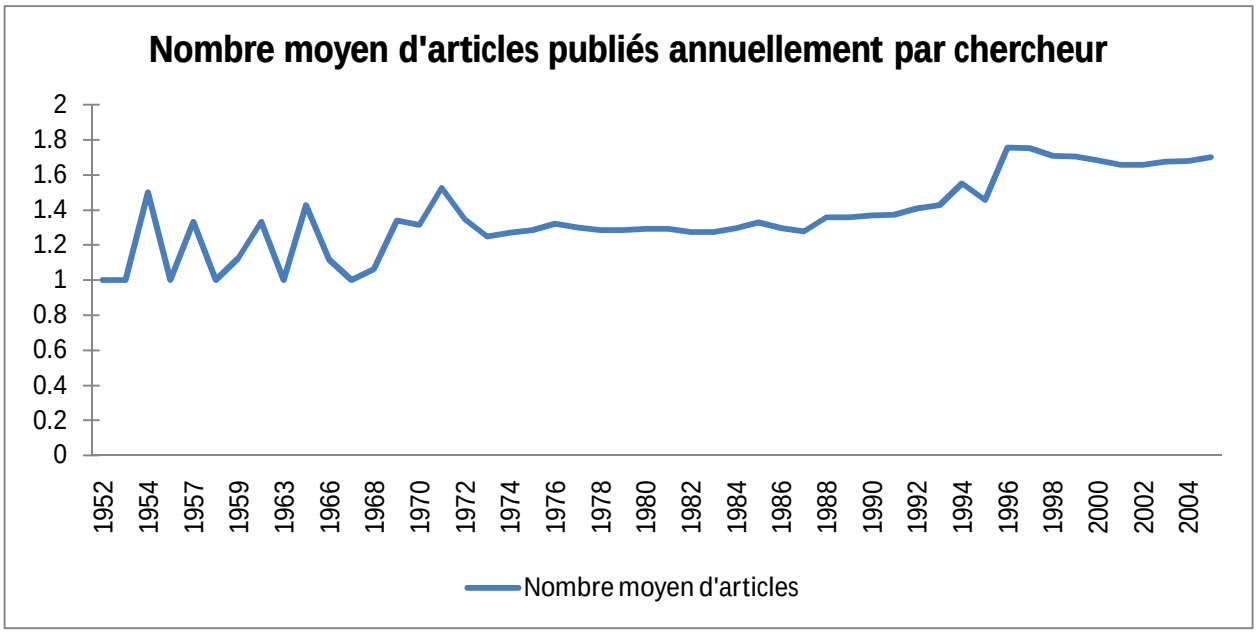


Figure 3.5 Nombre moyen d'articles publiés annuellement par chercheur

La Figure 3.5 donne un aperçu global de l'évolution du nombre d'articles publiés annuellement par chercheur sur une période de cinquante ans. On note une grande fluctuation des données pour la période 1952-1972. Voyons plus en détail la période 1985-2005 qui semble beaucoup plus stable.

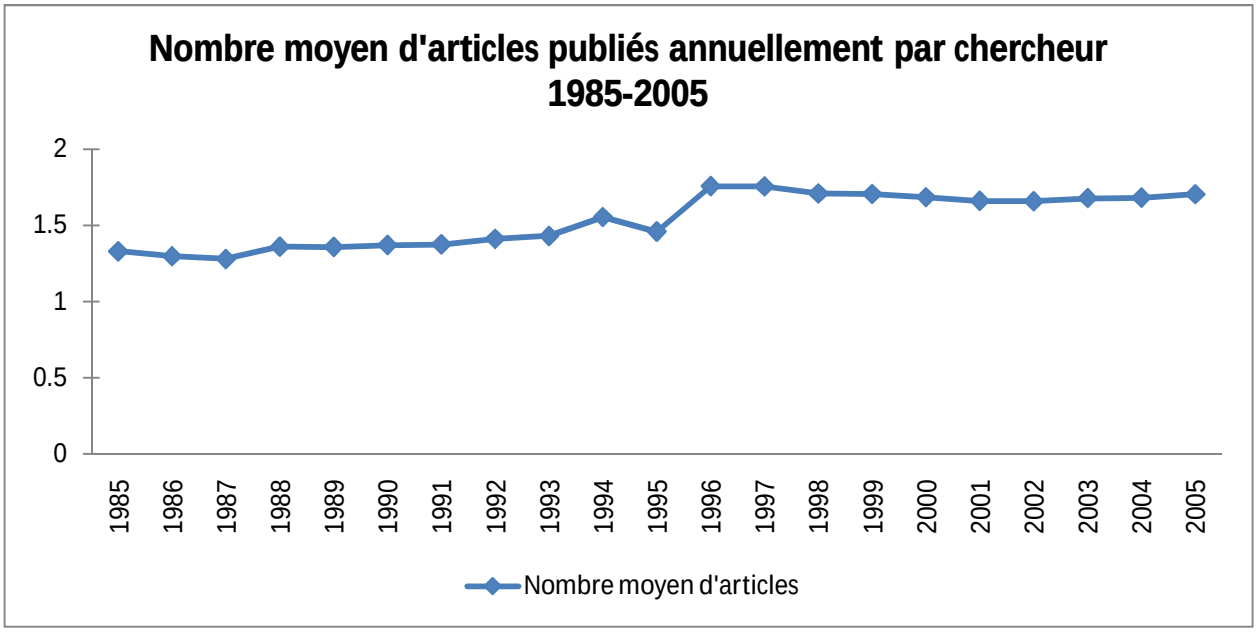


Figure 3.6 Nombre moyen d'articles publiés annuellement par chercheur 1985-2005

La courbe présentée à la Figure 3.6 suit une lente progression avec une pointe de croissance représentée en pourcentage de croissance (ou décroissance) entre les deux années spécifiées en 1995/1996 (120%) qui est constante avec le nombre annuel de publications présentées à la Figure 3.4. On note une légère décroissance du nombre moyen d'articles publiés pour les années 1985/1986 (98%), 1986/1987 (99%), 1994/1995 (94%), 1997/1998 (97%), 1999/2000 (99%) et 2000/2001 (99%).

Nous avons illustré à la Figure 3.4 la courbe générale du nombre total d'articles annuels, voyons maintenant le nombre moyen annuel de co-auteurs par article.

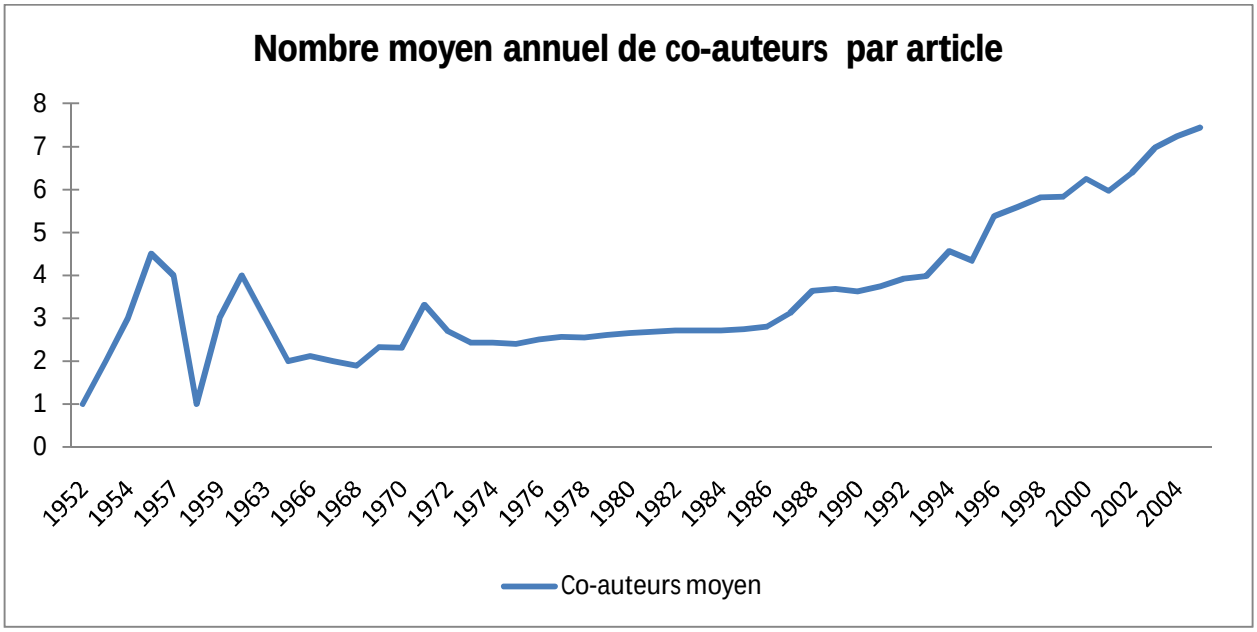


Figure 3.7 Nombre moyen annuel de co-auteurs par article

La Figure 3.7 présente une vue globale du nombre moyen de co-auteurs par article. Le nombre moyen de co-auteurs calculé en début d'échantillon souffre du faible nombre d'articles et par conséquent apparait en dent de scie. À partir de l'année 1974, la courbe présente une croissance presque constante du nombre moyen de co-auteurs avec une augmentation de la croissance à partir de 1985. Voyons plus en détails la période 1985-2005 à l'aide de la Figure 3.8.

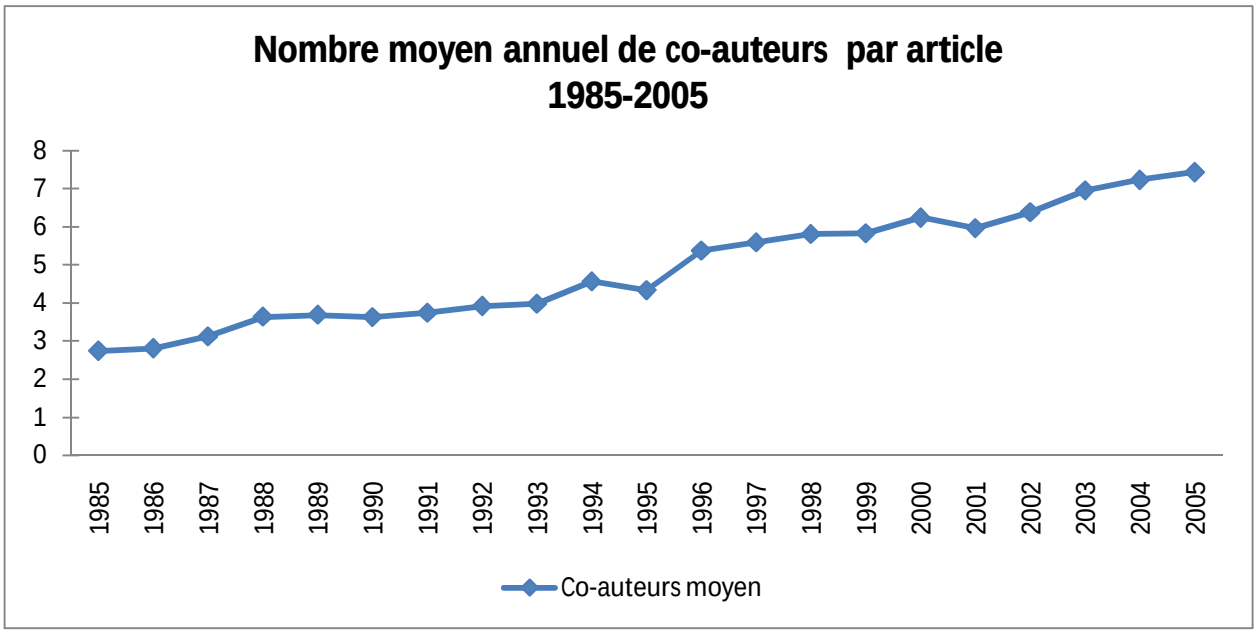


Figure 3.8 Nombre moyen annuel de co-auteurs par article 1985-2005

La Figure 3.8 présente une croissance générale du nombre moyen de co-auteurs par article en biotechnologie. En 1985, le nombre moyen de co-auteurs était de 2,7 alors qu'en 2005 celui-ci était de 7,7. Plusieurs phénomènes peuvent expliquer cette croissance. Premièrement, il y a l'éternel jeu du « publish or perish » qui est commun à tous les domaines de recherche et qui explique qu'individuellement chaque chercheur veut toujours publier plus d'articles. Deuxièmement, la nature même de la biotechnologie, décrite à la section 1.1, est complexe et multidisciplinaire ce qui nécessite de plus en plus le recourt à des *équipes* de recherche où chacun a un rôle spécifique. Au sein de l'équipe, le concept du « divide and conquer » prend tout son sens, car chacun en fait globalement moins mais on réussit à produire un résultat plus complexe. On peut lier cette augmentation de 5 co-auteurs en moyenne à la croissance des équipes de recherche. Par contre, il ne faut pas oublier que ce n'est pas nécessairement tous les membres d'une équipe qui publient ou qui bénéficient au même niveau du financement reçu par l'équipe ou même par un seul individu qui redistribue le financement sans écrire directement les articles. Plus les chercheurs travaillent au sein d'équipe de grande taille, plus la facilité d'être co-auteur augmente sans nécessairement avoir à dédier autant de ressources que si on était seul, car chaque expert est responsable de résoudre un problème spécifique. Donc une facilité existe mais la répartition des tâches est rarement égale. De plus, comme nous le verrons plus tard dans

l’analyse, le fait d’être co-auteur dans le cadre d’un projet permet de créer des liens facilitant ainsi la collaboration future.

Le graphique suivant représente la somme annuelle des articles, dont au moins un des auteurs possède une affiliation canadienne, distribués selon les grappes géographiques. Le concept de grappe utilisé pour l’analyse a été défini à la section 2.4.1.1 et fait référence à la catégorisation des villes selon deux groupes soit « Grappes liées au Québec » et « Grappes hors Québec ». Les données sont tirées de la base de données **BIO_Matrix** qui est la matrice finale générée par le processus défini à la Figure 2.8. Tel qu’illustrée la source des données provient partiellement de la base de données Scopus. La somme annuelle est basée sur le total annuel des publications scientifiques.

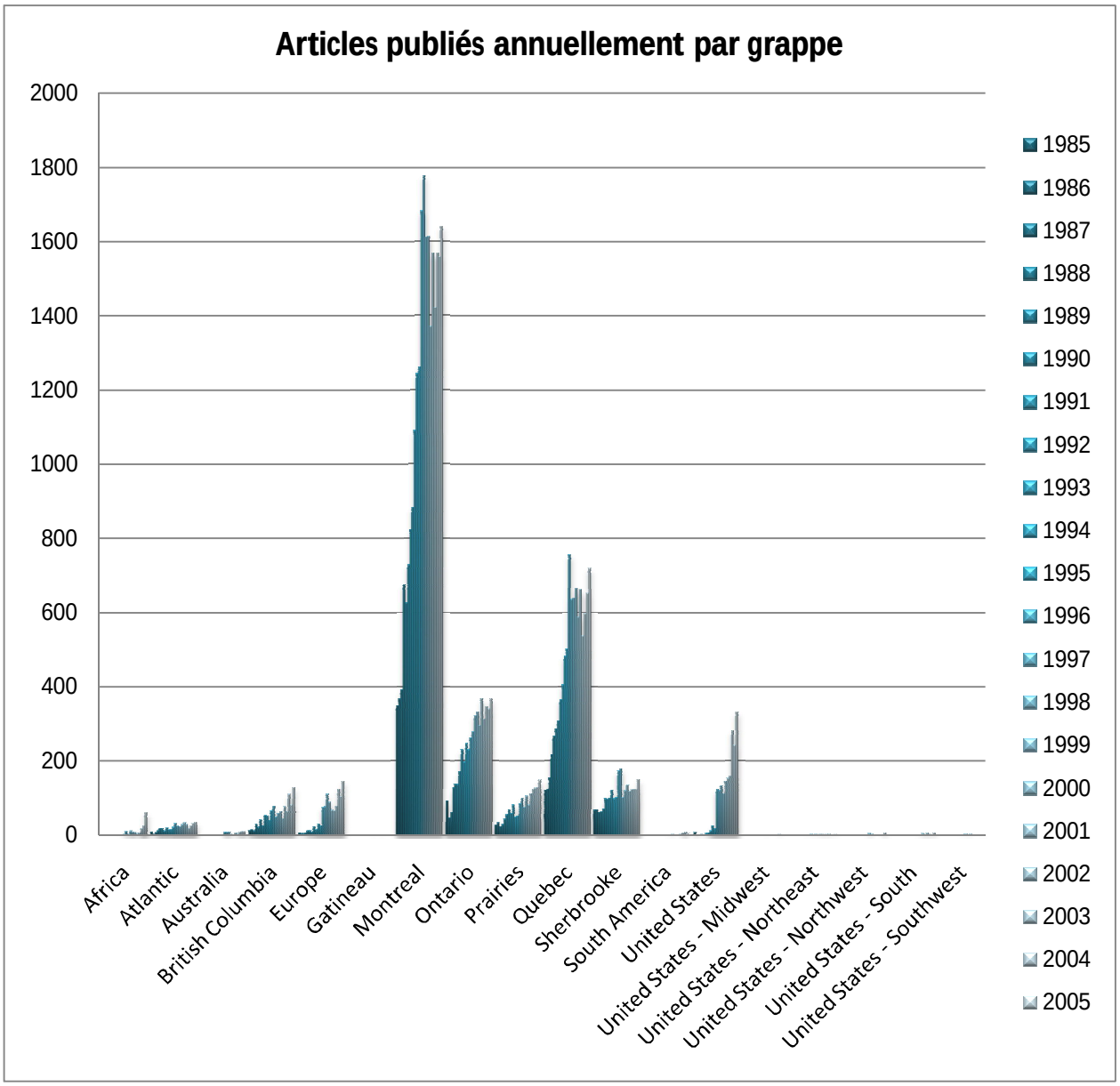


Figure 3.9 Articles publiés annuellement par grappe

Les résultats illustrés à la Figure 3.9 ne sont pas surprenants, car on s'attend à avoir plus de publications dans les grappes canadiennes, une affiliation canadienne étant requise pour faire partie des critères d'inclusion de la recherche. De plus, dans le cas où la grappe n'est pas définie, c'est-à-dire que son affiliation est manquante ou ne peut être localisée géographiquement l'article n'apparait pas dans les statistiques de la Figure 3.9 ce problème est présenté au Tableau 3.1.

Par contre, il est intéressant de voir l'ensemble des liens nationaux et internationaux existant à travers les différentes grappes.

Tableau 3.1 Différence au niveau de la comptabilisation des articles publiés annuellement

Années	*Articles annuels (basés sur les grappes)*	*Articles annuels (basés sur les articles uniques)*	*Variation*
1985	680	1629	42%
1986	663	1525	43%
1987	711	1512	47%
1988	1156	1839	63%
1989	1192	2007	59%
1990	1375	2474	56%
1991	1515	2612	58%
1992	1735	2954	59%
1993	1984	3220	62%
1994	2230	3630	61%
1995	2279	3582	64%
1996	3286	5748	57%
1997	3253	5719	57%
1998	3100	5499	56%
1999	3150	5513	57%
2000	2771	4530	61%
2001	3164	5212	61%
2002	2859	4795	60%
2003	3341	5768	58%
2004	3304	5725	58%
2005	3755	6568	57%

Le Tableau 3.1 présente un problème central à cette recherche, soit le manque de standardisation des données à la source, ce qui a comme résultat un nombre important de champs incomplets. Le résultat est évident, la colonne basée sur les grappes contient en moyenne 57% des données totales qui elles sont comptabilisées dans la colonne basée sur les articles uniques. Pour nos résultats, cela implique que des données sont imprécises quant à l'aspect géographique. La taille des bases de données et le manque de temps expliquent l'impossibilité de corriger manuellement ces données. Avec une variation médiane de 58% et une variation moyenne de 57% entre obtient une moyenne de 43% des données perdues. Concrètement, cette répartition géographique incomplète n'impactera que cette section des statistiques descriptives.

Regardons plus en détails les grappes québécoises de façon à mieux cerner l'évolution des publications dans les trois principales grappes québécoises, soit Montréal, Québec et Sherbrooke.

La première grappe, soit celle de Montréal, est la plus important de l'étude et elle est présentée à la Figure 3.10.

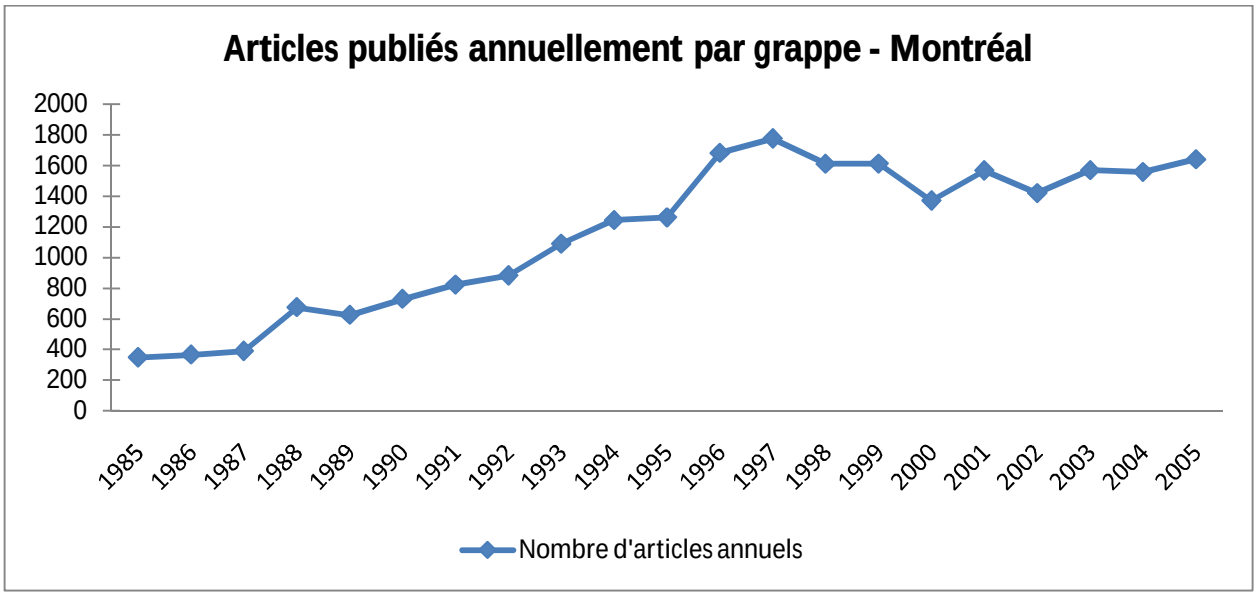


Figure 3.10 Articles publiés annuellement par grappe - Montréal

La courbe présentée à la Figure 3.10 pour la grappe Montréal suit la courbe générale de publication illustrée à la Figure 3.4. On note une décroissance des publications par rapport à l'année précédente pour 1988/1989 (93%), 1997/1998 (91%), 1999/2000 (85%), 2001/2002 (91%) et 2003/2004 (99%). Une croissance importante des publications est observée pour les années 1987/1988 (173%), 1995/1996 (133%) et 2000/2001 (114%). On note toutefois une certaine stabilisation des publications au cours de la dernière décennie. La deuxième grappe québécoise en importance, soit Québec, est présentée à la Figure 3.11.

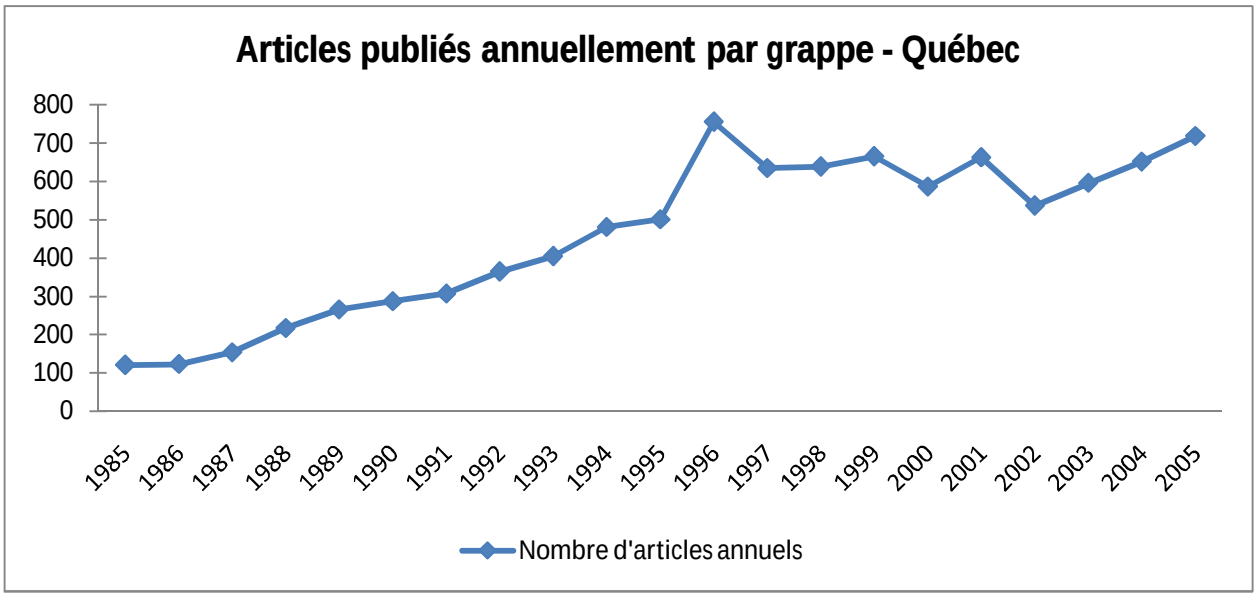


Figure 3.11 Articles publiés annuellement par grappe - Québec

La courbe présentée à la Figure 3.11 suit une croissance similaire aux autres courbes de grappes avec une pointe de croissance en 1995/1996 (133%) puis une décroissance jusqu'en 2001/2002. La dernière grappe étudiée sera celle de Sherbrooke présentée à la Figure 3.12.

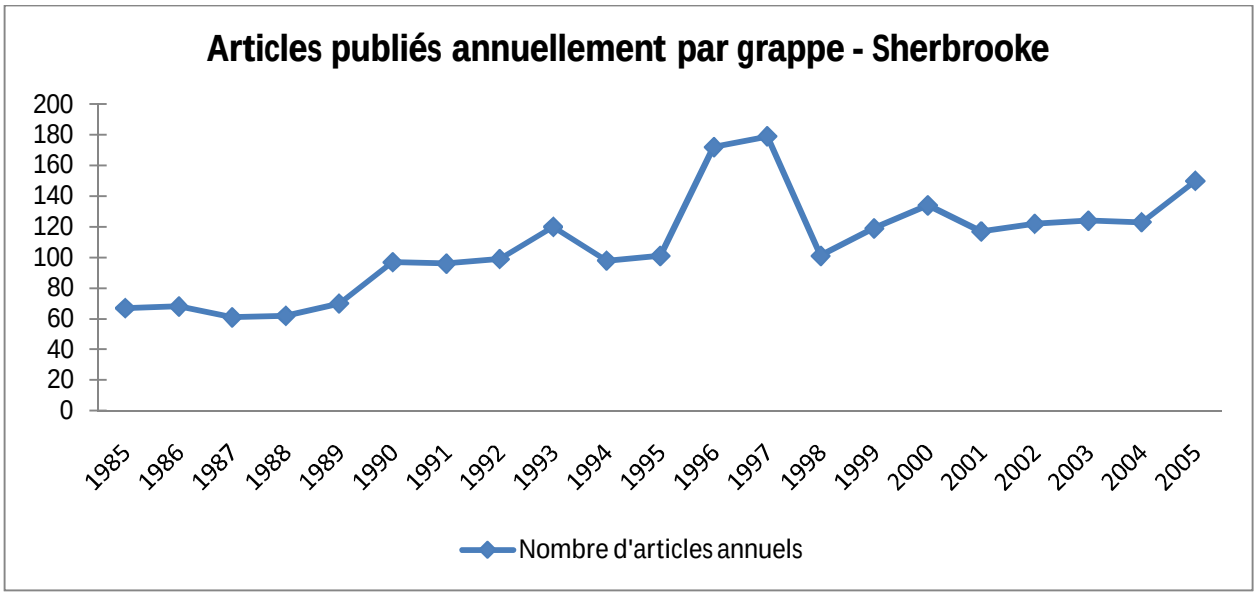


Figure 3.12 Articles publiés annuellement par grappe - Sherbrooke

La courbe présentée à la Figure 3.12 suit une croissance beaucoup plus faible que celles présentées pour les grappes de Montréal et Québec. On note un nombre annuel plus faible de publications scientifiques.

Les trois figures précédentes présentent toutes une même stabilisation à partir de l'année 1997 qui sera comparée à la section 3.3 aux montants du financement.

3.2.2 Chercheurs

La deuxième catégorie de statistiques tirées de Scopus traite des chercheurs présents dans les réseaux canadiens et des liens existants entre ces chercheurs. Ces liens entre co-auteurs sont capturés lors des publications.

La Figure 3.13 présente le nombre annuel de chercheurs actifs tirés de la base de données Scopus. Un chercheur est considéré actif dans un réseau annuel lorsqu'il est co-auteur d'une publication scientifique et qu'il y a unicité de la combinaison de l'année, du nom de famille complet et des initiales du prénom.

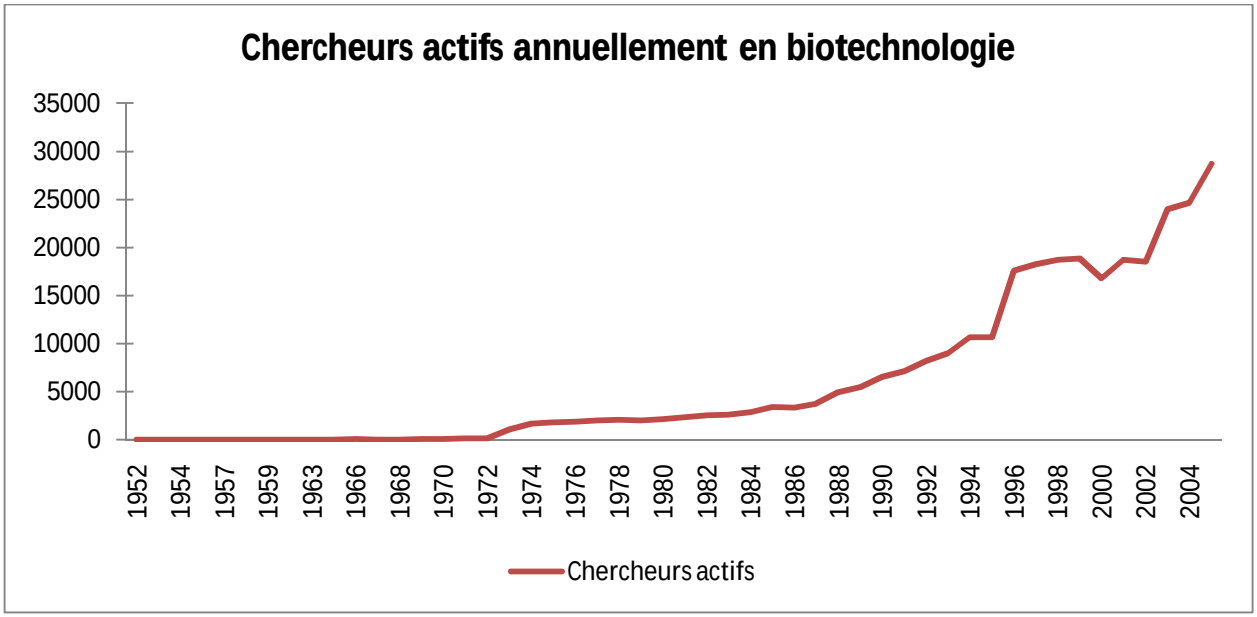


Figure 3.13 Chercheurs actifs annuellement en biotechnologie

Encore une fois, nous commençons par présenter une vue globale de l'évolution du nombre de chercheurs ayant publiés au moins un article scientifique. Une croissance lente à partir de 1972 (1056 chercheurs) pour croître plus rapidement jusqu'en 1995/1996 où on observe une pointe de croissance (165% c'est-à-dire le passage de 10643 à 17599 chercheurs) et terminer en 2005 avec 28678 chercheurs actifs. Il faut faire attention à cette figure et la suivante, car un chercheur est dit

actif s'il est co-auteur pour un article en biotechnologie. Cela inclus évidemment des chercheurs mais aussi des étudiants et des techniciens de laboratoire. Le nombre annuel peut donc être artificiellement gonflé et il est sensible aux mots-clés utilisés pour trouver les articles scientifiques. Voyons plus en détails la période 1985-2005.

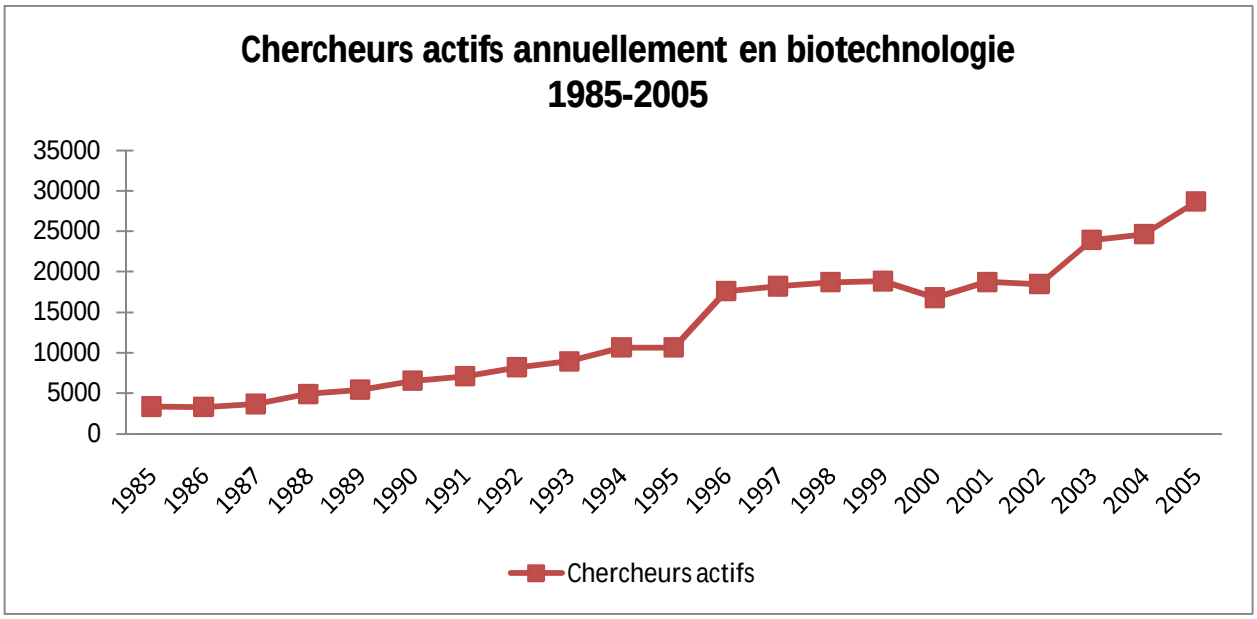


Figure 3.14 Chercheurs actifs annuellement en biotechnologie 1985-2005

La Figure 3.14 présente l'évolution du nombre de chercheurs à partir de l'année 1985 (3358 chercheurs jusqu'à l'année 2005 (28678 chercheurs). En étudiant la courbe, les points d'intérêts sont une pointe de croissance en 1987/1988 (133%), en 1995/1996 (165%) et en 2002/2003 (130%). On note une décroissance du nombre de chercheurs, par rapport à l'année précédente, aux années 1999/2000 (89%) et 2001/2002 (95%).

La variation du nombre de chercheurs actifs, donc qui publient, dans le réseau est le résultat d'un grand nombre de facteurs. La complexité des projets nécessitent des plus grosses équipes où chacun a un rôle important à jouer, ce qui inclu aussi les étudiants et les techniciens de laboratoires qui sont considérés comme des co-auteurs venant ainsi faire augmenter le nombre global de co-auteurs. Il y a un aspect cyclique qui est en lien avec le temps utilisé pour effectuer les recherches puis pour les publier. Le contexte historique vient jouer un rôle important notamment lorsqu'on fait référence à une augmentation de la popularité du domaine accompagné de financement important. On peut aussi parler du début des années 2000 où les activités de recherche ont pu être affectées indirectement par le crash de la bulle techno.

Après avoir étudié le nombre de chercheurs, analysons comment évoluent les liens créés entre eux. Les notions de ***collaborateur*** et de ***collaboration*** sont très importantes lorsqu'on parle de littérature scientifique. Le travail effectué dans le cadre d'un projet de recherche scientifique avec un ou plusieurs *collaborateurs* peuvent mener à des *collaborations*. Les mêmes collaborateurs peuvent produire différents articles scientifiques dans le cadre de plusieurs collaborations. La différence entre les mesures de ***co-auteurs*** et les collaborateurs vient du fait que l'unité de base n'est pas la même. Pour le nombre moyen annuel de co-auteurs, on se base sur le nombre moyen de chercheurs impliqués dans la rédaction d'articles scientifiques différents tandis que pour le nombre moyen annuel de collaborateurs, on se base sur le nombre de chercheurs différents ayant été impliqués avec les chercheurs. Par exemple, deux chercheurs ayant publié chacun 5 articles scientifiques au cours d'une année avec des chercheurs différents généreront un nombre moyen annuel de co-auteurs par article de 2 et un nombre moyen annuel de collaborateurs de 5. Voyons la tendance globale à l'aide de la Figure 3.15.

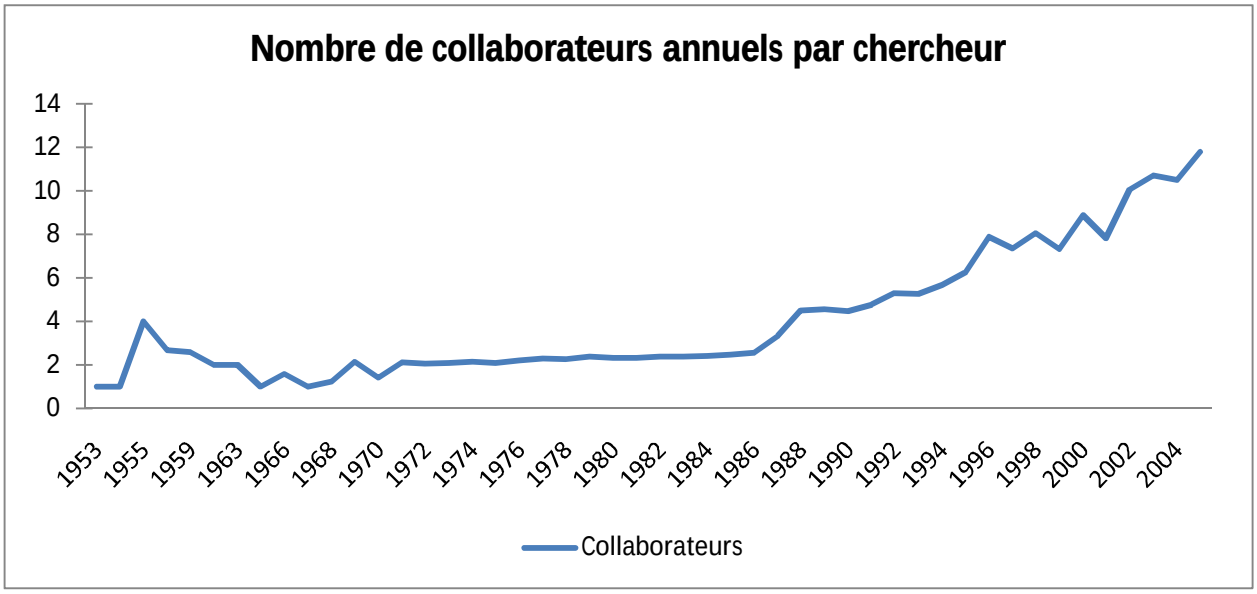


Figure 3.15 Nombre de collaborateurs annuels par chercheur

La Figure 3.15 présente une tendance de faible croissance entre 1972 avec 2,04 collaborateurs annuels jusqu'en 1986 avec 2,54 collaborateurs annuels. Par la suite, il y a une croissance plus forte mais en dents de scie pour terminer en 2005 avec 11,8 collaborateurs annuels. La Figure 3.16 présentera plus en détails la période 1985-2005.

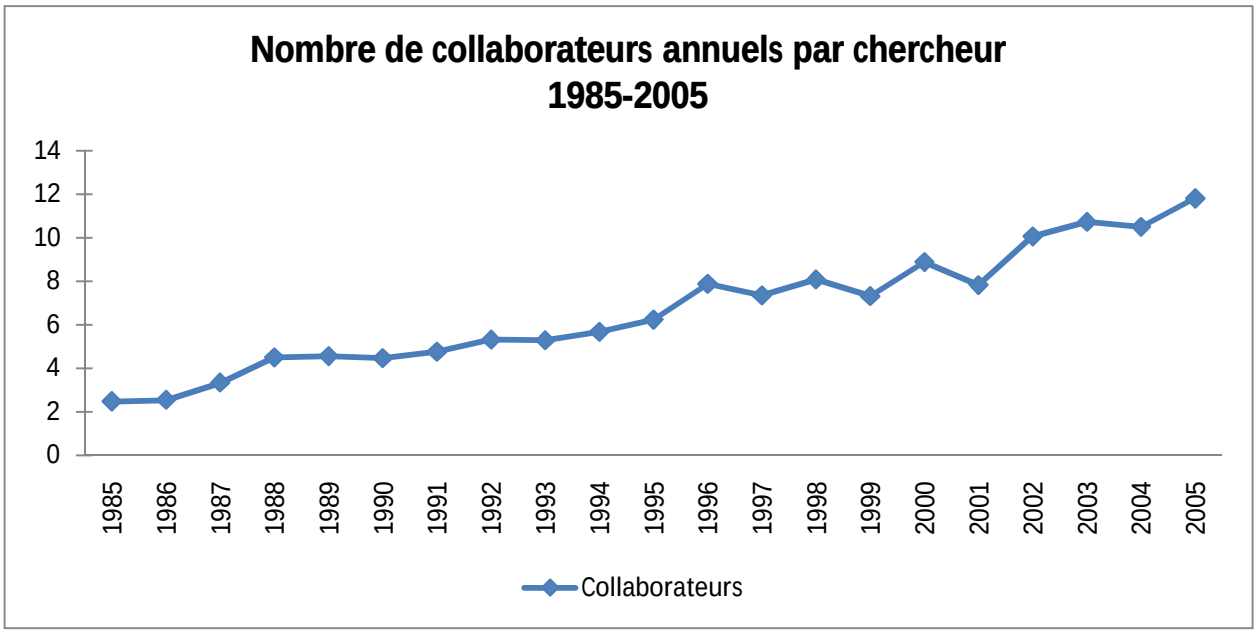


Figure 3.16 Nombre de collaborateurs annuels par chercheur 1985-2005

La Figure 3.16 présente une croissance de 2,47 collaborateurs en 1985 à 11,8 collaborateurs en 2005. Plusieurs pointes de croissance sont présentes pour les années 1986/1987 (131%), 1987/1988 (135%), 1995/1996 (126%), 1999/2000 (121%) et 2001/2002 (129%). La Figure 3.17 présente de son côté les collaborations annuelles. Note que les collaborations qui peuvent impliquer les mêmes collaborateurs. La taille des équipes de recherche a donc considérablement augmenté au cours des deux dernières décennies examinées. Cette augmentation peut s'expliquer par la nécessité de réaliser des projets de plus en plus complexes. De plus, l'équipe elle-même est composée de chercheurs mais aussi d'étudiants et de techniciens qui jouent un rôle *essentiel* à la recherche et donc sont nommés comme co-auteurs.

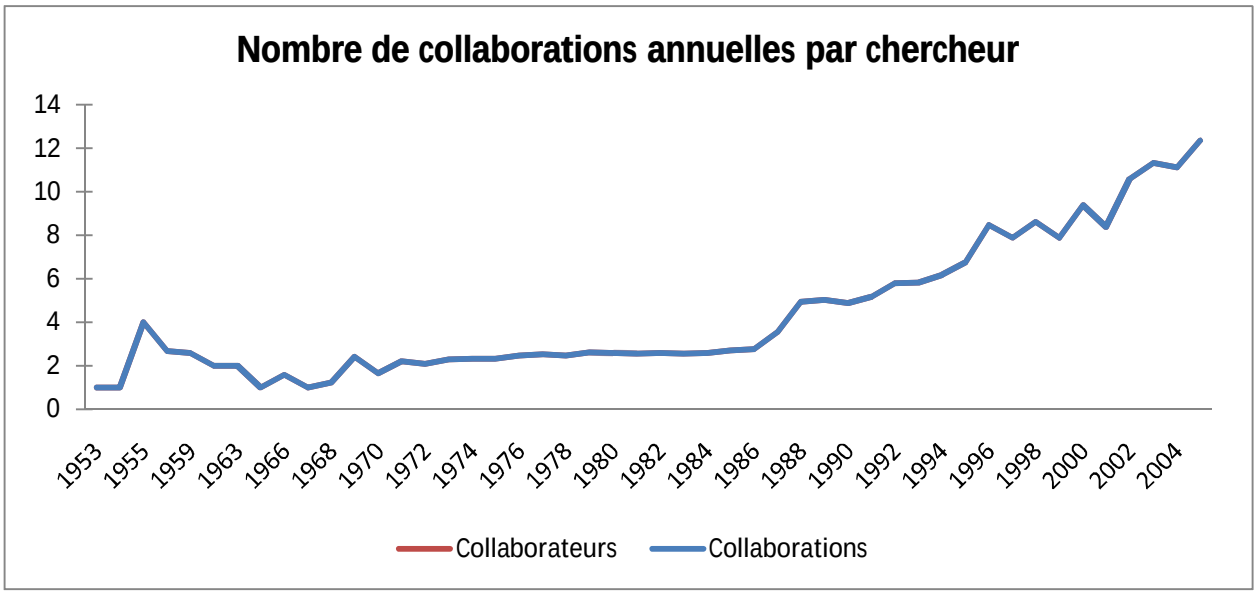


Figure 3.17 Nombre de collaborations annuelles par chercheur.

La Figure 3.17 illustrant les collaborations présente pratiquement la même courbe que la Figure 3.15 illustrant les collaborateurs mais avec des valeurs légèrement plus élevées. Il existe aussi une faible croissance entre 1972 avec 2,09 collaborations annuelles jusqu'en 1986 avec 2,77 collaborations annuelles. Par la suite, il y a une croissance plus forte mais en dents de scie pour terminer en 2005 avec 12,36 collaborations annuelles. La Figure 3.18 présentera plus en détails la période 1985-2005.

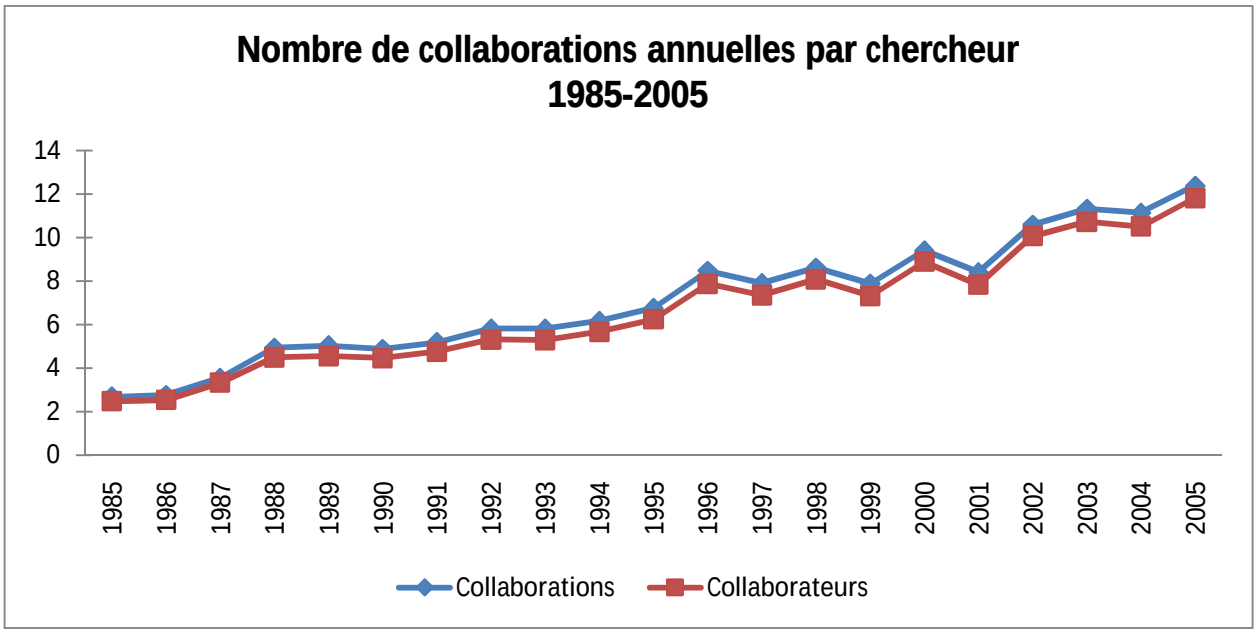


Figure 3.18 Nombre de collaborations annuelles par chercheur 1985-2005

La Figure 3.18 illustre plus précisément la faible différence existant entre les valeurs annuelles de collaborations qui sont plus élevées que les valeurs annuelles de collaborateurs. Cette situation signifie que certains chercheurs collaborent plus d'une fois avec les mêmes chercheurs au cours d'une année. Les mêmes constatations sont présentes pour les pointes de croissances des années 1986/1987 (128%), 1987/1988 (139%), 1995/1996 (125%), 1999/2000 (119%) et 2001/2002 (129%). Cette tendance à répéter les collaborations est mesurée de façon annuelle. Tel qu'observé, il est effectivement rare qu'un chercheur publie plus d'une fois avec le même co-auteur au cours de la même année. La confiance mutuelle entre deux chercheurs prend du temps et se mesure par les collaborations sur plusieurs années.

Après avoir vu quelques statistiques générales au niveau du financement public la deuxième partie présente maintenant des statistiques spécifiques sur les niveaux de financements reçus par les chercheurs universitaires.

3.3 Système d'information de la recherche universitaire

Cette section traite spécifiquement du financement reçu par les chercheurs universitaires œuvrant au sein d'universités québécoises. Pour l'analyse statistique qui suivra, nous traiterons donc, dans un premier temps, les données des contrats et des subventions regroupées puis, dans un deuxième temps, une séparation par type sera effectuée.

3.3.1 Aperçu

Une vue globale des montants reçus en financement public et privé, *tous domaines de recherche confondus* tirés de SIRU, pour la période 1985-2005 est présentée à la Figure 3.19.

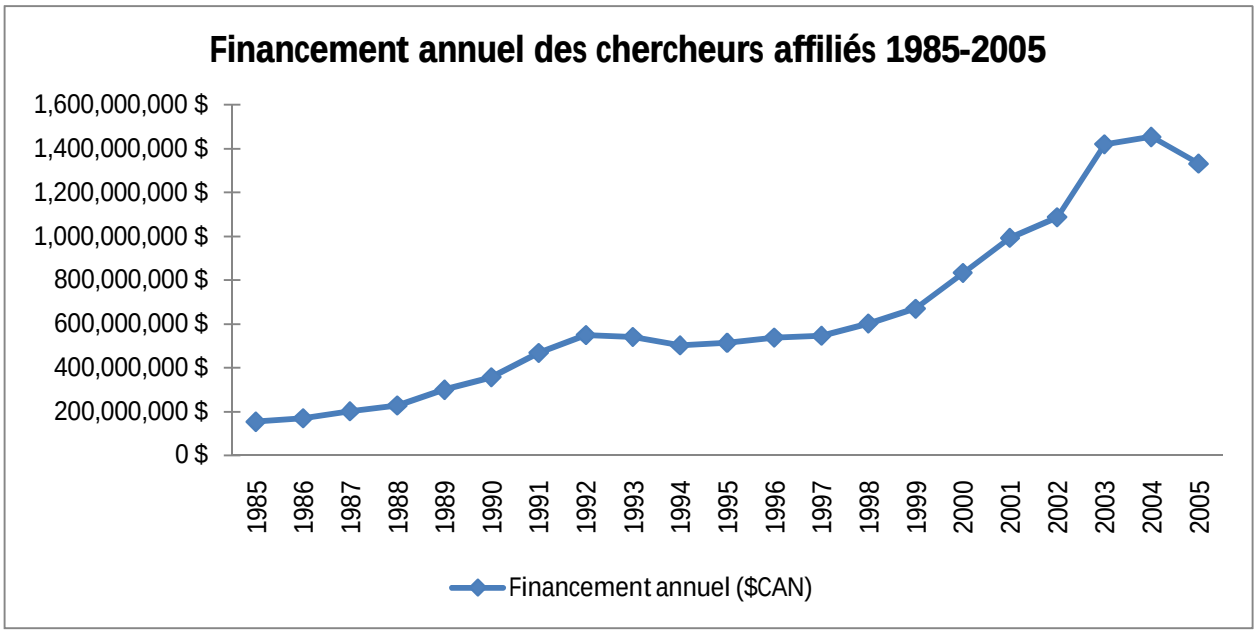


Figure 3.19 Financement annuel des chercheurs affiliés 1985-2005 - SIRU

La Figure 3.19 présente l'évolution du financement, en dollars constants de 2002, reçu à partir de l'année 1985 avec un montant de 154 627 439 $CAN pour croître jusqu'à l'année 1992 avec un montant de 550 171 267 $CAN. Par la suite, on observe une décroissance pour les années 1992/1993 (98%) et 1993/1994 (93%) pour reprendre une croissance à partir de l'année 1995. Cette croissance est maintenue jusqu'en 2003/2004 avec un montant de 1 454 150 059 $CAN puis une décroissance en 2004/2005 (92%) pour terminer avec 1 332 120 694 $CAN. On note que le ralentissement et la décroissance en fin de courbe suit évidemment les données présentées au début du chapitre à la Figure 3.1 au niveau du financement global des activités de recherche.

Il est intéressant d'analyser les données de la Figure 3.19 en séparant la proportion du financement en deux catégories, présentées dans le modèle général, soit les contrats et les subventions. La distinction entre le nombre et les montants du financement est essentielle pour saisir l'impact de chaque type de financement en plus d'aider à comprendre l'impact des variables du modèle général. Les montants des contrats et des subventions seront traités en premier, puis dans un second temps, le nombre de contrats et de subventions seront analysés.

3.3.2 Étude des montants des contrats et des subventions

La Figure 3.20 présente la distinction entre la proportion de contrats et la proportion de subventions.

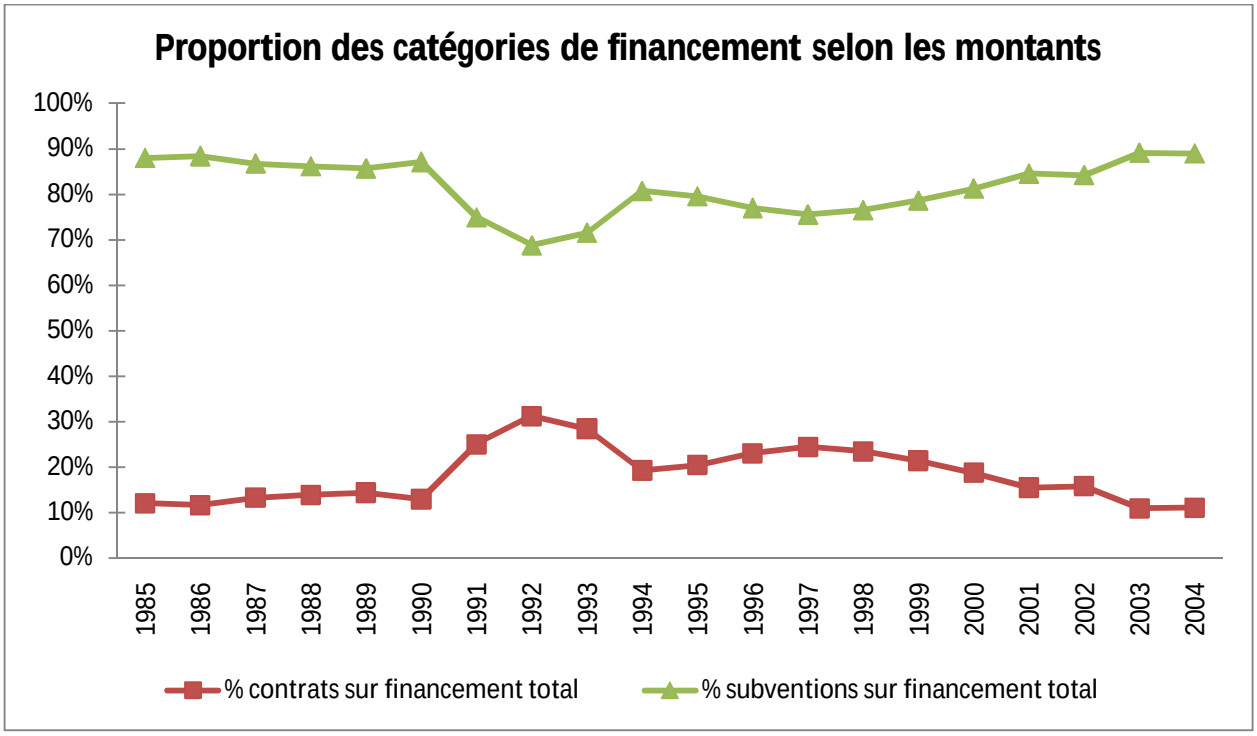


Figure 3.20 Proportion des catégories de financement selon les montants - SIRU

Les proportions illustrées à la Figure 3.20 sont relativement constantes pour les contrats (18%) et pour les subventions (82%) avec une variation entre les années 1990-1994 qui pourrait être attribuée a un changement de catégorisation du financement ou à l'augmentation des contrats dans le cadre l'un programme de financement.

La Figure 3.21 présente le détail des sous-catégories, soit l'appareillage, l'infrastructure et le fonctionnement pour les montants des contrats.

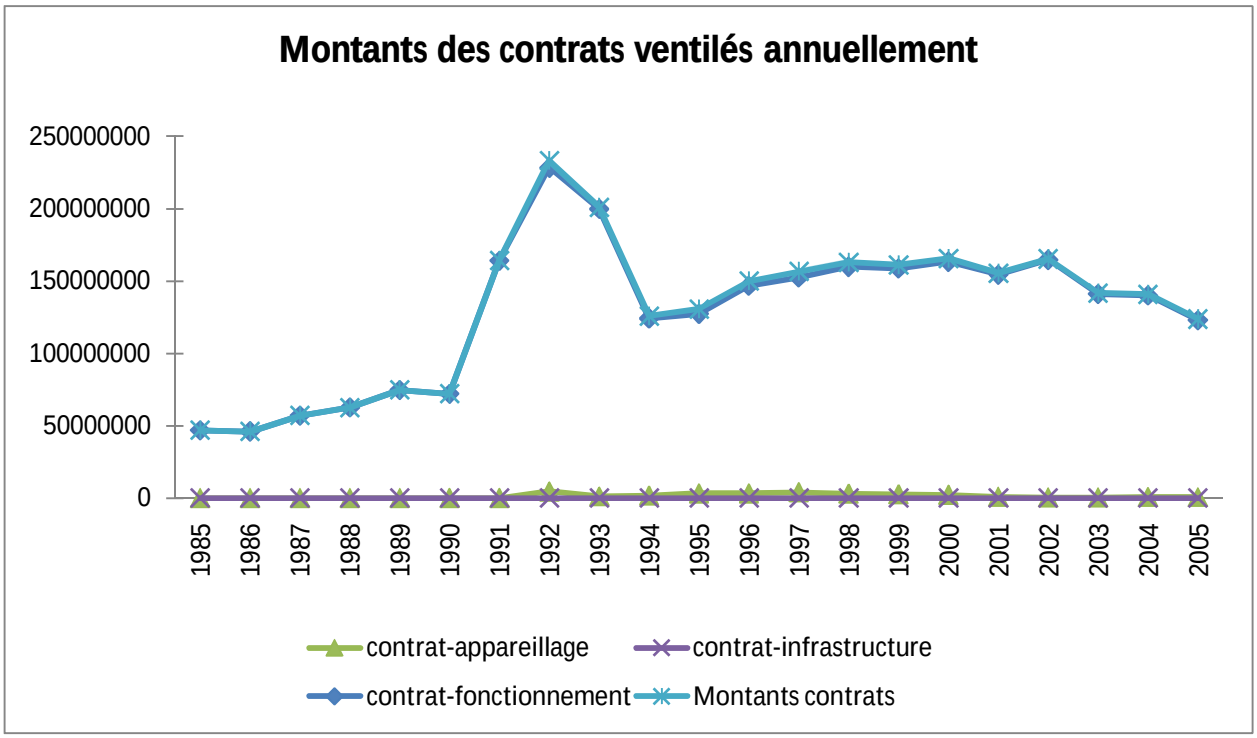


Figure 3.21 Montants des contrats ventilés annuellement - SIRU

La courbe présentée à la Figure 3.21 présente clairement que la majorité du financement, en dollars constants de 2002, par contrats provient de la sous catégorie « fonctionnement ». En effet, très peu de contrats sont accordés dans les autres catégories. Une pointe de croissance importante est observée pour les années 1990/1991 (241%) et 1991/1992 (144%) pour atteindre un maximum de 233 114 754 $CAN jamais atteint à nouveau. Un tel maximum pourrait être expliqué à plus court termes par une politique spécifique dans le cadre d'un mandat politique ou à plus long terme par une stratégie nationale comme la Stratégie Nationale en Biotechnologie établie en 1983 et dont les résultats porteraient fruits quelques années plus tard. Le reste de la courbe suit une croissance régulière de 1994 à 2002 pour décroître jusqu'en 2005.

La Figure 3.22 présente la même analyse mais pour les montants des subventions.

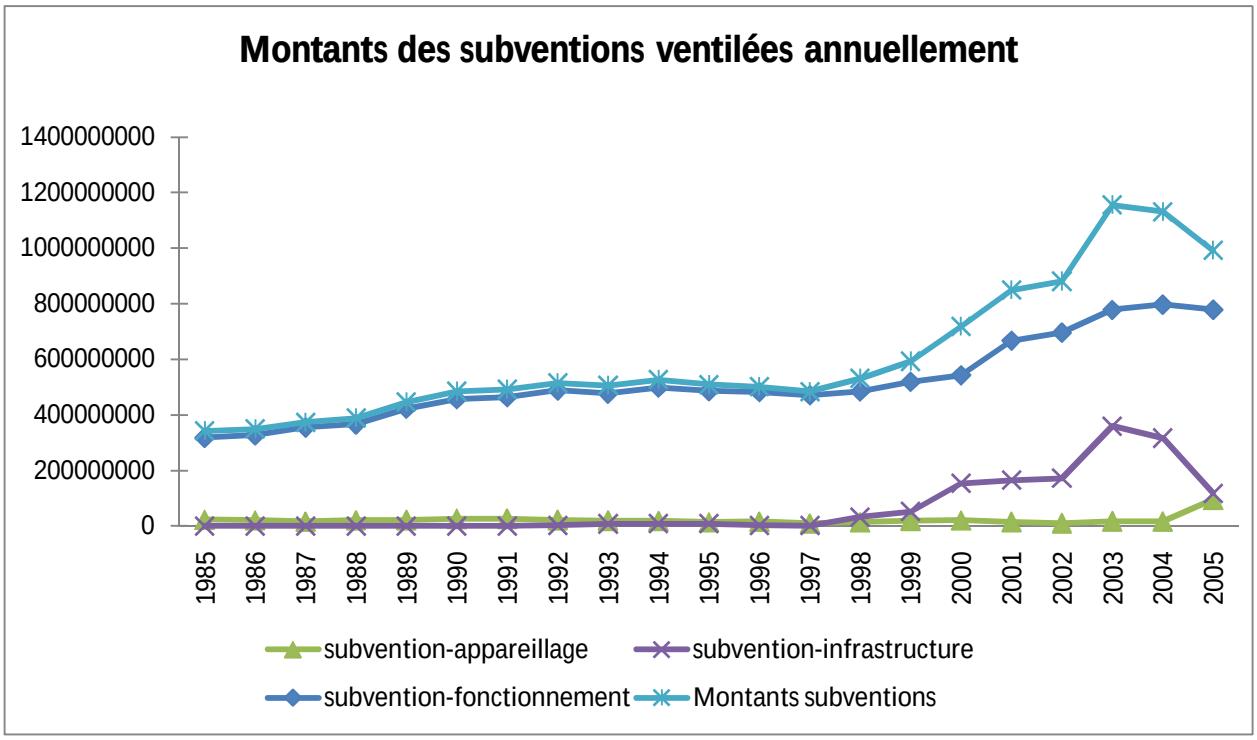


Figure 3.22 Montants des subventions ventilées annuellement - SIRU

La Figure 3.22 présente une courbe croissante dominée par les montants, en dollars constants de 2002, de la sous-catégorie « fonctionnement » jusqu'en 1997 où on note un changement majeur au niveau des sous-catégories. Alors que la sous-catégorie « fonctionnement » continue de croître un peu plus rapidement, on note que la sous-catégorie « infrastructure » croît rapidement de l'année 1998 jusqu'en 2003 pour redescendre en 2005 au niveau de la sous-catégorie « appareillage » qui avant commencée à croître en 2004. La courbe « infrastructure » peut être expliquée par des investissements massifs nécessaires pour pousser plus loin les limites de la recherche en biotechnologie.

La Figure 3.23 présente une vue intéressante sur les montants reçus par les universités québécoises *tous domaines de recherches confondus*.

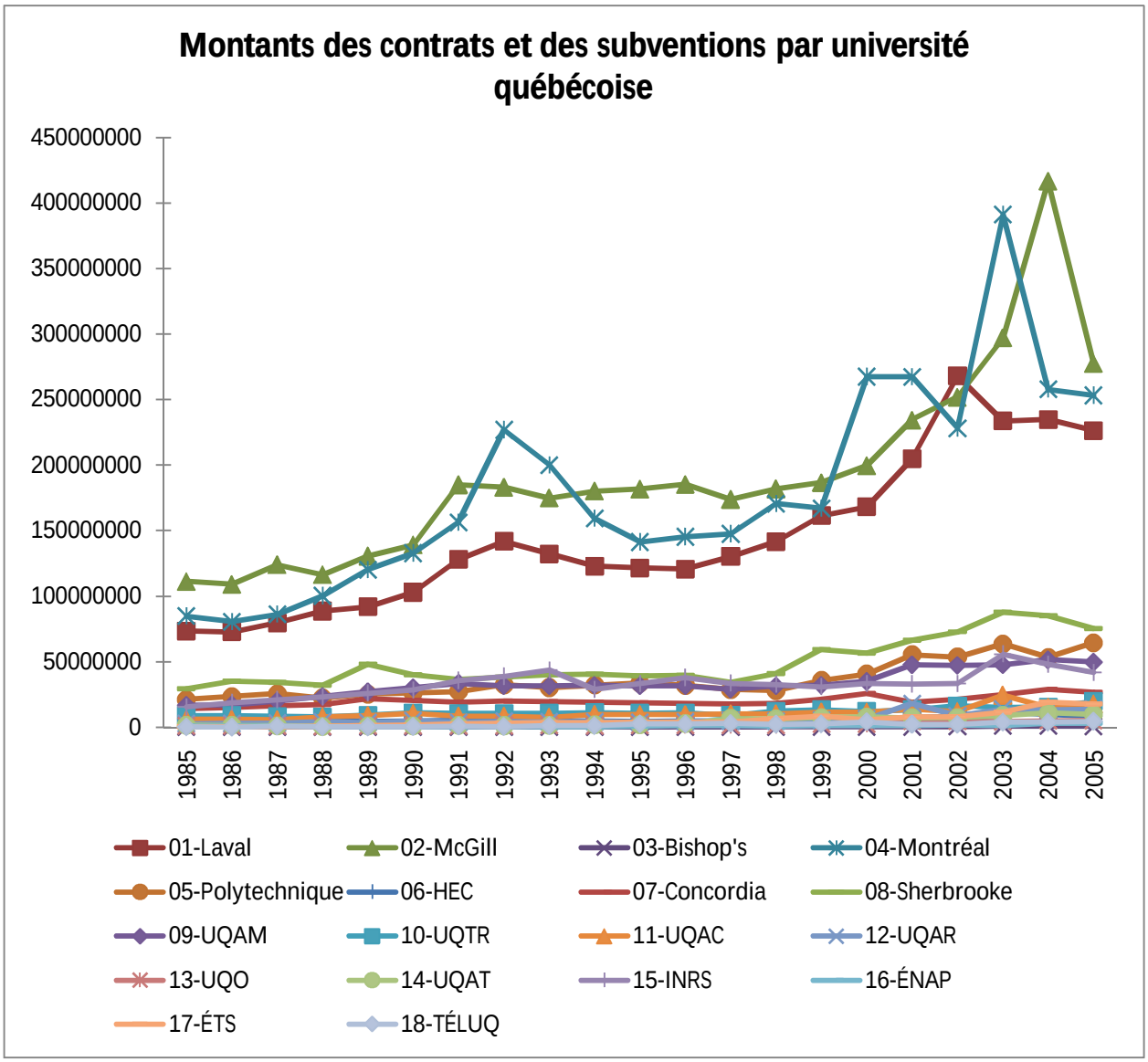


Figure 3.23 Montants des contrats et des subventions par université québécoise - SIRU

La Figure 3.23 illustre bien que les universités reçoivent globalement, en dollars constants de 2002, plus de financement avec certaines pointes au cours des années 1991-1993 et 2002-2004. Les universités Montréal, McGill et Laval dominent au niveau des montants reçus avec Sherbrooke qui se démarque un peu du reste des établissements québécois. Il y a donc au moins une université recevant plus de financement que la moyenne par grappe importante pour la publication d'articles scientifiques en biotechnologie. On représente ici l'ensemble des établissements québécois pour avoir un portrait complet, pour l'analyse nous ferons certains

regroupements. Il faut cependant noter que nous ne tenons pas compte de l'effet de la taille des universités.

3.3.3 Étude du nombre de contrats et de subventions

Appliquons la même analyse mais pour le **nombre** de contrats et de subventions. La Figure 3.24 présente cette distinction en pourcentage.

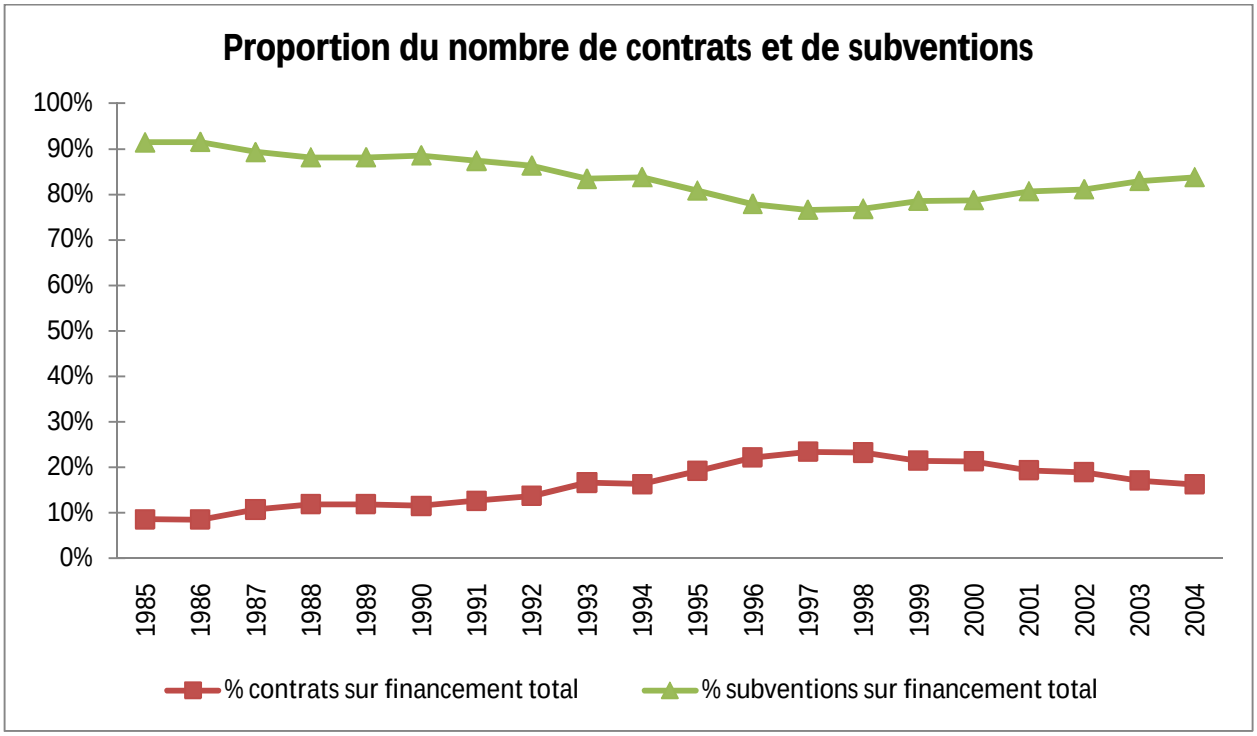


Figure 3.24 Proportion du nombre de contrats et de subventions - SIRU

La Figure 3.24 présente donc des proportions moyennes de contrats (16%) et de subventions (84%) relativement constantes sur une période de 20 ans. La majorité du financement provient donc des subventions et la proportion demeure relativement stable au fils des ans.

Toujours en travaillant avec les données de financement pour l'ensemble des chercheurs des universités québécoises, tous domaines confondus, les catégories seront maintenant ventilées en trois sous-catégories soit l'appareillage, l'infrastructure et le fonctionnement. Le graphe du financement selon le **nombre** de contrats ventilés sera présenté à la Figure 3.25 et le financement selon le **nombre** de subventions ventilées sera présenté à la Figure 3.26.

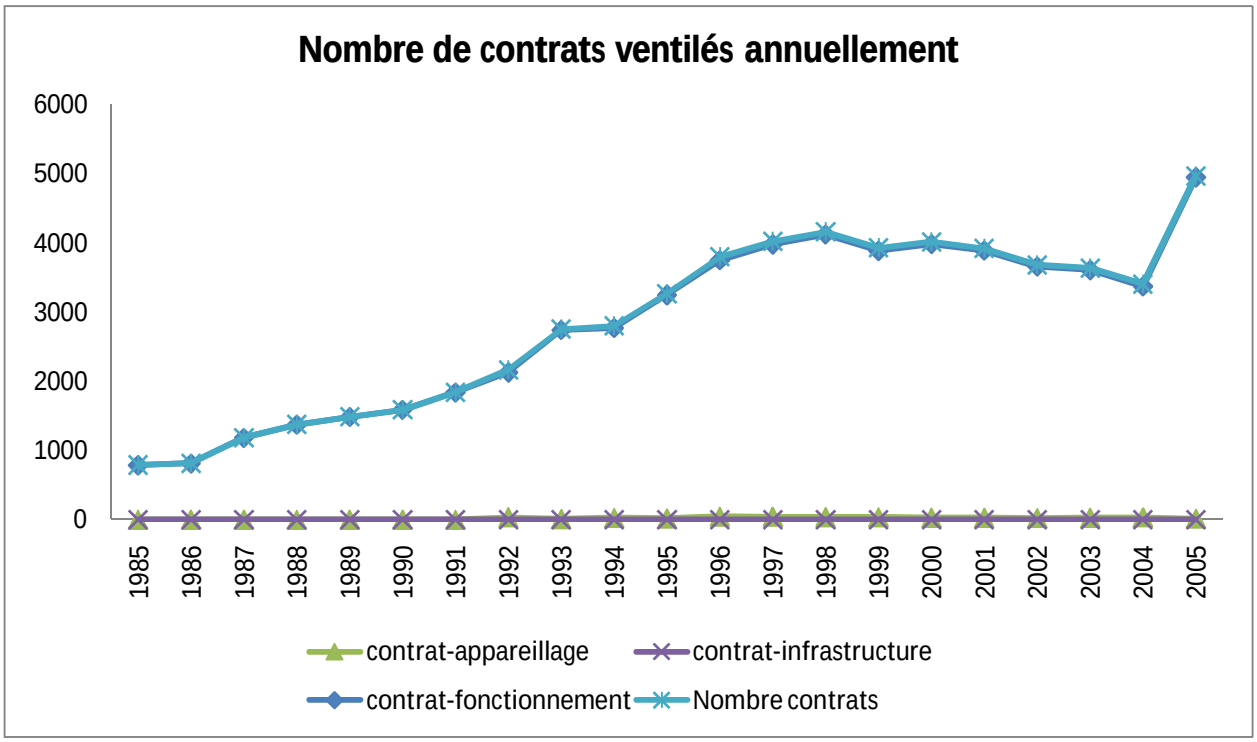


Figure 3.25 Nombre de contrats ventilés annuellement - SIRU

La Figure 3.25 présente le nombre de contrats ventilés selon l'appareillage, l'infrastructure et le fonctionnement. Les contrats d'appareillage et d'infrastructure sont pratiquement nuls tandis que ceux de fonctionnement constituent la majeure partie du financement sous la forme de contrats. Voyons maintenant le nombre de subventions ventilés selon les mêmes trois sous-catégories.

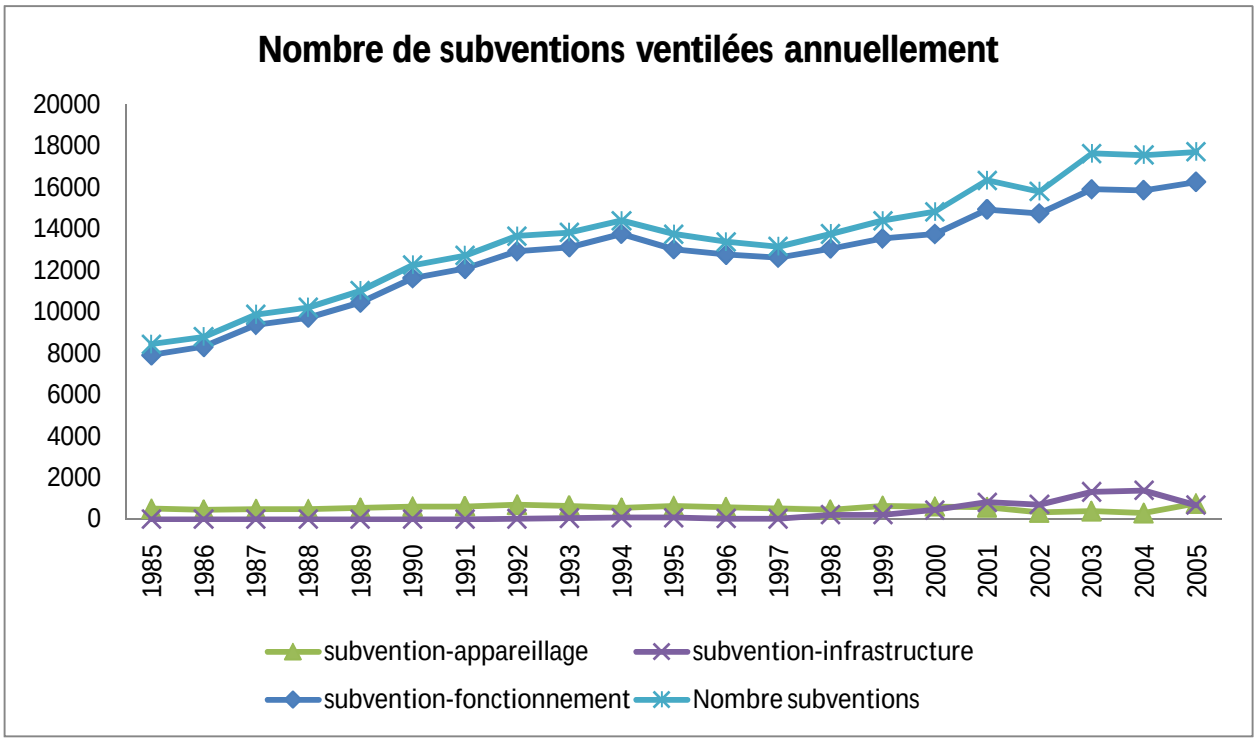


Figure 3.26 Nombre de subventions ventilées annuellement - SIRU

La Figure 3.26 présente le nombre de subventions ventilées selon l'appareillage, l'infrastructure et le fonctionnement. On retrouve le même comportement qu'à la Figure 3.25 où la majorité du financement provient de la sous-catégorie « fonctionnement ». Par contre, la sous-catégorie « appareillage » prend de l'importance avec le temps.

Le graphe présenté à la Figure 3.27 représente le nombre annuel de contrats et subventions par établissements québécois.

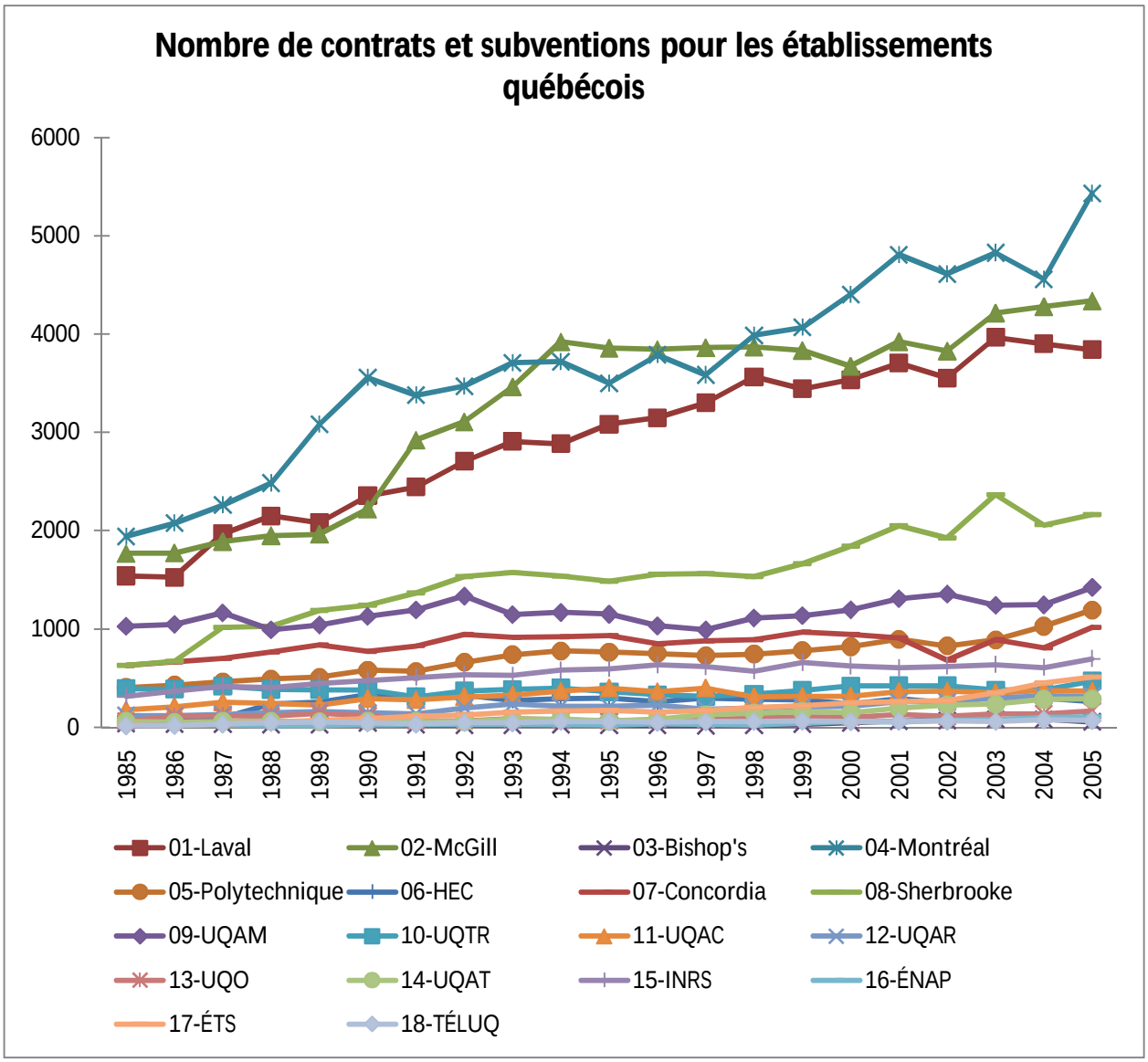


Figure 3.27 Nombre de contrats et subventions pour les établissements québécois - SIRU

Les courbes présentées à la Figure 3.27 illustrent les établissements en fonction du nombre de contrats et subventions. Les universités qui se démarquent sont Montréal, McGill, Laval et Sherbrooke. On retrouve donc des universités dans chacune des grappes actives présentées à la Figure 3.9 au niveau des publications scientifiques en biotechnologie. On représente ici l'ensemble des établissements québécois pour avoir un portrait complet, pour l'analyse nous ferons certains regroupements. Il faut cependant noter que nous ne tenons pas compte de l'effet de la taille des universités.

3.3.4 Notes

Il est intéressant ici de comparer la différence entre le graphe du nombre de subventions présenté à la Figure 3.26 et le graphe des montants des subventions présenté à la Figure 3.22. On note que le nombre de subventions suit une croissance moyenne de 3% sur toute la durée tandis que les montants des subventions suivent une croissance plus importante à partir de 1997. Pour le nombre de contrats présenté à la Figure 3.25 on observe une courbe décroissante tandis que les montants des contrats présentés à la Figure 3.21 on trouve une faible croissance suivie d'une décroissance.

En comparant ces résultats avec le graphe général des publications à la Figure 3.4 on s'aperçoit que le nombre de publications croit annuellement à un taux relativement stable depuis l'année 2000. Par contre, on constate qu'il n'y a pas globalement plus de contrats ou de subventions mais que les montants annuels sont de plus en plus importants pour les subventions, toutes catégories confondues, depuis 1997. Ainsi une croissance faible du nombre de publications en fonction des montants investis pourrait être expliquée, partiellement, par la redistribution des ressources pour la production de brevets qui ne sont pas comptabilisés dans cette étude.

3.4 Statistiques globales des réseaux

Nous avons présenté précédemment des statistiques tirées directement des bases de données sources. Cette section présente des statistiques sur les données tirées de la matrice finale générée pour fin d'analyses. Ces données permettent d'illustrer le comportement global des réseaux à l'aide de différentes propriétés de ceux-ci basées sur des fenêtres temporelles d'un, trois et cinq ans. Ces données ne sont pas présentes dans le modèle général mais permettent de qualifier son comportement qui sera analysé au CHAPITRE 4.

La Figure 3.28 présente le diamètre maximal du réseau, c'est-à-dire le chemin le plus long existant entre deux sommets du réseau. C'est donc le diamètre de la composante, dont le concept a été défini à la section 1.3.1.2, qui possède la plus grosse taille.

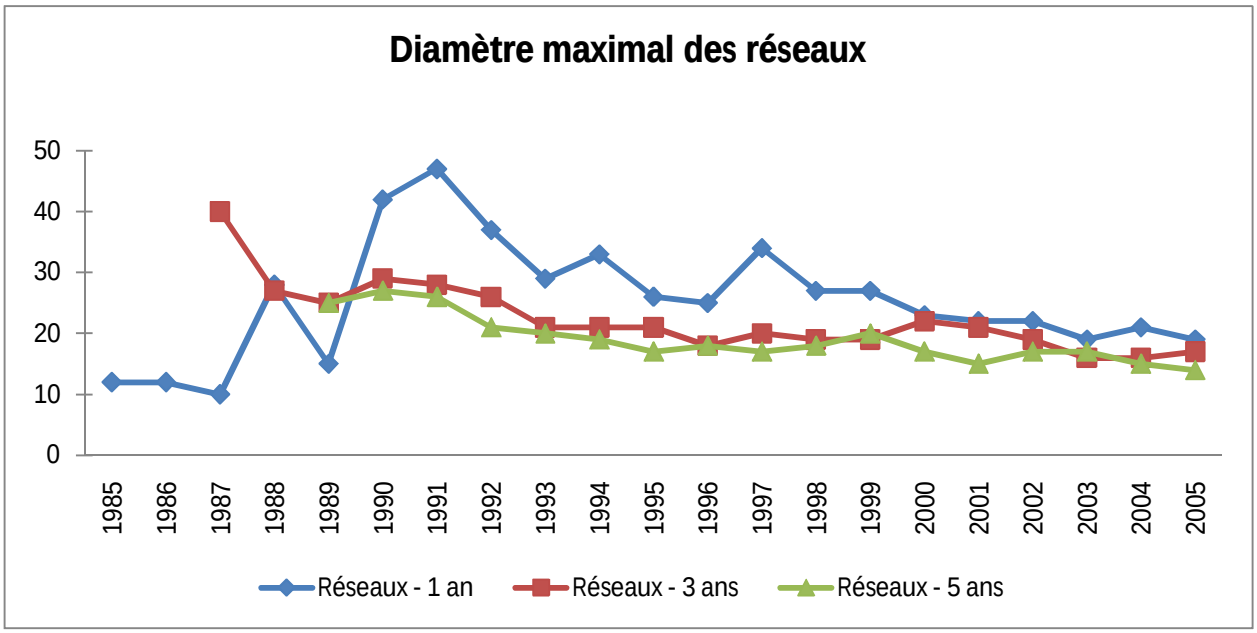


Figure 3.28 Diamètre maximal des réseaux

La Figure 3.28 illustre qu'il y a une faible décroissance au niveau de la taille de la plus grosse composante des réseaux calculés sur 1, 3 et 5 ans. Cette décroissance du diamètre suggère que les chercheurs sont plus interconnectés au sein du réseau et que par conséquent, la connaissance s'y transmet mieux et plus rapidement.

La Figure 3.29 présente le concept de centralisation défini à la section 1.3.1.4.

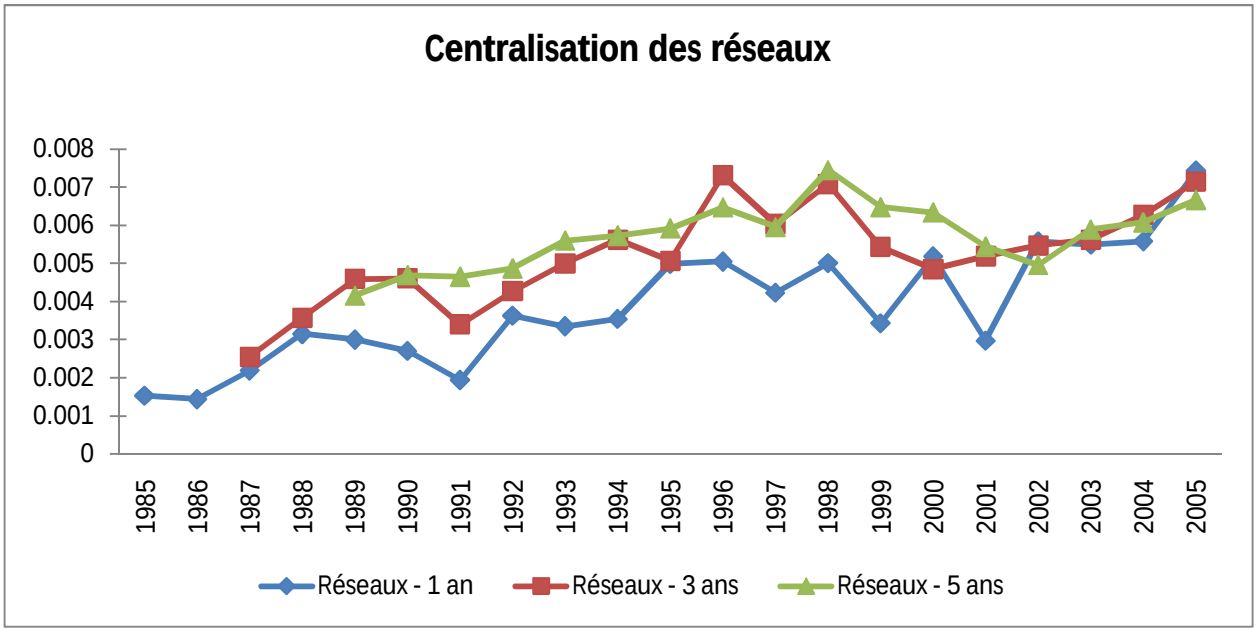


Figure 3.29 Centralisation des réseaux

La Figure 3.29 illustre l'évolution de la centralisation des réseaux. Cette mesure permet d'évaluer la relation entre le centre du réseau et la périphérie en termes de transfert d'information. Globalement pour l'ensemble des réseaux présentés, on note une tendance à la croissance de la centralisation malgré une évolution « en dents de scie », ce qui devrait avoir un impact positif sur l'échange d'informations au sein du réseau. Plus le noyau central d'un réseau est important, plus la transmission de la connaissance s'y fait rapidement, car le nombre d'intermédiaires diminue entre deux sommets.

Les mesures suivantes traitent de la cohésion des réseaux. Dans les réseaux non-dirigés, comme ceux étudiés, les composantes sont isolées les unes des autres, il n'existe pas de liens entre les sommets de différentes composantes[8].

La Figure 3.30 présente l'évolution du nombre de composantes présentes dans chacun des réseaux calculés sur des périodes d'un, trois et cinq ans.

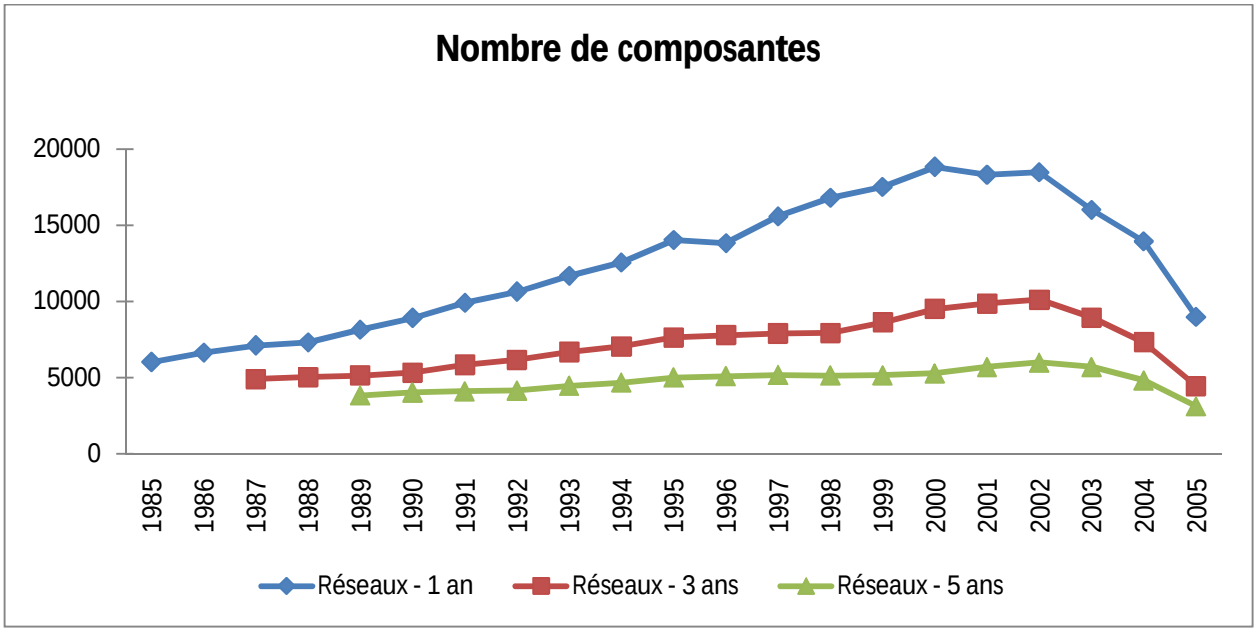


Figure 3.30 Nombre de composantes

À partir de la Figure 3.30 on observe le même comportement des courbes avec des amplitudes différentes. Il y a une croissance presque constante jusqu'en 2002 puis une décroissance du

[8] La commande [Net -> Components -> Weak (min size 1)] de Pajek nous permet de trouver l'ensemble des identificateurs de composantes liés aux sommets et d'en calculer les statistiques.

nombre de composantes à l'intérieur des réseaux. Nous observons donc une augmentation de la fragmentation du réseau jusqu'au tournant du siècle, puis un regroupement des composantes en réseaux plus imposants. Ce résultat est cohérant avec une augmentation du nombre de collaborateurs illustré à la Figure 3.15.

Les prochaines figures étudient l'évolution de la taille des composantes pour voir si une décroissance du nombre de singletons est expliquée par une croissance de la taille des autres composantes.

Pour les réseaux d'**un an**, on retrouve les trois figures suivantes :

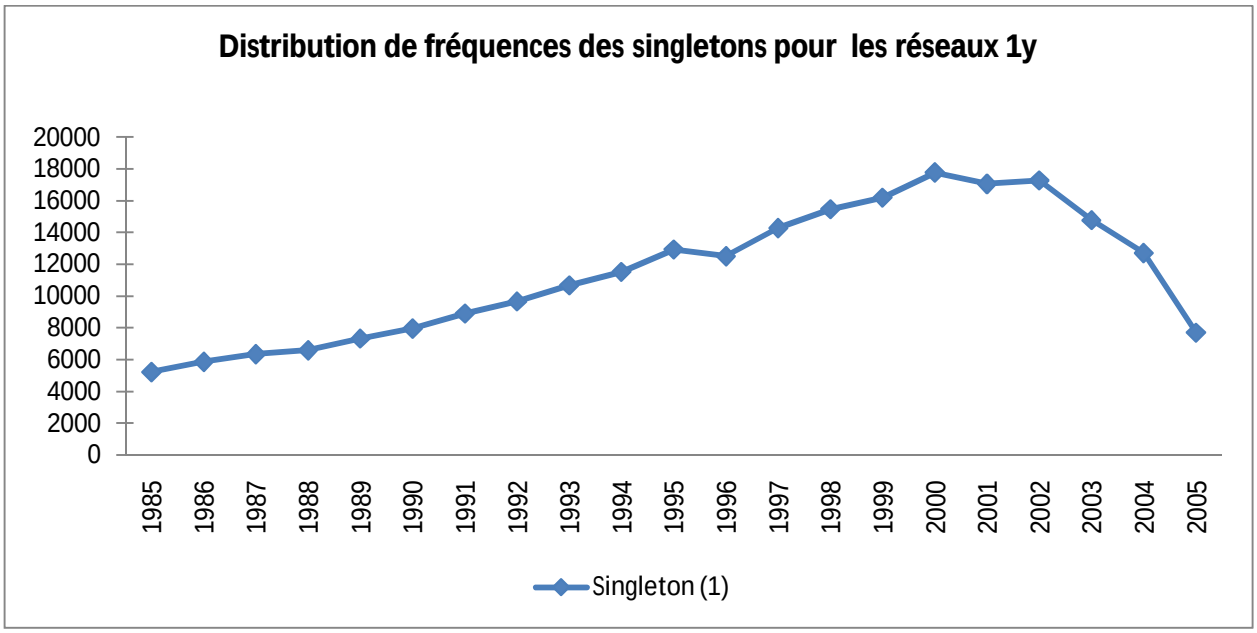


Figure 3.31 Distribution de fréquences des singletons pour les réseaux 1y

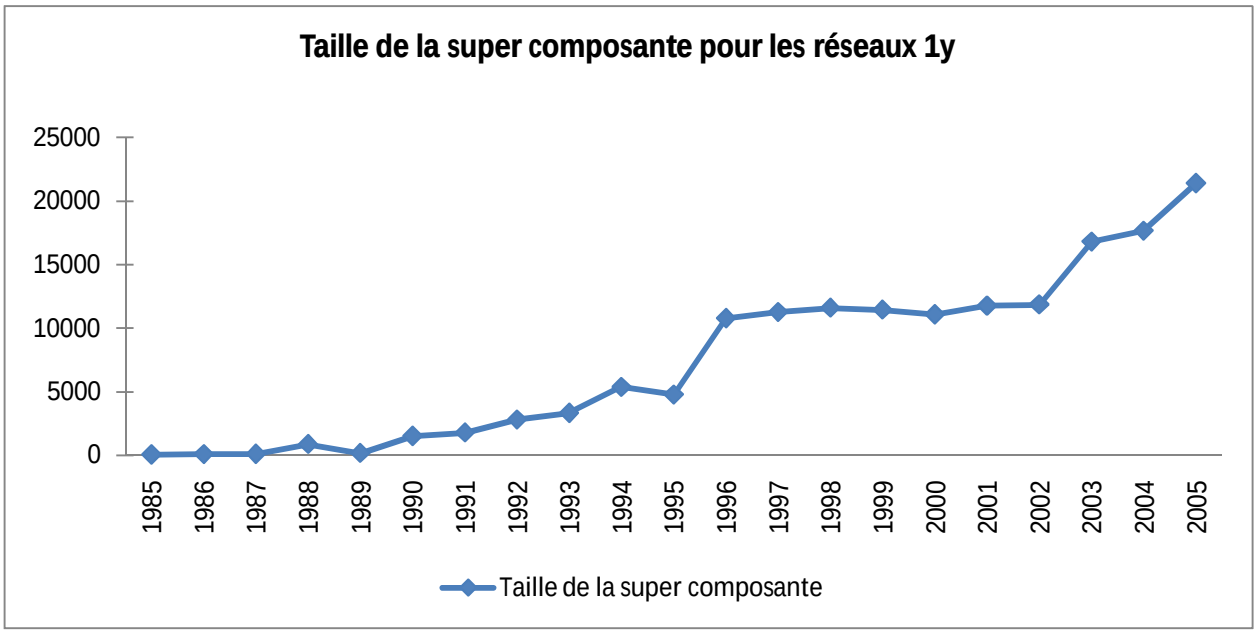


Figure 3.32 Taille de la super composante pour les réseaux 1y

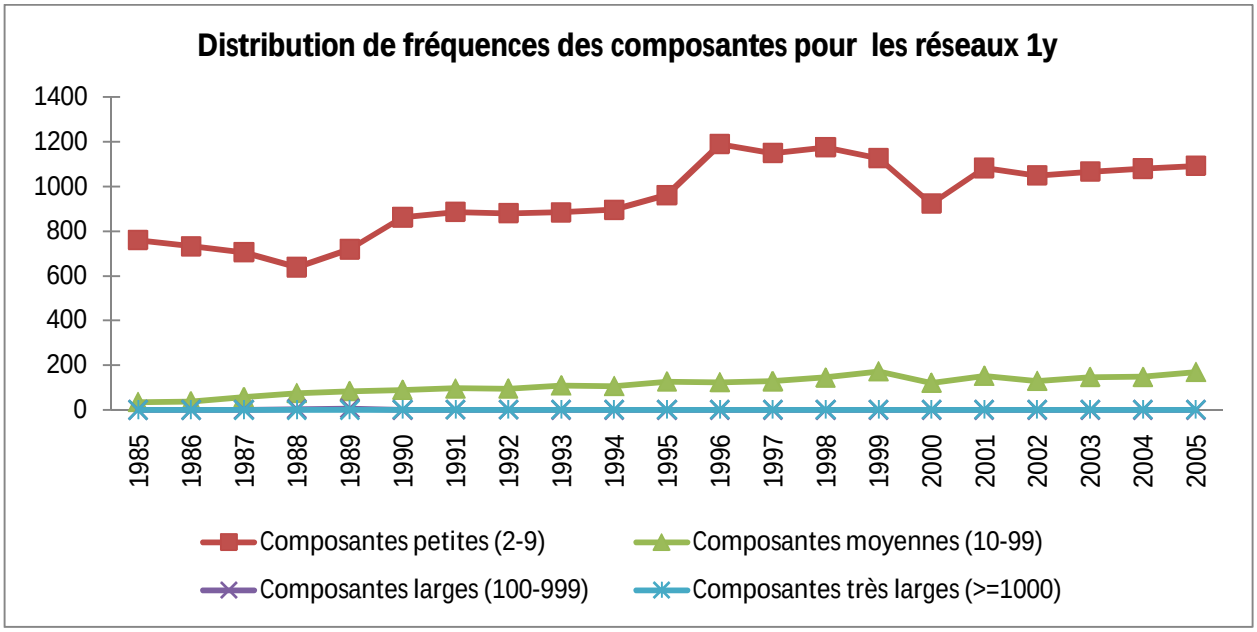


Figure 3.33 Distribution de fréquences des composantes pour les réseaux 1y

Pour les réseaux de **trois ans**, on retrouve les trois figures suivantes :

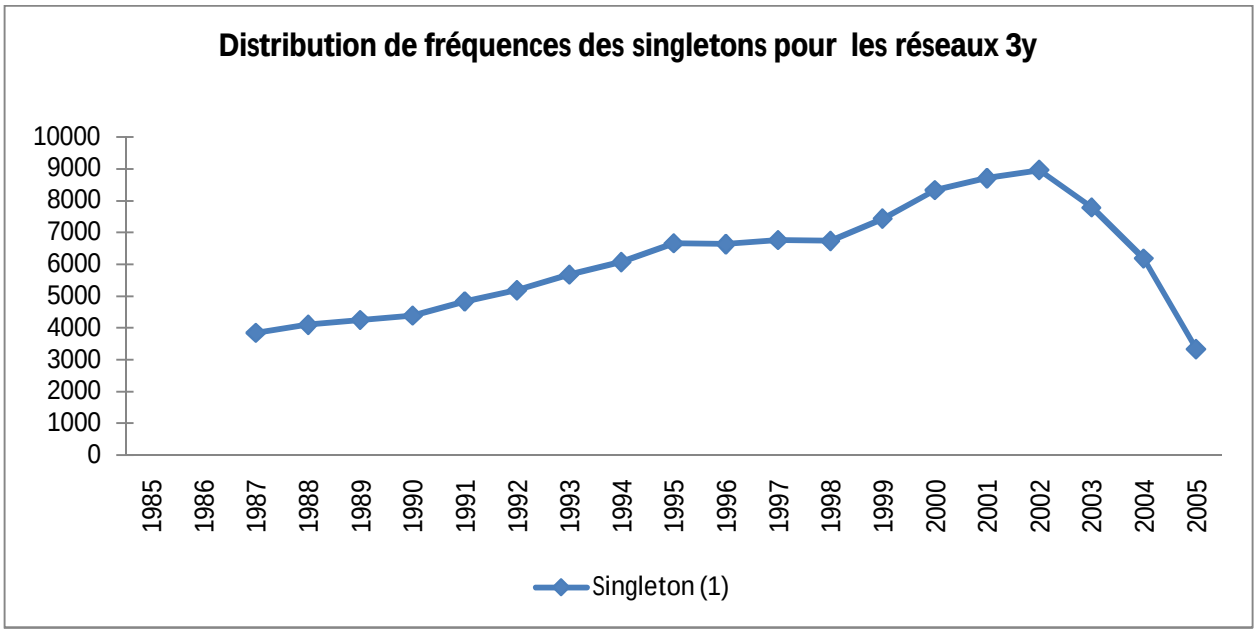


Figure 3.34 Distribution de fréquences des singletons pour les réseaux 3y

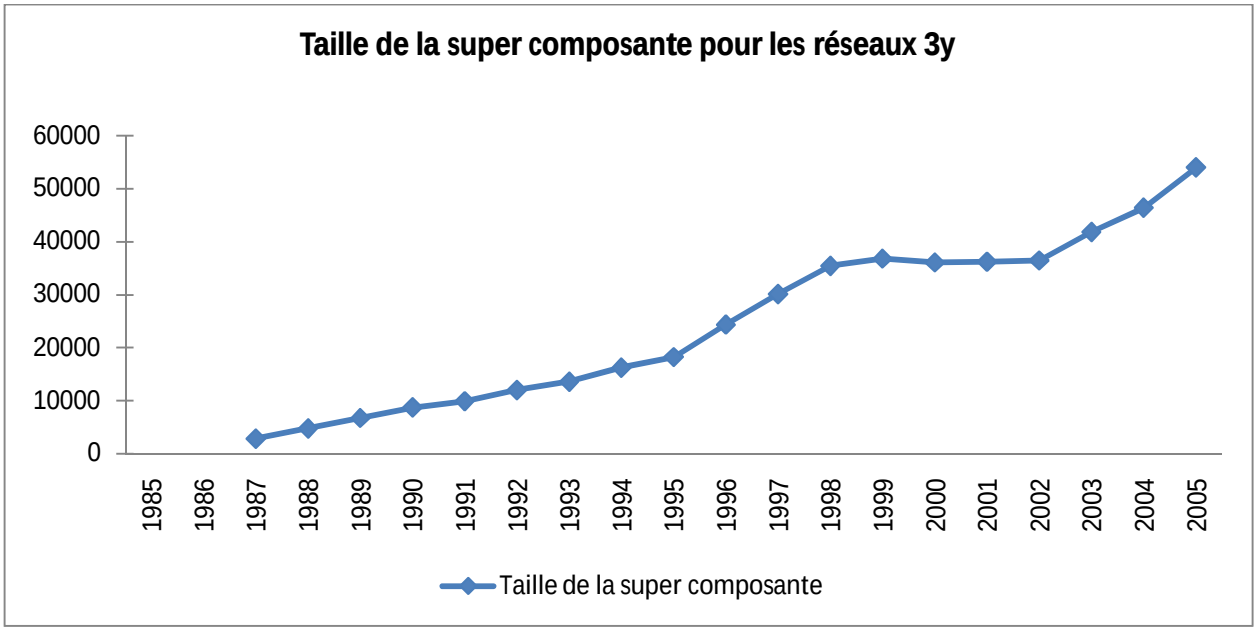


Figure 3.35 Taille de la super composante pour les réseaux 3y

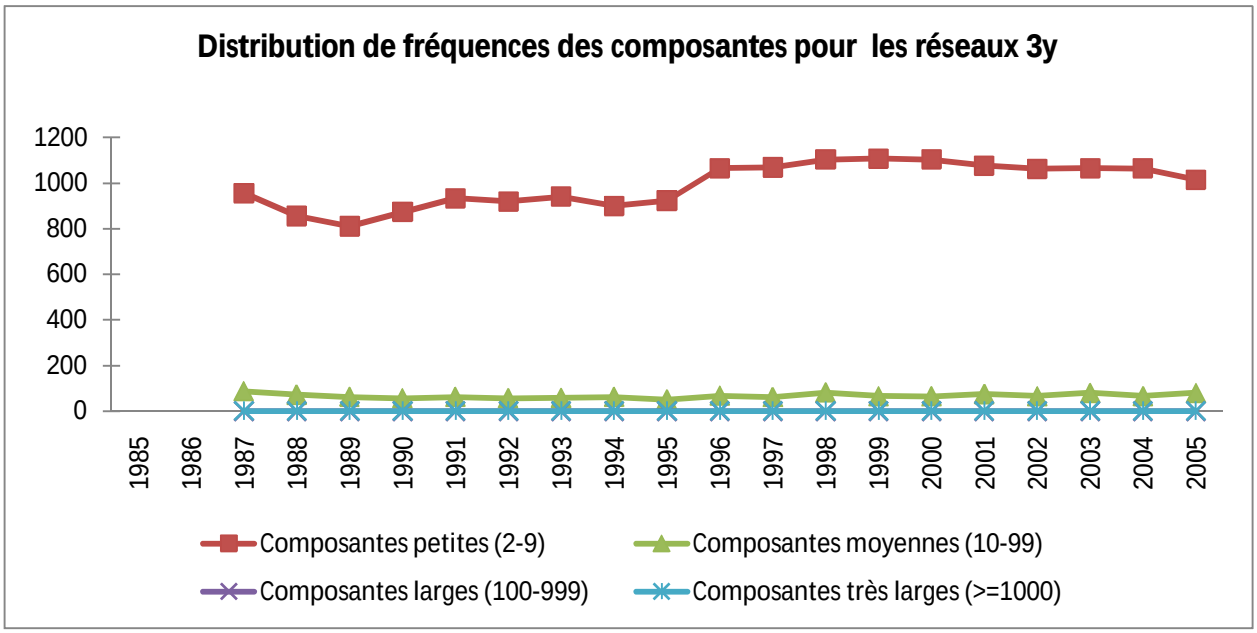


Figure 3.36 Distribution de fréquences des composantes pour les réseaux 3y

Pour les réseaux de **cinq ans**, on retrouve les trois figures suivantes :

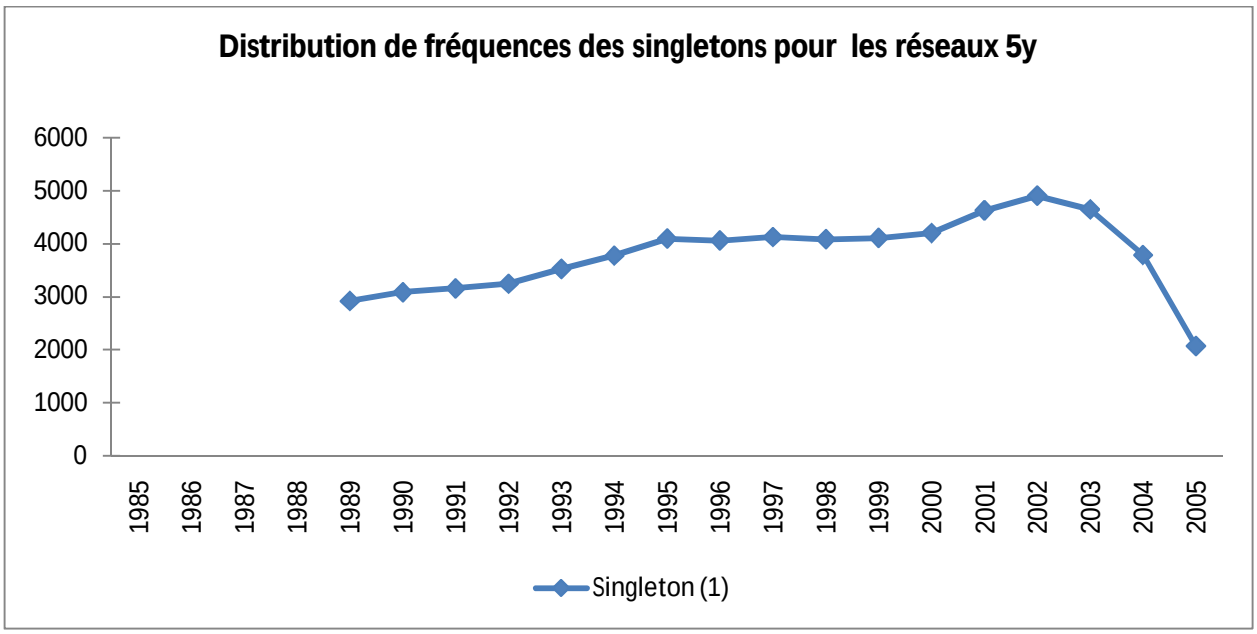


Figure 3.37 Distribution de fréquences des singletons pour les réseaux 5y

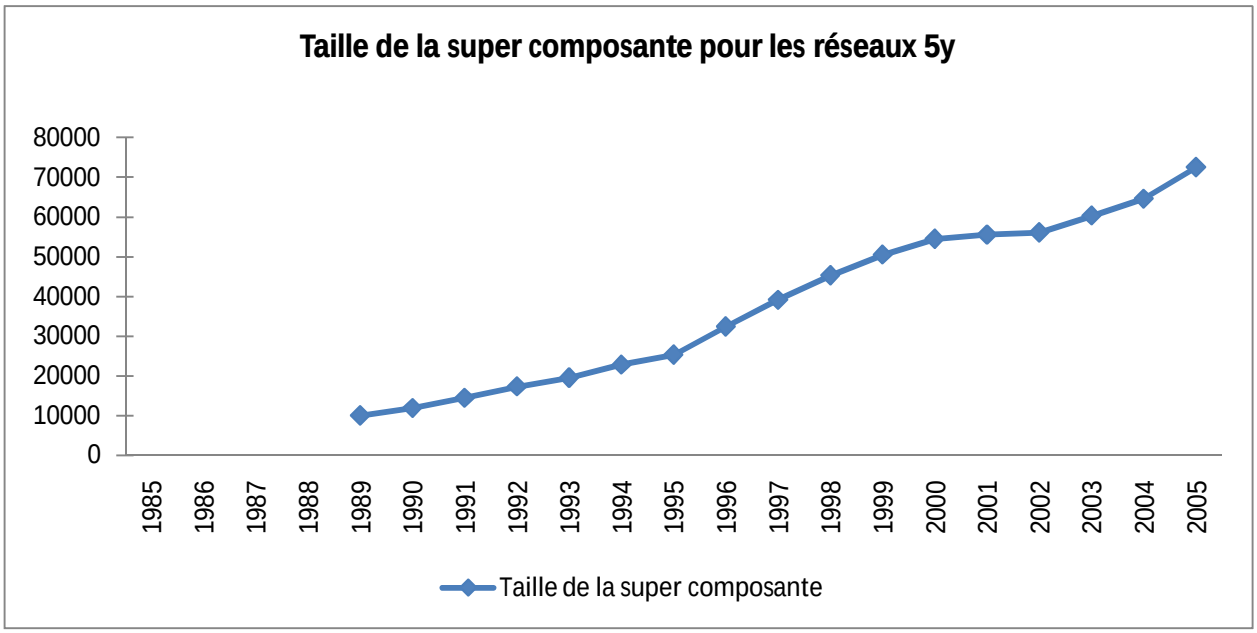


Figure 3.38 Taille de la super composante pour les réseaux 5y

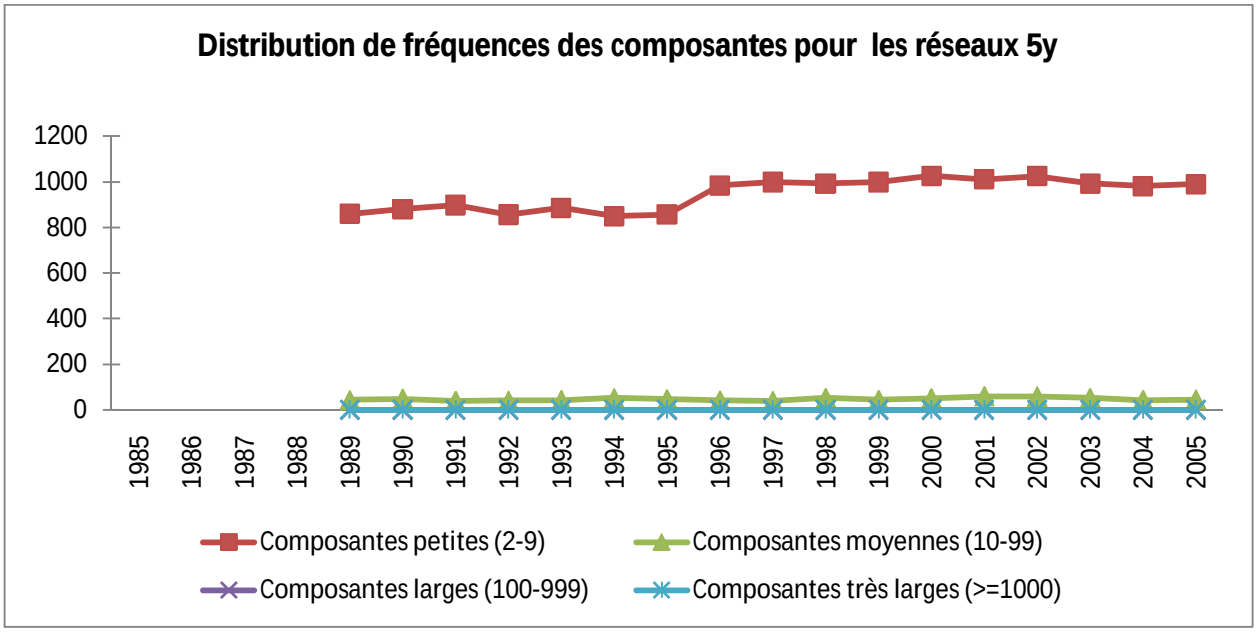


Figure 3.39 Distribution de fréquences des composantes pour les réseaux 5y

Les 9 dernières figures sont très importantes pour valider la tendance générale, observée jusqu'à présent, de regroupement des chercheurs pour la publication d'articles scientifiques en biotechnologie. Les figures présentant la distribution des fréquences des composantes illustrent donc des réseaux multi-centres avec un grand nombre de singletons et une très large composante.

Avec ces dernières figures, nous confirmons que, peu importe l'horizon d'analyse utilisé, le nombre de chercheurs publiant seuls diminue depuis la période 2000-2002 et qu'il y a de plus en plus de chercheurs qui sont connectés à travers une composante globale. On peut faire le lien entre l'augmentation presque constante de la « super » composante et une augmentation des publications scientifiques faites par plusieurs co-auteurs.

Les statistiques présentées dans ce chapitre permettent d'avoir un aperçu global de l'évolution temporelle des données et de certaines relations entre les variables du modèle. L'étape suivante, présentée dans le CHAPITRE 4, est donc l'analyse formelle des données à l'aide du modèle général et la présentation des résultats obtenus. Nous pouvons anticiper tout de suite que le modèle choisi devra avoir un horizon d'analyse plus grand qu'un an de façon à capturer l'effet essentiel de l'historique des réseaux.

CHAPITRE 4 ANALYSE ET RÉSULTATS

Ce chapitre a un double objectif. Premièrement, l'étude du comportement de plusieurs variations du modèle général permettra de faire émerger le modèle le plus robuste pouvant expliquer les données. Deuxièmement, de façon à répondre aux hypothèses posées à la section 2.1 l'étude des liens entre les variables du modèle permettra de quantifier l'impact du financement public sur la publication d'articles scientifiques.

Dans un premier temps, la section 4.1 détaillera comment l'analyse du modèle et des données seront effectuées. La section 4.1.1 présentera les variables d'analyses utilisées. La section 4.1.2 détaillera l'approche pour les analyses de régressions puis la section 4.1.3 traitera de l'analyse de corrélation sur l'ensemble des variables utilisées.

Dans un second temps, la partie 4.2 contiendra les résultats des analyses de régressions et de corrélation puis une discussion sur les résultats obtenus à la lumière de l'ensemble du projet.

4.1 Analyse des données

Ce chapitre analyse les données pour comprendre et modéliser les liens existants entre les paramètres du modèle. Avant de présenter les variables créées pour l'analyse, il est important de comprendre l'approche générale retenue qui est divisée en deux étapes. L'**analyse de régression** permet de trouver des relations entre les variables à l'aide de modèles avec des horizons de 1, 3 et 5 ans. De son côté, l'**analyse de corrélation** permet l'obtention d'une matrice de coefficients de corrélation dans le but d'évaluer l'association linéaire (la force du lien/le pouvoir explicatif) entre les variables des modèles retenus. Voyons maintenant comment les données tirées de la matrice générée, dont la génération a été détaillée à la section 2.4.4, seront utilisées pour l'analyse.

4.1.1 Variables d'analyse

Un ensemble de variables indépendantes a été crée pour décomposer et contrôler les effets possibles sur la production annuelle d'articles scientifiques. Ces variables sont liées aux réseaux construits sur 1, 3 et 5 ans. Le Tableau 4.1 résume les deux types de variables utilisées pour l'analyse, soit les variables nominales et les variables statistiques continues.

Tableau 4.1 Types de variables d'analyse

Types de variables	Description
Variable nominale	Variable qualitative dont la seule structure significative est la distinction des valeurs entre elles. (« dummy variable »)
Variable statistique continue	Variable statistique dont les valeurs possibles sont en nombre infini dans un intervalle donné. (« continuous variable »)

Différentes permutations de ces variables indépendantes constituent l'ensemble des modèles utilisés pour l'analyse. La création de modèles utilisant les différents types de variables comme paramètres a pour objectif la recherche de la *meilleure combinaison*. De plus, la robustesse du modèle final pourra être testée grâce aux variations de la présence des variables au sein de celui-ci.

Pour faciliter l'étude des différentes variations possibles du modèle général d'analyse, celui-ci sera décomposé en sept sections détaillées dans le Tableau 4.2. Chacune de ces sections représente un groupe de variables liées logiquement de façon à faciliter la présentation et la manipulation des données avec Stata 10. Les variables représentent les données brutes tirées de la matrice d'analyse sur lesquelles certaines transformations ont été appliquées pour faciliter leur analyse.

Tableau 4.2: Sections des modèles d'analyse

Sections du modèle d'analyse	Type de variables	Description
Mesures de subventions	Variables statistiques continues	· Nombre de subventions sur la période : **[nbsubv]** · Montant total des subventions sur la période : **[totsubv]** · Montant total moyen de subventions sur la période : **[totsubvmoy]** · Montant total moyen de subventions annuelles sur la période : **[totsubvmoyan]**
Mesures de contrats	Variables statistiques continues	· Nombre de contrats sur la période : **[nbcont]** · Montant total des contrats sur la période : **[totcont]** · Montant total moyen des contrats sur la période : **[totcontmoy]** · Montant total moyen des contrats annuels sur la période : **[totcontmoyan]** · Proportion des contrats par rapport aux subventions sur la période : **[propcont]**
Mesures de réseau	Variables statistiques continues	· Mesure de la centralité (« centrality ») : **[degcentind]** · Mesure de centralité de relation d'intermédiaire (« betweenness centrality ») : **[betcentind]** · Coefficient de groupement local (« local clustering coefficient ») : **[cliquind2]**
Universités	Variables nominales	Groupes comportant une ou plusieurs universités liées : **[d01Laval] [d02McGill] [d07Concordia] [d08SherbrookeG] [d09UQAMG]**
Années	Variables nominales	Années pour lesquelles ont veut contrôler : **[d1990] [d1991] [d1992] [d1993] [d1994] [d1995] …**
Contraintes pour la sélection des données	Ne s'applique pas	Voir Figure 4.6 pour les détails des contraintes de sélection des données.

L'utilisation des variables nominales est justifiée par le besoin de contrôler par les années et les universités d'attache des chercheurs. Dans ce cas, chacune des variables nominales est donc considérée indépendante des autres (Hilbe, 2007). Il y a donc autant de variables indépendantes qu'il y a de groupes d'universités et d'années.

La colonne « description » du Tableau 4.2 contient en plus de la définition de la variable, la terminologie utilisée pour faire référence aux variables lors des analyses de régression et de corrélation. Des préfixes et suffixes peuvent s'ajouter aux noms des variables de façon à préciser le contexte d'utilisation et des manipulations mathématiques effectuées pour les standardiser. Les chiffres en suffixe d'une variable **1**, **3** et **5** représentent l'horizon d'analyse. Le logarithme de la valeur d'une variable est représenté par le préfixe **l** tandis qu'une multiplication par 1000 de la valeur de la variable est représenté par le préfixe **m**. L'utilisation du logarithme naturel (ln) permet la réduction de l'échelle et la normalisation des valeurs.

La définition des variables d'analyse et leur raison d'être ont été posées, voyons maintenant en détail leur intégration au niveau des modèles utilisés pour les régressions.

4.1.2 Analyse de régression

Le choix du type de régression à effectuer sur les modèles est dépendant des données. Pour notre analyse, nous avons construit un ensemble de données de format panel (« panel data »). Le format des données est dit panel ou longitudinal, car l'ensemble de données suit un groupe d'individus à travers le temps (Hsiao, 2003). On retrouve donc plusieurs observations pour chacun des individus au cours de leur période d'activité en tant que chercheurs universitaires. L'objectif étant toujours de lier ces observations (lire variables) au nombre de publications scientifiques annuelles par chercheur. On parle donc d'un ensemble de données à deux dimensions (observations/temps).

Comme dans bien des ensembles de données réelles, il y a présence d'*hyperdispersion* qui sera validée à postériori dans la section 4.1.2.6. L'utilisation de la régression binomiale négative, plutôt que la régression de Poisson compense les problèmes d'hyperdispersion des données (Hilbe, 2007).

4.1.2.1 Conditions générales d'analyse des modèles

La construction de réseaux sur des périodes temporelles différentes (1-3-5 ans) et la recherche de relations entre les variables de ces périodes à travers différents modèles ont mené à la sélection d'une année de référence commune. Le processus de sélection de l'année 1989 comme référence est illustrée à la Figure 4.1. Pour l'année de référence, les données sont disponibles pour l'ensemble des réseaux d'un, trois et cinq ans.

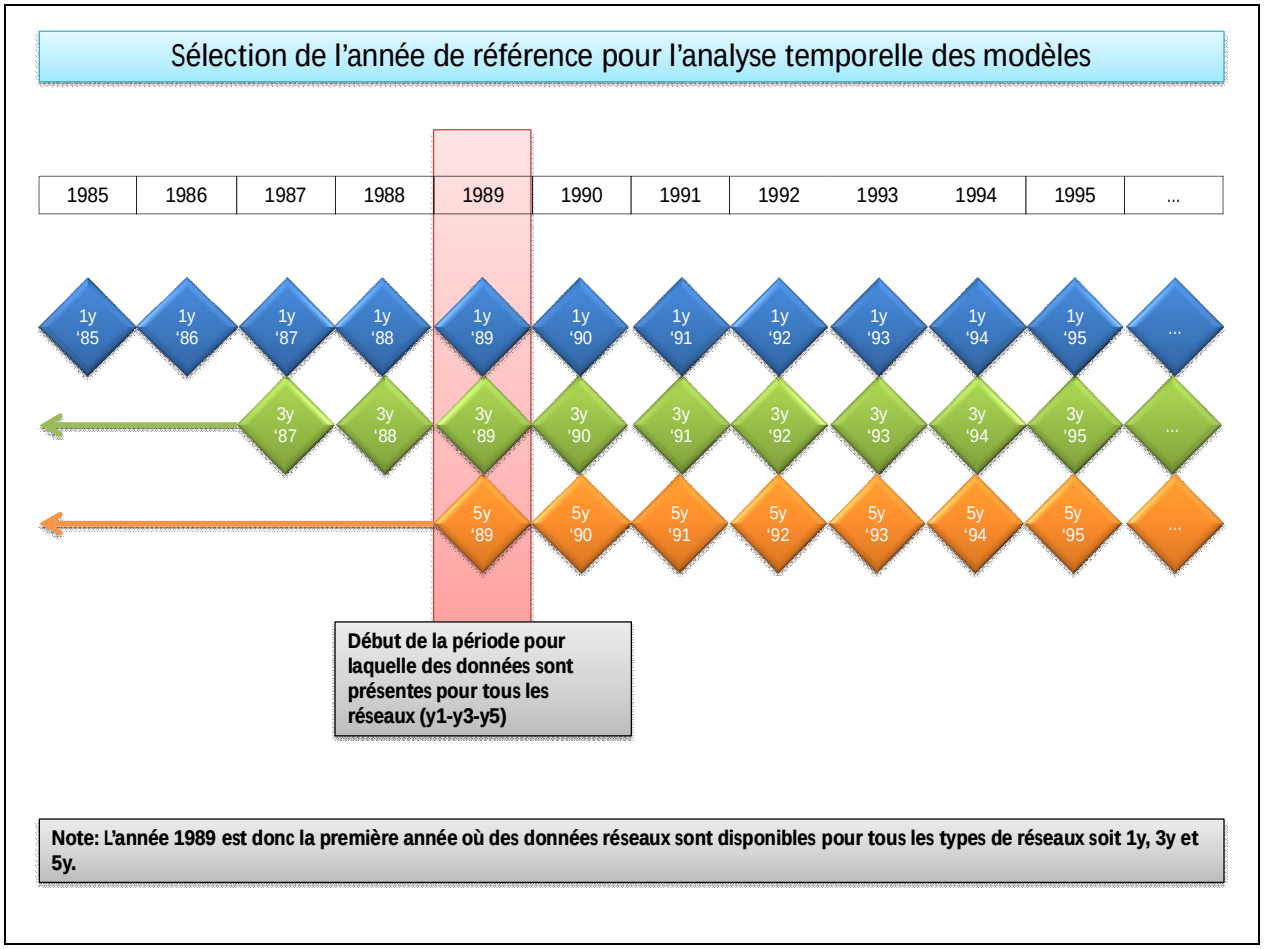


Figure 4.1 Processus de sélection de l'année de référence

La Figure 4.1 permet de visualiser le moment à partir duquel un réseau de cinq ans peut être introduit dans une régression à titre de variable représentant l'historique. Cette figure permet aussi de visualiser les trois types d'horizons de réseaux sélectionnés pour l'analyse.

Pour trouver le modèle ou la combinaison de modèles expliquant le mieux les données, différentes combinaisons de variables seront testées. Les modèles ainsi formés auront de légères

variations permettant de bien saisir la présence ou l'absence des différentes variables à l'aide des régressions. Cela représente 6 modèles différents pour les réseaux d'un an, et de 12 modèles pour les réseaux de 3 et de 5 ans. Le section 4.1.2.2 détaille les **variables statistiques continues** présentes dans chacun des modèles. On porte une attention particulière aux **variables nominales** d'années qui vont varier pour contrôler par les années en fonction de l'ensemble de données utilisé. Pour les variables nominales d'universités, comme on contrôle par l'université de Montréal/Polytechnique ont aura pour tous les modèles les variables suivantes : d01Laval, d02McGill, d07Concordia, d08SherbrookeG et d09UQAMG. Les trois sections suivantes présenteront la définition des modèles.

4.1.2.2 Description des modèles

Le Tableau 4.3 présente les 6 modèles pour les réseaux d'un an. Les variables évaluées sont liées aux mesures de subventions, aux mesures de contrats et aux mesures réseau.

Tableau 4.3 Modèles basés sur des réseaux d'un an

Modèles 1y	Mesures de subventions		Mesures de contrats		Mesures réseau		
Modèle B1	nbsubv0	ltotsubv	nbcont0	ltotcont	mdegcentindn1	mbetcentind1	lmcliquind21
Modèle B2	nbsubv0	ltotsubv	nbcont0	ltotcontmoy1	mdegcentindn1	mbetcentind1	lmcliquind21
Modèle B3	nbsubv0	ltotsubv	nbcont0	propcont	mdegcentindn1	mbetcentind1	lmcliquind21
Modèle B4	nbsubv0	ltotsubvmoy1	nbcont0	ltotcont	mdegcentindn1	mbetcentind1	lmcliquind21
Modèle B5	nbsubv0	ltotsubvmoy1	nbcont0	ltotcontmoy1	mdegcentindn1	mbetcentind1	lmcliquind21
Modèle B6	nbsubv0	ltotsubvmoy1	nbcont0	propcont	mdegcentindn1	mbetcentind1	lmcliquind21

En observant le Tableau 4.3, on note que les mesures de subventions et les mesures de contrats se divisent en une partie fixe et en une partie variable tandis que les mesures réseau sont fixes pour chacun des modèles. Les différents modèles décris représentent en fait des variations basées sur les variables de *financement*. Nous voulons trouver la combinaison de variables qui représentent le mieux les données. Ainsi, autant au niveau des contrats que des subventions, les modèles sont construits de façon à tester toutes les combinaisons du *total* et la *moyenne* des fonds reçus.

Le Tableau 4.4 présente les 12 modèles pour les réseaux sur trois ans. Les variables évaluées sont liées aux mesures de subventions, aux mesures de contrats et aux mesures réseau.

Tableau 4.4 Modèles basés sur des réseaux de trois ans

Modèles 3y	Mesures de subventions	Mesures de contrats	Mesures réseau
Modèle B1	nbsubv3 ltotsubv3	nbcont3 ltotcont3	mdegcentindn3 mbetcentind3 lmcliquind23
Modèle B2	nbsubv3 ltotsubv3	nbcont3 ltotcontmoy3	mdegcentindn3 mbetcentind3 lmcliquind23
Modèle B3	nbsubv3 ltotsubv3	nbcont3 ltotcontmoyan3	mdegcentindn3 mbetcentind3 lmcliquind23
Modèle B4	nbsubv3 ltotsubv3	nbcont3 propcont3	mdegcentindn3 mbetcentind3 lmcliquind23
Modèle B5	nbsubv3 ltotsubvmoy3	nbcont3 ltotcont3	mdegcentindn3 mbetcentind3 lmcliquind23
Modèle B6	nbsubv3 ltotsubvmoy3	nbcont3 ltotcontmoy3	mdegcentindn3 mbetcentind3 lmcliquind23
Modèle B7	nbsubv3 ltotsubvmoy3	nbcont3 ltotcontmoyan3	mdegcentindn3 mbetcentind3 lmcliquind23
Modèle B8	nbsubv3 ltotsubvmoy3	nbcont3 propcont3	mdegcentindn3 mbetcentind3 lmcliquind23
Modèle B9	nbsubv3 ltotsubvmoyan3	nbcont3 ltotcont3	mdegcentindn3 mbetcentind3 lmcliquind23
Modèle B10	nbsubv3 ltotsubvmoyan3	nbcont3 ltotcontmoy3	mdegcentindn3 mbetcentind3 lmcliquind23
Modèle B11	nbsubv3 ltotsubvmoyan3	nbcont3 ltotcontmoyan3	mdegcentindn3 mbetcentind3 lmcliquind23
Modèle B12	nbsubv3 ltotsubvmoyan3	nbcont3 propcont3	mdegcentindn3 mbetcentind3 lmcliquind23

Le Tableau 4.5 présente les 12 modèles pour les réseaux sur cinq ans. Les variables évaluées sont liées aux mesures de subventions, aux mesures de contrats et aux mesures réseau.

Tableau 4.5 Modèles basés sur des réseaux de cinq ans

Modèles 5y	Mesures de subventions	Mesures de contrats	Mesures réseau
Modèle B1	nbsubv5 ltotsubv5	nbcont5 ltotcont5	mdegcentindn5 mbetcentind5 lmcliquind25
Modèle B2	nbsubv5 ltotsubv5	nbcont5 ltotcontmoy5	mdegcentindn5 mbetcentind5 lmcliquind25
Modèle B3	nbsubv5 ltotsubv5	nbcont5 ltotcontmoyan5	mdegcentindn5 mbetcentind5 lmcliquind25
Modèle B4	nbsubv5 ltotsubv5	nbcont5 propcont5	mdegcentindn5 mbetcentind5 lmcliquind25
Modèle B5	nbsubv5 ltotsubvmoy5	nbcont5 ltotcont5	mdegcentindn5 mbetcentind5 lmcliquind25
Modèle B6	nbsubv5 ltotsubvmoy5	nbcont5 ltotcontmoy5	mdegcentindn5 mbetcentind5 lmcliquind25
Modèle B7	nbsubv5 ltotsubvmoy5	nbcont5 ltotcontmoyan5	mdegcentindn5 mbetcentind5 lmcliquind25
Modèle B8	nbsubv5 ltotsubvmoy5	nbcont5 propcont5	mdegcentindn5 mbetcentind5 lmcliquind25
Modèle B9	nbsubv5 ltotsubvmoyan5	nbcont5 ltotcont5	mdegcentindn5 mbetcentind5 lmcliquind25
Modèle B10	nbsubv5 ltotsubvmoyan5	nbcont5 ltotcontmoy5	mdegcentindn5 mbetcentind5 lmcliquind25
Modèle B11	nbsubv5 ltotsubvmoyan5	nbcont5 ltotcontmoyan5	mdegcentindn5 mbetcentind5 lmcliquind25
Modèle B12	nbsubv5 ltotsubvmoyan5	nbcont5 propcont5	mdegcentindn5 mbetcentind5 lmcliquind25

En observant Tableau 4.4 et le Tableau 4.5, on note encore une fois que les mesures de subventions et les mesures de contrats se divisent en une partie fixe et en une partie variable tandis que les mesures de caractéristiques de réseau sont fixes pour chacun des modèles. La même logique de variation du modèle s'applique ici, par contre, comme nous travaillons avec des réseaux ayant un horizon de 3 ans et 5 ans, une nouvelle variable s'ajoute : la *moyenne annuelle*

des fonds reçus. L'ajout de cette nouvelle variable explique qu'il y a le double des variations présentées au Tableau 4.3 soit 12 modèles.

Les modèles étant définis pour chacun des réseaux d'un, trois et cinq ans, il s'agit maintenant d'appliquer, à ceux-ci, les retards de phase.

4.1.2.3 Retards de phase appliqués aux modèles

La justification de l'utilisation des retards de phase, fait préalablement à la section 1.3, représente concrètement, au niveau des modèles d'analyse, des retards de phase (« lag ») d'une année (t-1) (L1) et de deux années (t-2) (L2).

La combinaison des retards de phases (L1 et L2) avec les différents modèles sont regroupés en 9 analyses. Les choix effectués pour construire ces 9 analyses se trouvent en Annexe 3 mais la Figure 4.2 présente les détails pour l'analyse #5 qui sera centrale à notre étude.

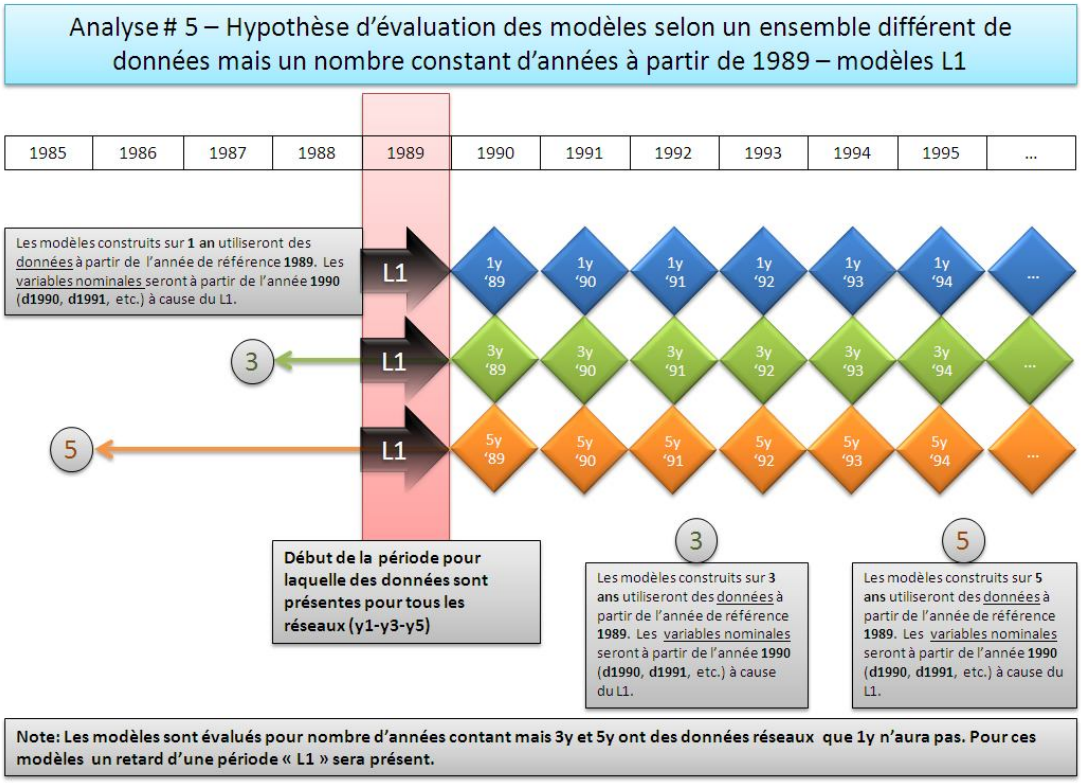


Figure 4.2 Analyse # 5

L'ensemble de ces analyses totalise **243** modèles, résumés par la Figure 4.4 et la Figure 4.5.

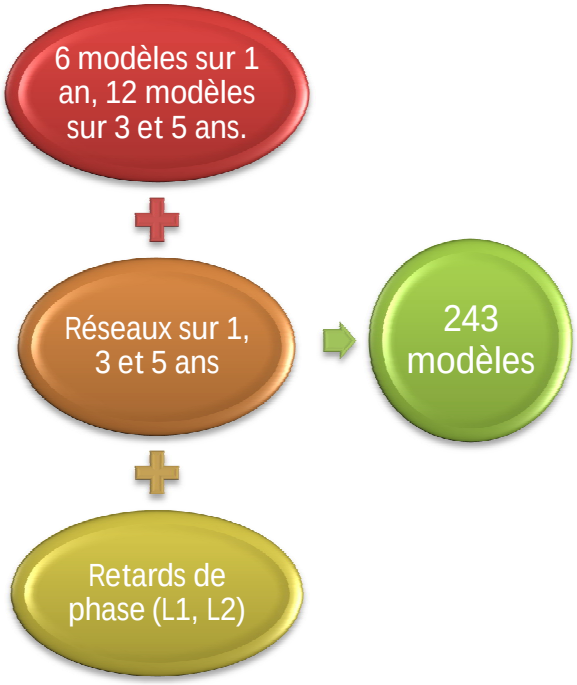


Figure 4.3 Construction des modèles

Celles-ci permettent de valider différentes combinaisons de modèles autant au niveau temporel qu'au niveau des données et des retards de phase.

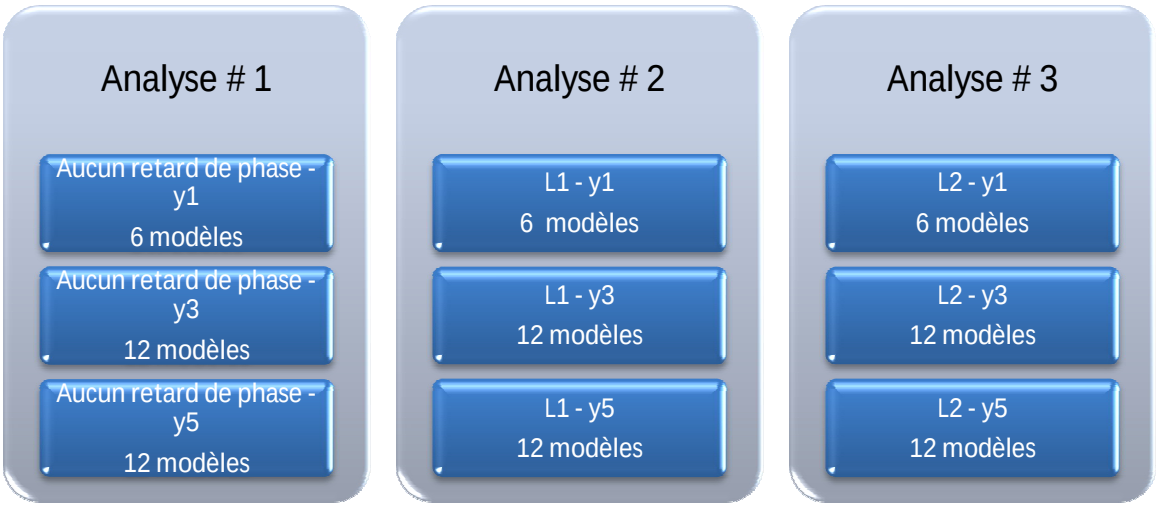


Figure 4.4 Premier groupe d'analyses des modèles

Les analyses #1, #2 et #3, illustrées à la Figure 4.4, représentent le **premier groupe d'analyses** et utilisent un *même ensemble* de données. Le fait de considérer un retard de phase de deux périodes plutôt qu'une seule de diminue donc pas le nombre d'observations.

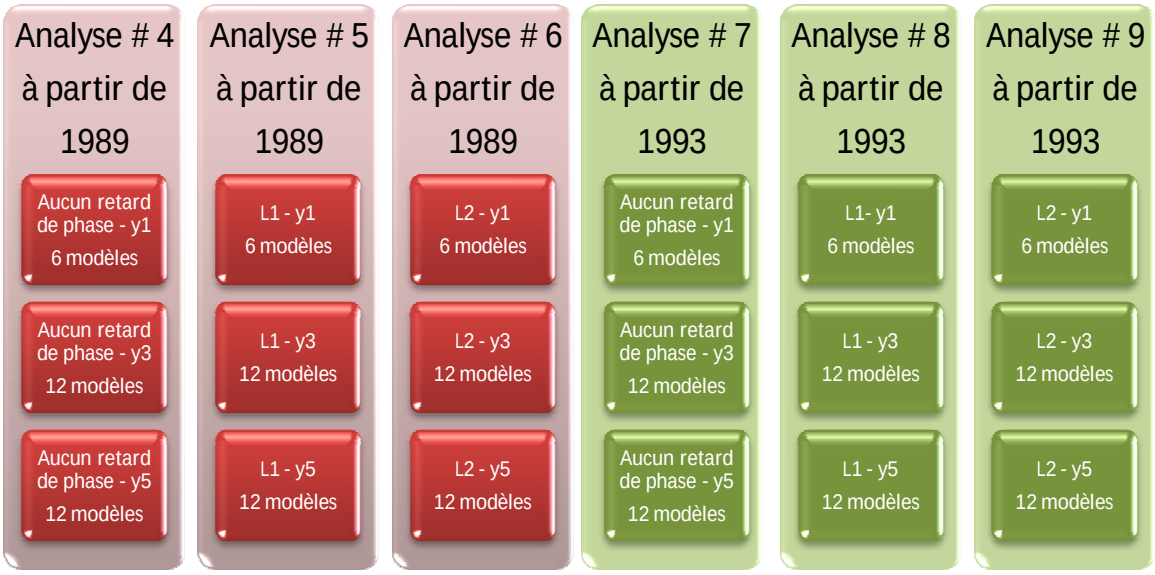


Figure 4.5 Deuxième et troisième groupes d'analyses des modèles

Les analyses #4, #5 et #6 illustrées à la Figure 4.5 représentent le **deuxième groupe** d'analyses et utilisent des données à partir de l'année de référence 1989. Les analyses #7, #8 et #9 toujours illustrées à la Figure 4.5 représentent le **troisième groupe** d'analyses et utilisent des données à partir de l'année de référence 1993. Le troisième groupe d'analyse va permettre de valider qu'il n'y a pas d'effet dû au début de la période de l'échantillon.

4.1.2.4 Présélection des modèles

Comme un nombre important de modèles sont étudiés, il est impératif de faire une première sélection basée sur une mesure fiable de comparaison des modèles. Les statistiques utilisées seront le critère d'information Akaike « Akaike information criterion » (AIC) et le critère d'information bayésien « Bayesian information criterion » **(**BIC). Ce sont deux mesures utilisées pour comparer les modèles entre eux. On parle donc d'outils de sélection de modèles statistiques à partir d'un ensemble de modèles potentiels. Les mesures AIC et le BIC sont des mesures populaires pour comparer des modèles de vraisemblance maximale (« maximum likelihood »). Les définitions des mesures AIC et de BIC sont présentées à l'Équation 4.1.

$$mesure = -2 * \ln(likelihood) + complexité$$

$$AIC = -2 * \ln(likelihood) + 2 * k$$

$$BIC = -2 * \ln(likelihood) + \ln(N) * k$$

***k* = degrés de liberté du modèle**

***N* = nombre d'observations**

Équation 4.1 Mesures AIC et BIC

Ces deux mesures combinent à la fois *l'ajustement* (« fit ») et la *complexité* du modèle. L'ajustement est mesuré négativement avec $-2 * \ln(likelihood)$ et une grande valeur signifie un mauvais ajustement. La complexité est mesurée positivement par $2 * k$ pour AIC et par $\ln(N) * k$ pour BIC. La différence entre AIC et BIC vient de la constante utilisée pour pondérer les degrés de liberté représenté par *k*. AIC utilise la constante 2 pour pondérer *k* alors que BIC utilise $\ln(N)$. Il est important de noter que la mesure de vraisemblance (« likelihood ») doit mesurer <u>les mêmes événements</u> pour permettre la comparaison des modèles à l'aide des mesures AIC/BIC.

Considérant différents modèles ajustés sur le même ensemble de données, le modèle possédant la plus petite valeur du AIC/BIC est considéré comme étant le meilleur. Ainsi pour chacune des analyses (#1 à #9) le même ensemble de données était utilisé et les degrés de liberté sont les mêmes, il est donc valide d'utiliser le AIC/BIC pour la sélection du modèle le plus approprié au sein d'un même sous-groupe d'analyse.

Une autre analyse contribuera à faire la présélection de modèles. L'analyse du test statistique z et le niveau de signification « p-value », qui sont liés au test de l'hypothèse nulle qu'un coefficient de la régression soit zéro en tenant compte que les autres variables sont présentes dans le modèle, permet d'évaluer que les variables sont significatives. Le test statistique *z* est le ratio du coefficient sur l'erreur standard tel que présenté à l'Équation 4.2 et il permet de représenter la « force » et le « sens » du coefficient.

$$z = \frac{Coef.}{Std.Err.}$$

Équation 4.2 Test statistique z

La valeur du « *p*-value » varie de 0 à 1. Une variable qui aura une valeur de probabilité supérieure à 0.1 du « *p*-value » sera dite **non significative**. Ainsi une valeur du « *p*-value » de 10% ou moins est le point de rejet de l'hypothèse nulle. Autrement dit, c'est avec une probabilité de 90% ou plus que la variable a un effet sur le modèle. Plus la valeur du « *p*-value » s'approche de 0, plus le résultat est **significatif**.

4.1.2.5 Analyse de régressions avec Stata

L'outil statistique Stata 10.0 est utilisé pour l'analyse de régressions et la commande qui permet d'estimer les coefficients de régressions est **xtnbreg**. La Figure 4.6 présente un exemple de régression d'un modèle binomial négatif en format Stata 10.

```
* Local variables
local lag "L2."
local clusters "d01Laval d02McGill d07Concordia d08SherbrookeG d09UQAMG"
local years "d1991 d1992 d1993 d1994 d1995 d1996 d1997 d1998 d1999 d2000 d2001 d2002 d2003 d2004 d2005"
local conditions  "if DA1 != 1 & DA2_4 != 1 & totsubvF0 > 0 & year >= 1989 & dpQc == 1"
local parameters "re irr"
local model "B1"

* Fits random-effects overdispersion models with parameters:
xtnbreg nbarticleB          /* Subventions - 1 AN */                                   ///
                            `lag'nbsubv0 `lag'ltotsubv                                  ///
                            /* Contrats - 1 AN */                                       ///
                            `lag'nbcont0 `lag'ltotcont                                  ///
                            /* Réseau individus - 1 AN */                               ///
                            `lag'mdegcentindn1 `lag'mbetcentind1 `lag'lmcliquind21      ///
                            /*Clusters universités */                                   ///
                            `clusters'                                                  ///
                            /* Années */                                                ///
                            `years'                                                     ///
                            /* Data selection conditions */                             ///
                            `conditions'                                                ///
                            /* Function (xtnbreg) parameters */                         ///
                            ,`parameters'
* Saves the current (active) estimation results under the current name
estimates store `model'

* Postestimation statistics - Obtain information criteria: AIC and BIC
estat ic
```

Figure 4.6 Exemple de régression d'un modèle binomial négatif en format Stata 10

Les conditions présentes dans le modèle de la Figure 4.6 nous assurent que :

1. Les chercheurs présents ont publié 5 articles ou plus au cours des 21 années examinées;
2. Les chercheurs présents ont reçu du financement;
3. Les données proviennent d'un sous-ensemble à partir d'une année spécifique;
4. Les chercheurs sont affiliés à un établissement québécois.

La section 4.2.1 présentera les résultats des régressions des modèles décrits dans cette section.

4.1.2.6 Détection à postériori de l'hyperdispersion

En suivant l'approche présentée dans (Hilbe, 2007), nous validons à postériori le choix de la modélisation des données par un modèle binomial négatif. Le modèle binomial négatif devrait compenser pour l'hyperdispersion de façon à minimiser le biais. En utilisant une modélisation Poisson à l'aide de la commande Stata **glm fam(poi)** pour le modèle final avec nos données binomiales négative cela permettra l'évaluation de la statistique de dispersion Pearson. Une valeur de cette statistique plus grande que 1.0 nous indique la présence d'hyperdispersion dans le modèle. La détection de l'hyperdispersion dans une modélisation Poisson des données confirme notre choix de l'utilisation d'un modèle plus général qui est la méthode de régression binomiale négative pour modéliser ces données. La section 4.2.1.3 présentera les résultats du test décrit dans cette section.

L'analyse des régressions nous permet donc d'évaluer les différents modèles pour en sélectionner un nombre restreint qui expliqueront le plus adéquatement les données. Par la suite, l'ensemble des variables utilisées dans les modèles sélectionnés seront analysées à l'aide d'une matrice de corrélation pour valider les relations existantes entre celles-ci. La section 4.1.3 décrira cette étape.

4.1.3 Analyse de corrélation

Pour obtenir un modèle réaliste et stable, les variables les plus corrélés ne doivent pas être utilisées. À partir d'un fichier de données Stata contenant les variables d'analyse, celles-ci sont analysés sommairement d'un point de vue statistique et, dans un second temps, la matrice des coefficients de corrélation est générée pour toutes les variables faisant parties des modèles sélectionnés.

4.1.3.1 Analyse de corrélations avec Stata

La commande **pwcorr** de Stata produit une matrice de coefficients de corrélation qui permet d'évaluer le pouvoir explicatif des variables indépendantes sur les variables dépendantes. Dans le cas de la commande *pwcorr* qui est utilisée pour notre analyse, les points sont retirés du calcul de corrélation uniquement s'il y a au moins une valeur manquante.

Quelques commentaires sur les valeurs présentes dans la matrice de corrélation:

- On mesure la corrélation avec un *coefficient de corrélation* (aussi appelé coefficient de corrélation Pearson);
- La mesure de corrélation indique la *force* et la *direction* de la relation linéaire entre deux variables, on peut aussi la voir comme la capacité à pouvoir prévoir la valeur de l'autre variable;
- Les valeurs possibles des coefficients varient de -1 à 1 [-1, +1];
- La corrélation entre une variable et elle-même est toujours 1;
- Une corrélation positive indique que lorsque la valeur du coefficient d'une variable augmente, l'autre fera de même;
- Il ne faut pas induire automatiquement une relation de cause à effet entre deux variables à partir d'un coefficient de corrélation;
- Le coefficient de corrélation ne peut pas généralement caractériser complètement la relation entre deux variables.

Une fois la matrice générée, il sera possible de valider la force de corrélation existante entre les variables sélectionnées et possiblement raffiner la sélection. La section 4.2.2 présentera les résultats de corrélations.

4.2 Résultats et discussion

Cette section présente les résultats des analyses décrites à la section 4.1 et discute des résultats.

4.2.1 Résultats des régressions

4.2.1.1 Choix du modèle

Un nombre important de modèles a été généré et tel que décrit à la section 4.1.2.4 une présélection s'effectue en comparant directement les mesures de AIC/BIC des modèles au sein d'un même sous-groupe. Les résultats complets de cette analyse sont présentés en Annexe 4.

Le Tableau 4.6 présente le résultat de la sélection du modèle possédant le BIC/AIC le plus petit pour chacune des analyses (#1 à #9) pour les réseaux d'un an.

Tableau 4.6 Statistiques AIC et BIC de sélection de modèles pour les réseaux d'un an

Analyse	*Model*	*Obs*	*ll(null)*	*ll(model)*	*df*	*AIC*	*BIC*
#1_no_lag_y1	(not concave) les algorithmes ne peuvent pas converger.						
#2_L1y1	B1	17100	.	-24708.2	31	49478.3	49718.5
#3_L2y1	B1	16696	.	-24190.8	30	48441.7	48673.4
#4_no_lag_y1	(not concave) les algorithmes ne peuvent pas converger.						
#5_L1y1	B1	17100	.	-24708.2	31	49478.3	49718.5
#6_L2y1	B1	16696	.	-24190.8	30	48441.7	48673.4
#7_no_lag_y1	(not concave) les algorithmes ne peuvent pas converger.						
#8_L1y1	B1	14465	.	-21489.1	27	43032.2	43236.8
#9_L2y1	B1	14266	.	-21167.5	26	42387	42583.7

Tableau 4.7 Statistiques AIC et BIC de sélection de modèles pour les réseaux de trois ans

Analyse	*Model*	*Obs*	*ll(null)*	*ll(model)*	*df*	*AIC*	*BIC*
#1_no_lag_y3	B11	16556	.	-22549.6	29	45157.3	45381
#2_L1y3	B11	16270	.	-23493.7	29	47045.4	47268.6
#3_L2y3	B11	15853	.	-22985.6	28	46027.1	46241.9
#4_no_lag_y3	B11	17803	.	-23910.8	31	47883.7	48125.1
#5_L1y3	B11	17426	.	-24890	31	49842.1	50082.8
#6_L2y3	B11	16909	.	-24287.2	30	48634.4	48866.5
#7_no_lag_y3	B11	14978	.	-20758.1	27	41570.2	41775.8
#8_L1y3	B11	14788	.	-21671.5	27	43396.9	43602.2
#9_L2y3	B11	14477	.	-21257.5	26	42567	42764.1

Tableau 4.8 Statistiques AIC et BIC de sélection de modèles pour les réseaux de cinq ans

Analyse	*Model*	*Obs*	*ll(null)*	*ll(model)*	*df*	*AIC*	*BIC*
#1_no_lag_y5	B11	15192	.	-21252.34	27	42558.68	42764.65
#2_L1y5	B11	14932	.	-21730.46	27	43514.93	43720.43
#3_L2y5	B11	14580	.	-21303.42	26	42658.83	42856.11
#4_no_lag_y5	B11	18019	.	-24533	31	49128	49369.78
#5_L1y5	B11	17041	.	-24337.84	30	48735.68	48967.98
#6_L2y5	B11	15958	.	-23041.19	29	46140.39	46363.04
#7_no_lag_y5	B11	15192	.	-21252.34	27	42558.68	42764.65
#8_L1y5	B11	14932	.	-21730.46	27	43514.93	43720.43
#9_L2y5	B11	14580	.	-21303.42	26	42658.83	42856.11

Trois derniers tableaux résument donc les 27 meilleurs modèles en se basant sur la valeur du AIC/BIC. Pour chacun des groupes d'analyses, le nombre d'observations est décroissant en fonction de la période de décalage (sans décalage, L1 et L2). Cette observation est cohérente avec le mécanisme de retard de période car on enlève une année à chacune des périodes enlevées. Les degrés de libertés qui ont la même valeur pour une même sous-analyse sont différents entre classes de sous-analyses comme l'illustre les trois derniers tableaux.

À la suite des régressions des différentes analyses (#1 à #9), le modèle **B1** pour les réseaux d'un an et le modèle **B11** pour les réseaux de trois et cinq ans sont les modèles qui représentent le mieux les données. Ces résultats, obtenus à l'aide de la comparaison des mesures AIC/BIC, sont résumés par ces trois tableaux. Ils représentent donc un *premier raffinement* de l'analyse quant au *choix du modèle* qui produit 24 modèles à analyser.

Un *deuxième raffinement* permet de diminuer la liste de modèles potentiels. Celui-ci consiste à sélectionner la valeur optimale pour nos analyses de la taille des fenêtres réseaux (1 an, 3 ans ou 5 ans) et du retard de période (sans retard, L1 ou L2) de façon à évaluer le plus précisément possible l'effet des variables indépendants sur la variable indépendante, soit le nombre d'articles.

Les réseaux calculés sur 1 an sont écartés dès le départ, car ils ne peuvent pas capturer l'effet de l'historique du financement avec assez de richesse. Un des points d'intérêt de cette étude est l'importance accordée à l'historique du financement passé pour expliquer les actions ou les résultats présents. Une fenêtre d'analyse d'un an pour les réseaux est simplement trop courte pour étudier l'impact du financement au niveau des publications. Par contre, les réseaux d'un an permettent d'illustrer plusieurs statistiques descriptives qui, elles, sont calculées sur une base annuelle. Les auteurs consultés et référencés à la section 1.3 travaillent normalement avec des fenêtres de temps de 3 ou 5 ans. De plus, tel qu'observé au Tableau 4.6 les algorithmes ne peuvent converger lorsqu'on tente d'évaluer les coefficients du modèle sur 1 an sans retard de phase. En dernier lieu, on peut aussi ajouter que les subventions sont souvent réparties sur plusieurs années tel que discuté à la section 1.3 donc en considérant les réseaux d'un an cela reviendrait à considérer du financement partiel des projets. En enlevant les réseaux d'un an de l'analyse, le modèle B1 est donc automatiquement éliminé de l'analyse. L'horizon d'analyse doit donc utiliser des fenêtres d'une durée de 3 ou 5 ans avec le modèle B11, cela représente encore 18 modèles potentiels. Les modèles utilisées dans les analyses #1, #2 et #3 ont été ajoutés pour permettre une comparaison avec un même ensemble de données mais cela a impliqué une modification au niveau des années de références et donc ces analyses seront utilisées pour tester la robustesse du modèle final. Il ne reste donc que 12 modèles potentiels à analyser.

Un *troisième raffinement* de la sélection des modèles est encore possible car la différence entre le 2e groupe d'analyse (les analyses #4, #5, et #6) et le 3e groupe d'analyse (les analyses #7, #8 et #9) provient de la quantité de données utilisées. C'est-à-dire que le 2e groupe a comme année de référence 1989 alors que le 3e groupe a comme année de référence 1993. Le 2e groupe possède plus de données donc, théoriquement, une historique plus riche. Le choix de l'année de référence 1989 pour le 2e groupe avait été illustré à la Figure 4.1 et est basé sur le fait que l'année 1989 est la première année où tous les réseaux ont des données complètes. Le choix de l'année de référence 1993 en tant que deuxième année de référence pour le 3e groupe est basé sur le fait que c'est la première année où toutes les données réseaux sont disponibles à partir de l'année 1989.

Les résultats sont présentés sous le format des ratios des taux d'incidences (« incidence-rate ratios ») de façon à faciliter leur analyse. On exprime donc les coefficients de régression en termes du log du taux de changement de la variable dépendante[9]. L'augmentation d'une unité d'une variable indépendante aura un impact croissant ou décroissant sur la variable dépendante en conservant les autres variables du modèle constantes. Par exemple, un contrat de plus au cours des trois dernières années augmentera d'un facteur de 1.002 le nombre d'articles scientifiques publiés aujourd'hui.

Les 6 tableaux suivants présentent la comparaison des analyses deux par deux (#4 & #7, #5 & #8 et # 6 & #9) pour les réseaux de 3 et 5 ans. L'objectif étant de valider qu'il est mieux de travailler avec un ensemble de données plus important donc d'utiliser les résultats des analyses #4, #5 et #6. L'absence de la constante dans la liste de variables provient du fait que ce n'est pas les coefficients qui sont affichés mais bien les ratios des taux d'incidences.

---

[9] On peut utiliser l'IRR pour la présentation du coefficient, car $\beta = \log(\mu_{x0+1} / \mu_{x0})$ où β est le coefficient de régression, μ est la valeur de la variable dépendante au temps où moment où la variable indépendante a été évaluée soit à $_{x0}$ et $_{x0+1}$ représentant un changement d'une unité de mesure.

Tableau 4.9 Comparaison des analyses #4 et #7 (aucun décalage) - 3 ans

	[Analyse #4 y3]			[Analyse #7 y3]		
	IRR		**S.E.**	**IRR**		**S.E.**
nbsubv3	1.0012270		(0.0011427)	1.0022720	*	(0.0012870)
ltotsubvmoyan3	1.0634490	****	(0.0093575)	1.0625460	****	(0.0102764)
nbcont3	1.0021730		(0.0017618)	1.0014010		(0.0018953)
ltotcontmoyan3	1.0034370	*	(0.0019855)	1.0040060	*	(0.0021621)
mdegcentindn3	1.9943000	****	(0.0587498)	1.9174640	****	(0.0657308)
mbetcentind3	0.9943594		(0.0077328)	1.0089710		(0.0096812)
lmcliquind23	1.3414190	****	(0.0153179)	1.3427950	****	(0.0176654)
d01Laval	1.0549920		(0.0348348)	1.0330980		(0.0363873)
d02McGill	1.0142060		(0.0321798)	1.0071890		(0.0339804)
d07Concordia	0.9204336		(0.0666768)	0.9211560		(0.0704924)
d08SherbrookeG	1.0490320		(0.0513861)	1.0286870		(0.0552348)
d09UQAMG	0.9954945		(0.0383815)	0.9816029		(0.0406576)
d1989						
d1990	1.0776290		(0.0643135)			
d1991	1.0593390		(0.0624345)			
d1992	1.2592930	****	(0.0725868)			
d1993	1.4087370	****	(0.0798291)			
d1994	1.6309010	****	(0.0894369)	1.1565590	***	(0.0523949)
d1995	1.5638090	****	(0.0851375)	1.1038870	**	(0.0496719)
d1996	2.4122180	****	(0.1271198)	1.6973090	****	(0.0726605)
d1997	2.2659230	****	(0.1188308)	1.5923630	****	(0.0676148)
d1998	2.2098100	****	(0.1173587)	1.5521650	****	(0.0671040)
d1999	2.0751660	****	(0.1098498)	1.4544110	****	(0.0624378)
d2000	1.7308320	****	(0.0931329)	1.2175440	****	(0.0535721)
d2001	1.9748920	****	(0.1045746)	1.3864970	****	(0.0595578)
d2002	1.7515470	****	(0.0938909)	1.2295900	****	(0.0538555)
d2003	2.0815320	****	(0.1099931)	1.4574930	****	(0.0624901)
d2004	2.0236720	****	(0.1080304)	1.4164780	****	(0.0617523)
d2005	2.2629340	****	(0.1194267)	1.5852830	****	(0.0683122)

L'erreur standard (S. E.) est entre parenthèses

* $p < 0.1$

** $p < 0.05$

*** $p < 0.01$

**** $p < 0.001$

Tableau 4.10 Comparaison des analyses #5 et #8 (décalage un an) - 3 ans

	[Analyse #5 y3]			[Analyse #8 y3]		
	IRR		**S.E.**	**IRR**		**S.E.**
nbsubv3	1.0072570	****	(0.0011204)	1.0079150	****	(0.0012593)
ltotsubvmoyan3	1.0101640	***	(0.0034247)	1.0086400	**	(0.0036638)
nbcont3	1.0049390	**	(0.0019648)	1.0045550	**	(0.0020841)
ltotcontmoyan3	1.0050740	**	(0.0022470)	1.0067470	***	(0.0023989)
mdegcentindn3	1.4075540	****	(0.0476771)	1.3670610	****	(0.0523556)
mbetcentind3	1.0145570	*	(0.0088099)	1.0212980	**	(0.0102198)
lmcliquind23	1.0308630	***	(0.0098719)	1.0147500		(0.0111345)
d01Laval	1.0365570		(0.0394897)	1.0122550		(0.0410432)
d02McGill	1.0069150		(0.0367761)	0.9979058		(0.0386925)
d07Concordia	0.9293707		(0.0780449)	0.9806471		(0.0868071)
d08SherbrookeG	1.0130010		(0.0566632)	0.9748109		(0.0593128)
d09UQAMG	0.9311812		(0.0411764)	0.9235828	*	(0.0437221)
d1989						
d1990	1.0546040		(0.0741983)			
d1991	1.0457220		(0.0718718)			
d1992	1.1276090	*	(0.0757816)			
d1993	1.1962660	***	(0.0788423)			
d1994	1.3517880	****	(0.0863045)	1.1235340	**	(0.0576543)
d1995	1.3451850	****	(0.0851983)	1.1122330	**	(0.0563937)
d1996	1.8173720	****	(0.1106255)	1.4992950	****	(0.0714543)
d1997	1.8240030	****	(0.1109744)	1.4955300	****	(0.0713839)
d1998	1.6633490	****	(0.1014212)	1.3639580	****	(0.0651872)
d1999	1.6641810	****	(0.1017366)	1.3586860	****	(0.0652336)
d2000	1.4232260	****	(0.0883815)	1.1619340	***	(0.0570725)
d2001	1.6061470	****	(0.0981422)	1.3146850	****	(0.0629714)
d2002	1.4238190	****	(0.0880968)	1.1630700	***	(0.0567895)
d2003	1.6347910	****	(0.0997232)	1.3359570	****	(0.0638367)
d2004	1.5620040	****	(0.0960081)	1.2730320	****	(0.0616082)
d2005	1.8302110	****	(0.1110646)	1.4864660	****	(0.0704921)

L'erreur standard (S. E.) est entre parenthèses

* $p < 0.1$

** $p < 0.05$

*** $p < 0.01$

**** $p < 0.001$

Tableau 4.11 Comparaison des analyses #6 et #9 (décalage deux ans) - 3 ans

	[Analyse #6 y3]			[Analyse #9 y3]		
	IRR		**S.E.**	**IRR**		**S.E.**
nbsubv3	1.0067670	****	(0.0011862)	1.0070420	****	0.0012911
ltotsubvmoyan3	1.0124500	****	(0.0029200)	1.0097420	***	0.0030941
nbcont3	1.0054950	***	(0.0020802)	1.0051940	**	0.0021713
ltotcontmoyan3	1.0044700	*	(0.0023127)	1.0056280	**	0.0024431
mdegcentindn3	1.3226790	****	(0.0467558)	1.3316270	****	0.053023
mbetcentind3	1.0104190		(0.0091809)	1.0170310		0.0105919
lmcliquind23	1.0350200	****	(0.0094990)	1.0241060	**	0.0107256
d01Laval	1.0365270		(0.0408721)	1.0142020		0.0421083
d02McGill	1.0106840		(0.0382893)	1.0023560		0.0399123
d07Concordia	0.9481968		(0.0841848)	1.0063890		0.0925452
d08SherbrookeG	1.0181560		(0.0586626)	0.9802668		0.060908
d09UQAMG	0.9362634		(0.0425960)	0.9319737		0.0449491
d1989						
d1990						
d1991	1.0197450		(0.0594919)			
d1992	1.0495650		(0.0591909)			
d1993	1.1269900	**	(0.0613722)			
d1994	1.2534480	****	(0.0650462)			
d1995	1.2607660	****	(0.0645590)	1.0472180		0.0445798
d1996	1.7211300	****	(0.0826730)	1.4212740	****	0.0550344
d1997	1.6636720	****	(0.0796950)	1.3726300	****	0.0530719
d1998	1.5730460	****	(0.0760028)	1.2914030	****	0.0507108
d1999	1.5254620	****	(0.0734517)	1.2517660	****	0.0487882
d2000	1.3013820	****	(0.0646930)	1.0667670		0.0436567
d2001	1.4604390	****	(0.0705978)	1.1977540	****	0.0468759
d2002	1.2918520	****	(0.0636603)	1.0614760		0.0426716
d2003	1.4666620	****	(0.0708384)	1.2070760	****	0.0470795
d2004	1.3956450	****	(0.0681929)	1.1469540	***	0.0455565
d2005	1.6227550	****	(0.0777169)	1.3279040	****	0.0511786

L'erreur standard (S. E.) est entre parenthèses

* $p < 0.1$

** $p < 0.05$

*** $p < 0.01$

**** $p < 0.001$

Tableau 4.12 Comparaison des analyses #4 et #7 (aucun décalage) - 5 ans

	[Analyse #4 y5]			[Analyse #7 y5]		
	IRR		**S.E.**	**IRR**		**S.E.**
nbsubv5	1.0010160		(0.0008211)	1.0012060		(0.0009062)
ltotsubvmoyan5	1.0629000	****	(0.0099239)	1.0636650	****	(0.0108847)
nbcont5	1.0019230		(0.0012390)	1.0011100		(0.0013298)
ltotcontmoyan5	1.0011670		(0.0021053)	1.0026050		(0.0022871)
mdegcentindn5	2.7287900	****	(0.0895619)	2.8978850	****	(0.1146510)
mbetcentind5	0.9380578	****	(0.0077147)	0.9277313	****	(0.0099082)
lmcliquind25	1.1603720	****	(0.0158966)	1.1192080	****	(0.0179694)
d01Laval	1.0609110	*	(0.0355600)	1.0334930		(0.0368497)
d02McGill	1.0034770		(0.0324791)	0.9945859		(0.0342186)
d07Concordia	0.9433919		(0.0694770)	0.9652240		(0.0750755)
d08SherbrookeG	1.0384380		(0.0517088)	1.0244640		(0.0557085)
d09UQAMG	0.9956698		(0.0389683)	0.9833692		(0.0412379)
d1989	0.9527180		(0.0572766)			
d1990						
d1991	0.9706936		(0.0541127)			
d1992	1.0342450		(0.0562914)			
d1993	1.1406800	**	(0.0610443)			
d1994	1.2953330	****	(0.0667744)	1.1365630	***	(0.0516082)
d1995	1.2747120	****	(0.0655031)	1.1139260	**	(0.0502467)
d1996	1.7911550	****	(0.0884143)	1.5507250	****	(0.0661160)
d1997	1.7668350	****	(0.0869028)	1.5295460	****	(0.0650959)
d1998	1.6851100	****	(0.0846265)	1.4440380	****	(0.0628758)
d1999	1.6682040	****	(0.0834729)	1.4329910	****	(0.0621353)
d2000	1.4181020	****	(0.0726473)	1.2159920	****	(0.0542932)
d2001	1.5487030	****	(0.0775266)	1.3309170	****	(0.0577293)
d2002	1.3544550	****	(0.0685324)	1.1674600	****	(0.0514699)
d2003	1.5974490	****	(0.0798966)	1.3690540	****	(0.0593215)
d2004	1.5009520	****	(0.0760578)	1.2794730	****	(0.0563744)
d2005	1.6630980	****	(0.0833778)	1.4108310	****	(0.0613256)

L'erreur standard (S. E.) est entre parenthèses

* $p < 0.1$

** $p < 0.05$

*** $p < 0.01$

**** $p < 0.001$

Tableau 4.13 Comparaison des analyses #5 et #8 (décalage un an) - 5 ans

	[Analyse #5 y5]			[Analyse #8 y5]		
	IRR		**S.E.**	**IRR**		**S.E.**
nbsubv5	1.0046190	****	(0.0008127)	1.0048670	****	(0.0008672)
ltotsubvmoyan5	1.0135080	****	(0.0037756)	1.0114150	***	(0.0039608)
nbcont5	1.0032490	**	(0.0014133)	1.0027190	*	(0.0014824)
ltotcontmoyan5	1.0036690		(0.0023806)	1.0052580	**	(0.0025217)
mdegcentindn5	1.5873630	****	(0.0600111)	1.6120290	****	(0.0696880)
mbetcentind5	0.9956794		(0.0092441)	0.9959041		(0.0112571)
lmcliquind25	1.0008140		(0.0126671)	0.9789086		(0.0141485)
d01Laval	1.0360310		(0.0396976)	1.0183270		(0.0408235)
d02McGill	0.9983843		(0.0366859)	0.9972994		(0.0383109)
d07Concordia	0.9424181		(0.0791636)	0.9773600		(0.0852618)
d08SherbrookeG	0.9951715		(0.0562482)	0.9767376		(0.0587612)
d09UQAMG	0.9256592	*	(0.0411114)	0.9308822		(0.0435201)
d1989						
d1990						
d1991	0.9879913		(0.0627034)			
d1992	1.0333090		(0.0639799)			
d1993	1.0690110		(0.0648388)			
d1994	1.2105540	***	(0.0707669)	1.1264760	**	(0.0574708)
d1995	1.1989370	***	(0.0695564)	1.1113310	**	(0.0560961)
d1996	1.6186120	****	(0.0899051)	1.4931590	****	(0.0710100)
d1997	1.6072050	****	(0.0893188)	1.4745550	****	(0.0702444)
d1998	1.4709890	****	(0.0819643)	1.3499070	****	(0.0645340)
d1999	1.4626130	****	(0.0822421)	1.3337630	****	(0.0644557)
d2000	1.2598880	****	(0.0720323)	1.1486050	***	(0.0567809)
d2001	1.4058960	****	(0.0792469)	1.2806510	****	(0.0620027)
d2002	1.2441640	****	(0.0709051)	1.1346770	**	(0.0557507)
d2003	1.4274970	****	(0.0798737)	1.3068070	****	(0.0626786)
d2004	1.3500560	****	(0.0766293)	1.2292340	****	(0.0600714)
d2005	1.5547960	****	(0.0870142)	1.4121660	****	(0.0676544)

L'erreur standard (S. E.) est entre parenthèses

* $p < 0.1$

** $p < 0.05$

*** $p < 0.01$

**** $p < 0.001$

Tableau 4.14 Comparaison des analyses #6 et #9 (décalage deux ans) - 5 ans

	[Analyse #6 y5]			[Analyse #9 y5]		
	IRR		**S.E.**	**IRR**		**S.E.**
nbsubv5	1.0043230	****	(0.0008849)	1.0044190	****	(0.0009236)
ltotsubvmoyan5	1.0112980	****	(0.0031661)	1.0098380	***	(0.0032794)
nbcont5	1.0031960	**	(0.0015144)	1.0028240	*	(0.0015574)
ltotcontmoyan5	1.0041760	*	(0.0024907)	1.0047370	*	(0.0025889)
mdegcentindn5	1.4302970	****	(0.0557407)	1.4713300	****	(0.0632544)
mbetcentind5	1.0098460		(0.0096914)	1.0066360		(0.0112972)
lmcliquind25	1.0082480		(0.0126907)	0.9954289		(0.0137126)
d01Laval	1.0395420		(0.0417314)	1.0153050		(0.0419340)
d02McGill	1.0103420		(0.0389408)	1.0025140		(0.0397687)
d07Concordia	0.9645132		(0.0856405)	0.9863575		(0.0895515)
d08SherbrookeG	0.9802575		(0.0579890)	0.9719117		(0.0599369)
d09UQAMG	0.9227493	*	(0.0427798)	0.9337534		(0.0448499)
d1989						
d1990						
d1991	0.9689806		(0.0590606)			
d1992						
d1993	1.0477900		(0.0596352)			
d1994	1.1491880	**	(0.0627455)			
d1995	1.1498670	**	(0.0622160)	1.0384700		(0.0440653)
d1996	1.5627330	****	(0.0799716)	1.4061720	****	(0.0543573)
d1997	1.5063720	****	(0.0771782)	1.3553760	****	(0.0525071)
d1998	1.4054190	****	(0.0726529)	1.2610540	****	(0.0495342)
d1999	1.3726520	****	(0.0707393)	1.2315830	****	(0.0481802)
d2000	1.1712850	***	(0.0624513)	1.0468580		(0.0433263)
d2001	1.3177210	****	(0.0686060)	1.1785100	****	(0.0468225)
d2002	1.1622240	***	(0.0617391)	1.0387680		(0.0426356)
d2003	1.3177510	****	(0.0686046)	1.1795920	****	(0.0467377)
d2004	1.2513850	****	(0.0656777)	1.1213590	***	(0.0450518)
d2005	1.4450450	****	(0.0748796)	1.2905340	****	(0.0507611)

L'erreur standard (S. E.) est entre parenthèses

* $p < 0.1$
** $p < 0.05$
*** $p < 0.01$
**** $p < 0.001$

En comparant les résultats présentés dans les 6 derniers tableaux, on note premièrement des résultats très similaires en termes des coefficients de régressions, exprimés en termes de ratios des taux d'incidences, et en termes de « *p*-value » témoignant des variables significatives des

modèles. Par contre, en analysant le degré de signification des variables, représentés par un nombre variant entre une (*) et quatre (****) étoiles, ont remarque que dans plusieurs des cas les résultats sont plus significatifs pour les modèles utilisant des données à partir de 1989 au lieu de 1993. Les cas les plus souvent observés dans les tableaux ci-dessus sont le passage à une catégorie de probabilité inférieure. Par exemple, dans le Tableau 4.11 la variable « nbcont3 » passe de « $p < 0.05$ » à « $p < 0.01$ ». Ce résultat supporte notre affirmation générale énonçant qu'utiliser des ensembles de données plus important permettaient d'obtenir des résultats plus significatifs. En omettant les analyses #7, #8 et #9 de la suite de l'analyse, il reste donc 6 modèles potentiels soit basés sur les analyses #4, #5 et #6. Ces analyses utilisent donc l'ensemble de données à partir de l'année 1989 jusqu'en 2005. Comme le même modèle est comparé, soit toujours le modèle B11, la différence réside au niveau du retard de phase utilisé (sans retard, une année « L1 » ou deux années « L2 ») et au niveau de l'horizon utilisé pour les réseaux (3 ans ou 5 ans). Les deux tableaux suivant permettent de faire une comparaison entre les différentes combinaisons d'horizon réseau et de retard de phase.

Tableau 4.15 Analyses #4, #5 et #6 pour les réseaux sur 3 ans

	Analyse #4 - 3 ans - sans retard		**Analyse #5 - 3 ans - L1**		**Analyse #6 - 3 ans - L2**	
	IRR	***S.E.***	***IRR***	***S.E.***	***IRR***	***S.E.***
nbsubv3	1.0012270	(0.0011427)	1.0072570****	(0.0011204)	1.0067670****	(0.0011862)
ltotsubvmoyan3	1.0634490****	(0.0093575)	1.0101640***	(0.0034247)	1.0124500****	(0.0029200)
nbcont3	1.0021730	(0.0017618)	1.0049390**	(0.0019648)	1.0054950***	(0.0020802)
ltotcontmoyan3	1.0034370*	(0.0019855)	1.0050740**	(0.0022470)	1.0044700*	(0.0023127)
mdegcentindn3	1.9943000****	(0.0587498)	1.4075540****	(0.0476771)	1.3226790****	(0.0467558)
mbetcentind3	0.9943594	(0.0077328)	1.0145570*	(0.0088099)	1.0104190	(0.0091809)
lmcliquind23	1.3414190****	(0.0153179)	1.0308630***	(0.0098719)	1.0350200****	(0.0094990)
d01Laval	1.0549920	(0.0348348)	1.0365570	(0.0394897)	1.0365270	(0.0408721)
d02McGill	1.0142060	(0.0321798)	1.0069150	(0.0367761)	1.0106840	(0.0382893)
d07Concordia	0.9204336	(0.0666768)	0.9293707	(0.0780449)	0.9481968	(0.0841848)
d08SherbrookeG	1.0490320	(0.0513861)	1.0130010	(0.0566632)	1.0181560	(0.0586626)
d09UQAMG	0.9954945	(0.0383815)	0.9311812	(0.0411764)	0.9362634	(0.0425960)
d1989						
d1990	1.0776290	(0.0643135)	1.0546040	(0.0741983)		
d1991	1.0593390	(0.0624345)	1.0457220	(0.0718718)	1.0197450	(0.0594919)
d1992	1.2592930****	(0.0725868)	1.1276090*	(0.0757816)	1.0495650	(0.0591909)
d1993	1.4087370****	(0.0798291)	1.1962660***	(0.0788423)	1.1269900**	(0.0613722)
d1994	1.6309010****	(0.0894369)	1.3517880****	(0.0863045)	1.2534480****	(0.0650462)
d1995	1.5638090****	(0.0851375)	1.3451850****	(0.0851983)	1.2607660****	(0.0645590)
d1996	2.4122180****	(0.1271198)	1.8173720****	(0.1106255)	1.7211300****	(0.0826730)
d1997	2.2659230****	(0.1188308)	1.8240030****	(0.1109744)	1.6636720****	(0.0796950)
d1998	2.2098100****	(0.1173587)	1.6633490****	(0.1014212)	1.5730460****	(0.0760028)
d1999	2.0751660****	(0.1098498)	1.6641810****	(0.1017366)	1.5254620****	(0.0734517)
d2000	1.7308320****	(0.0931329)	1.4232260****	(0.0883815)	1.3013820****	(0.0646930)
d2001	1.9748920****	(0.1045746)	1.6061470****	(0.0981422)	1.4604390****	(0.0705978)
d2002	1.7515470****	(0.0938909)	1.4238190****	(0.0880968)	1.2918520****	(0.0636603)
d2003	2.0815320****	(0.1099931)	1.6347910****	(0.0997232)	1.4666620****	(0.0708384)
d2004	2.0236720****	(0.1080304)	1.5620040****	(0.0960081)	1.3956450****	(0.0681929)
d2005	2.2629340****	(0.1194267)	1.8302110****	(0.1110646)	1.6227550****	(0.0777169)
	L'erreur standard (S. E.) est entre parenthèses					
	*$p < 0.1$					
	**$p < 0.05$					
	***$p < 0.01$					
	****$p < 0.001$					

Tableau 4.16 Analyses #4, #5 et #6 pour les réseaux sur 5 ans

	Analyse #4 - 5 ans - sans retard		**Analyse #5 - 5 ans - L1**		**Analyse #6 - 5 ans - L2**	
	IRR	***S.E.***	***IRR***	***S.E.***	***IRR***	***S.E.***
nbsubv5	1.0010160	(0.0008211)	1.0046190****	(0.0008127)	1.0043230****	(0.0008849)
ltotsubvmoyan5	1.0629000****	(0.0099239)	1.0135080****	(0.0037756)	1.0112980****	(0.0031661)
nbcont5	1.0019230	(0.0012390)	1.0032490**	(0.0014133)	1.0031960**	(0.0015144)
ltotcontmoyan5	1.0011670	(0.0021053)	1.0036690	(0.0023806)	1.0041760*	(0.0024907)
mdegcentindn5	2.7287900****	(0.0895619)	1.5873630****	(0.0600111)	1.4302970****	(0.0557407)
mbetcentind5	0.9380578****	(0.0077147)	0.9956794	(0.0092441)	1.0098460	(0.0096914)
lmcliquind25	1.1603720****	(0.0158966)	1.0008140	(0.0126671)	1.0082480	(0.0126907)
d01Laval	1.0609110*	(0.0355600)	1.0360310	(0.0396976)	1.0395420	(0.0417314)
d02McGill	1.0034770	(0.0324791)	0.9983843	(0.0366859)	1.0103420	(0.0389408)
d07Concordia	0.9433919	(0.0694770)	0.9424181	(0.0791636)	0.9645132	(0.0856405)
d08SherbrookeG	1.0384380	(0.0517088)	0.9951715	(0.0562482)	0.9802575	(0.0579890)
d09UQAMG	0.9956698	(0.0389683)	0.9256592*	(0.0411114)	0.9227493*	(0.0427798)
d1989	0.9527180	(0.0572766)				
d1990						
d1991	0.9706936	(0.0541127)	0.9879913	(0.0627034)	0.9689806	(0.0590606)
d1992	1.0342450	(0.0562914)	1.0333090	(0.0639799)		
d1993	1.1406800**	(0.0610443)	1.0690110	(0.0648388)	1.0477900	(0.0596352)
d1994	1.2953330****	(0.0667744)	1.2105540***	(0.0707669)	1.1491880**	(0.0627455)
d1995	1.2747120****	(0.0655031)	1.1989370***	(0.0695564)	1.1498670**	(0.0622160)
d1996	1.7911550****	(0.0884143)	1.6186120****	(0.0899051)	1.5627330****	(0.0799716)
d1997	1.7668350****	(0.0869028)	1.6072050****	(0.0893188)	1.5063720****	(0.0771782)
d1998	1.6851100****	(0.0846265)	1.4709890****	(0.0819643)	1.4054190****	(0.0726529)
d1999	1.6682040****	(0.0834729)	1.4626130****	(0.0822421)	1.3726520****	(0.0707393)
d2000	1.4181020****	(0.0726473)	1.2598880****	(0.0720323)	1.1712850***	(0.0624513)
d2001	1.5487030****	(0.0775266)	1.4058960****	(0.0792469)	1.3177210****	(0.0686060)
d2002	1.3544550****	(0.0685324)	1.2441640****	(0.0709051)	1.1622240***	(0.0617391)
d2003	1.5974490****	(0.0798966)	1.4274970****	(0.0798737)	1.3177510****	(0.0686046)
d2004	1.5009520****	(0.0760578)	1.3500560****	(0.0766293)	1.2513850****	(0.0656777)
d2005	1.6630980****	(0.0833778)	1.5547960****	(0.0870142)	1.4450450****	(0.0748796)
	L'erreur standard (S. E.) est entre parenthèses					
	*$p < 0.1$					
	**$p < 0.05$					
	***$p < 0.01$					
	****$p < 0.001$					

À la suite de l'analyse des probabilités des ratios des taux d'incidences des variables présentées dans le Tableau 4.15 et le Tableau 4.16, une analyse se démarque particulièrement. L'**analyse #5**

– 3 ans – L1, présentée au Tableau 4.15, possède la *totalité* de ses variables statistiques continues *significatives*. Ce sera donc cette analyse du modèle B11 qui sera retenue. Par contre, le Tableau 4.16 utilisant des données réseaux avec un horizon de 5 ans, présente des variables autant, sinon plus, significatives pour **nbsubv5**, **ltotsubvmoyan5** et **mdegcentindn5** toujours en considérant un même retard de phase L1. Pour ces trois variables, regardons en détails les résultats des analyses de régression pour **#5 – 3 ans – L1** et **#5 – 5 ans – L1**, particulièrement la valeur du test statistique *z* qui nous informe sur le sens et la force du coefficient. Les résultats complets de Stata sont présentés dans le Tableau 4.17 et le Tableau 4.18.

Tableau 4.17 Résultats de régression pour l'analyse #5 - 3 ans - L1

nbarticleB	IRR	Std. Err.	z	P>|z|	[95% Conf. Interval]	[95% Conf. Interval]2
nbsubv3.L1	1.00725700	0.00112040	6.500	0.000	1.0050640	1.0094560
ltotsubvmoyan3.L1	1.01016400	0.00342470	2.980	0.003	1.0034740	1.0168990
nbcont3.L1	1.00493900	0.00196480	2.520	0.012	1.0010960	1.0087980
ltotcontmoyan3.L1	1.00507400	0.00224700	2.260	0.024	1.0006800	1.0094880
mdegcentindn3.L1	1.40755400	0.04767710	10.090	0.000	1.3171430	1.5041710
mbetcentind3.L1	1.01455700	0.00880990	1.660	0.096	0.9974362	1.0319720
lmcliquind23.L1	1.03086300	0.00987190	3.170	0.002	1.0116950	1.0503950
d01Laval	1.03655700	0.03948970	0.940	0.346	0.9619774	1.1169180
d02McGill	1.00691500	0.03677610	0.190	0.850	0.9373541	1.0816370
d07Concordia	0.92937070	0.07804490	-0.870	0.383	0.7883307	1.0956440
d08SherbrookeG	1.01300100	0.05666320	0.230	0.817	0.9078141	1.1303750
d09UQAMG	0.93118120	0.04117640	-1.610	0.107	0.8538754	1.0154860
d1989						
d1990	1.05460400	0.07419830	0.760	0.450	0.9187597	1.2105340
d1991	1.04572200	0.07187180	0.650	0.515	0.9139318	1.1965170
d1992	1.12760900	0.07578160	1.790	0.074	0.9884458	1.2863640
d1993	1.19626600	0.07884230	2.720	0.007	1.0513030	1.3612190
d1994	1.35178800	0.08630450	4.720	0.000	1.1927900	1.5319810
d1995	1.34518500	0.08519830	4.680	0.000	1.1881480	1.5229780
d1996	1.81737200	0.11062550	9.810	0.000	1.6129850	2.0476580
d1997	1.82400300	0.11097440	9.880	0.000	1.6189650	2.0550080
d1998	1.66334900	0.10142120	8.350	0.000	1.4759850	1.8744960
d1999	1.66418100	0.10173660	8.330	0.000	1.4762640	1.8760190
d2000	1.42322600	0.08838150	5.680	0.000	1.2601280	1.6074330
d2001	1.60614700	0.09814220	7.750	0.000	1.4248640	1.8104940
d2002	1.42381900	0.08809680	5.710	0.000	1.2612120	1.6073920
d2003	1.63479100	0.09972320	8.060	0.000	1.4505690	1.8424090
d2004	1.56200400	0.09600810	7.260	0.000	1.3847240	1.7619800
d2005	1.83021100	0.11106460	9.960	0.000	1.6249750	2.0613680
/ln_r	2.87184800	0.06639590			2.7417150	3.0019820
/ln_s	1.22485600	0.05156340			1.1237930	1.3259180
r	17.66965000	1.17319300			15.5135600	20.1253900
s	3.40367500	0.17550500			3.0765020	3.7656410

Tableau 4.18 Résultats de régression pour l'analyse #5 - 5 ans - L1

nbarticleB	IRR	Std. Err.	z	P>|z|	[95% Conf. Interval]	[95% Conf. Interval]2
nbsubv5.L1	1.00461900	0.00081270	5.700	0.000	1.0030280	1.0062130
ltotsubvmoyan5.L1	1.01350800	0.00377560	3.600	0.000	1.0061350	1.0209350
nbcont5.L1	1.00324900	0.00141330	2.300	0.021	1.0004820	1.0060220
ltotcontmoyan5.L1	1.00366900	0.00238060	1.540	0.123	0.9990139	1.0083460
mdegcentindn5.L1	1.58736300	0.06001110	12.220	0.000	1.4739950	1.7094500
mbetcentind5.L1	0.99567940	0.00924410	-0.470	0.641	0.9777252	1.0139630
lmcliquind25.L1	1.00081400	0.01266710	0.060	0.949	0.9762919	1.0259510
d01Laval	1.03603100	0.03969760	0.920	0.356	0.9610748	1.1168330
d02McGill	0.99838430	0.03668590	-0.040	0.965	0.9290095	1.0729400
d07Concordia	0.94241810	0.07916360	-0.710	0.480	0.7993597	1.1110790
d08SherbrookeG	0.99517150	0.05624820	-0.090	0.932	0.8908141	1.1117540
d09UQAMG	0.92565920	0.04111140	-1.740	0.082	0.8484898	1.0098470
d1989	0.98799130	0.06270340	-0.190	0.849	0.8724312	1.1188580
d1990	1.03330900	0.06397990	0.530	0.597	0.9152211	1.1666340
d1991	0.98799130	0.06270340	-0.190	0.849	0.8724312	1.1188580
d1992	1.03330900	0.06397990	0.530	0.597	0.9152211	1.1666340
d1993	1.06901100	0.06483880	1.100	0.271	0.9491923	1.2039550
d1994	1.21055400	0.07076690	3.270	0.001	1.0795050	1.3575130
d1995	1.19893700	0.06955640	3.130	0.002	1.0700740	1.3433180
d1996	1.61861200	0.08990510	8.670	0.000	1.4516540	1.8047720
d1997	1.60720500	0.08931880	8.540	0.000	1.4413400	1.7921560
d1998	1.47098900	0.08196430	6.930	0.000	1.3188030	1.6407370
d1999	1.46261300	0.08224210	6.760	0.000	1.3099870	1.6330230
d2000	1.25988800	0.07203230	4.040	0.000	1.1263300	1.4092830
d2001	1.40589600	0.07924690	6.040	0.000	1.2588480	1.5701220
d2002	1.24416400	0.07090510	3.830	0.000	1.1126730	1.3911950
d2003	1.42749700	0.07987370	6.360	0.000	1.2792260	1.5929530
d2004	1.35005600	0.07662930	5.290	0.000	1.2079180	1.5089190
d2005	1.55479600	0.08701420	7.890	0.000	1.3932720	1.7350460
/ln_r	2.90712400	0.06888930			2.7721030	3.0421440
/ln_s	1.23542400	0.05293760			1.1316680	1.3391800
r	18.30407000	1.26095600			15.9922300	20.9501200
s	3.43983500	0.18209680			3.1008240	3.8159110

En comparant le Tableau 4.17 et le Tableau 4.18, on trouve que les variables liées au nombre de subventions, au total annuel des subventions et aux mesures du degré de centralité ont des valeurs d'IRR semblables et sont très significatives sur des horizons de trois ans et cinq ans. Par contre, on note une différence au niveau du test statistique *z* qui indique que les variables du total annuel

et du degré de centralité, liées aux réseaux avec un horizon de 5 ans ont plus de « force » donc auraient une influence positive plus grande que les variables avec un horizon de 3 ans.

Pour analyser l'effet des ces trois variables sur le modèle B11 avec des réseaux sur trois ans combiné avec un retard de période d'un an, quatre analyses de régressions seront effectuées. Le Tableau 4.19 présente ces analyses.

Tableau 4.19 Variations du modèle B11

Modèles	*Mesures de subventions*		*Mesures de contrats*		*Mesures réseau*		
Modèle B11_1	nbsubv3	ltotsubvmoyan3	nbcont3	ltotcontmoyan3	mdegcentindn3	mbetcentind3	lmcliquind23
Modèle B11_2	nbsubv3	ltotsubvmoyan5	nbcont3	ltotcontmoyan3	mdegcentindn3	mbetcentind3	lmcliquind23
Modèle B11_3	nbsubv3	ltotsubvmoyan3	nbcont3	ltotcontmoyan3	mdegcentindn5	mbetcentind3	lmcliquind23
Modèle B11_4	nbsubv3	ltotsubvmoyan5	nbcont3	ltotcontmoyan3	mdegcentindn5	mbetcentind3	lmcliquind23
Modèle B11_5	nbsubv5	ltotsubvmoyan3	nbcont3	ltotcontmoyan3	mdegcentindn3	mbetcentind3	lmcliquind23
Modèle B11_6	nbsubv5	ltotsubvmoyan5	nbcont3	ltotcontmoyan3	mdegcentindn3	mbetcentind3	lmcliquind23
Modèle B11_7	nbsubv5	ltotsubvmoyan3	nbcont3	ltotcontmoyan3	mdegcentindn5	mbetcentind3	lmcliquind23
Modèle B11_8	nbsubv5	ltotsubvmoyan5	nbcont3	ltotcontmoyan3	mdegcentindn5	mbetcentind3	lmcliquind23

Le Tableau 4.19 présente les analyses de l'impact des subventions annuelles moyennes et du degré de centralité de l'horizon de 5 ans sur le modèle basé sur un horizon de 3 ans. Ce sont donc les combinaisons possibles avec ces deux variables de façon à capturer l'impact sur le modèle final. Le tableau suivant présente les mesures AIC/BIC pour les modèles définis au Tableau 4.19.

Tableau 4.20 Mesures AIC/BIC pour les variations du modèle B11

Model	*Obs*	*ll(null)*	*ll(model)*	*df*	*AIC*	*BIC*
B11_1	17426	.	-24890.0	31	49842.070	50082.810
B11_2	17432	.	-24899.2	31	49860.330	50101.080
B11_3	16895	.	-24228.8	30	48517.590	48749.640
B11_4	16901	.	-24238.0	30	48536.010	48768.060
B11_5	17426	.	-24887.8	31	49837.500	50078.230
B11_6	17432	.	-24897.7	31	49857.310	50098.060
B11_7	16895	.	-24227.8	30	48515.620	48747.660
B11_8	16901	.	-24237.7	30	48535.360	48767.410

Les résultats du Tableau 4.20 présentent, pour un nombre d'observation relativement similaires, une variation importante vers le bas du AIC/BIC lorsqu'on introduit la mesure du degré de centralité provenant des réseaux avec un horizon de 5 ans (modèles B11_3, B11_4, B11_7 et

B11_8). On note donc un écart lorsqu'on passe du modèle B11_2 ou B11_6 au modèle B11_3 ou B11_7. Voyons plus en détails les résultats complets générés par Stata où la colonne « ***p-value level*** » a été ajouté pour permettre l'évaluation rapide de la probabilité du « *p*-value ».

Tableau 4.21 Résultats de régressions pour le modèle B11_1

nbarticleB	*IRR*	*Std. Err.*	*z*	*P>|z|*	*p-value level*	*[95% Conf. Interval]*	
nbsubv3.L1	1.00725700	0.00112040	6.500	0.000	****	1.00506400	1.00945600
ltotsubvmoyan3.L1	1.01016400	0.00342470	2.980	0.003	***	1.00347400	1.01689900
nbcont3.L1	1.00493900	0.00196480	2.520	0.012	**	1.00109600	1.00879800
ltotcontmoyan3.L1	1.00507400	0.00224700	2.260	0.024	**	1.00068000	1.00948800
mdegcentindn3.L1	1.40755400	0.04767710	10.090	0.000	****	1.31714300	1.50417100
mbetcentind3.L1	1.01455700	0.00880990	1.660	0.096	*	0.99743620	1.03197200
lmcliquind23.L1	1.03086300	0.00987190	3.170	0.002	***	1.01169500	1.05039500
d01Laval	1.03655700	0.03948970	0.940	0.346		0.96197740	1.11691800
d02McGill	1.00691500	0.03677610	0.190	0.850		0.93735410	1.08163700
d07Concordia	0.92937070	0.07804490	-0.870	0.383		0.78833070	1.09564400
d08SherbrookeG	1.01300100	0.05666320	0.230	0.817		0.90781410	1.13037500
d09UQAMG	0.93118120	0.04117640	-1.610	0.107		0.85387540	1.01548600
d1989							
d1990	1.05460400	0.07419830	0.760	0.450		0.91875970	1.21053400
d1991	1.04572200	0.07187180	0.650	0.515		0.91393180	1.19651700
d1992	1.12760900	0.07578160	1.790	0.074	*	0.98844580	1.28636400
d1993	1.19626600	0.07884230	2.720	0.007	***	1.05130300	1.36121900
d1994	1.35178800	0.08630450	4.720	0.000	****	1.19279000	1.53198100
d1995	1.34518500	0.08519830	4.680	0.000	****	1.18814800	1.52297800
d1996	1.81737200	0.11062550	9.810	0.000	****	1.61298500	2.04765800
d1997	1.82400300	0.11097440	9.880	0.000	****	1.61896500	2.05500800
d1998	1.66334900	0.10142120	8.350	0.000	****	1.47598500	1.87449600
d1999	1.66418100	0.10173660	8.330	0.000	****	1.47626400	1.87601900
d2000	1.42322600	0.08838150	5.680	0.000	****	1.26012800	1.60743300
d2001	1.60614700	0.09814220	7.750	0.000	****	1.42486400	1.81049400
d2002	1.42381900	0.08809680	5.710	0.000	****	1.26121200	1.60739200
d2003	1.63479100	0.09972320	8.060	0.000	****	1.45056900	1.84240900
d2004	1.56200400	0.09600810	7.260	0.000	****	1.38472400	1.76198000
d2005	1.83021100	0.11106460	9.960	0.000	****	1.62497500	2.06136800
				*	$p < 0.1$		
				**	$p < 0.05$		
				***	$p < 0.01$		
				****	$p < 0.001$		

Tableau 4.22 Résultats de régressions pour le modèle B11_2

nbarticleB	*IRR*	*Std. Err.*	*z*	*P>|z|*	*p-value level*	*[95% Conf. Interval]*	
nbsubv3.L1	1.00698100	0.00111630	6.280	0.000	****	1.00479600	1.00917100
ltotsubvmoyan5.L1	1.01386500	0.00371020	3.760	0.000	****	1.00662000	1.02116300
nbcont3.L1	1.00491300	0.00196420	2.510	0.012	**	1.00107100	1.00877000
ltotcontmoyan3.L1	1.00479900	0.00224830	2.140	0.032	**	1.00040200	1.00921500
mdegcentindn3.L1	1.39887500	0.04726730	9.930	0.000	****	1.30923400	1.49465400
mbetcentind3.L1	1.01557300	0.00879750	1.780	0.074	*	0.99847550	1.03296300
lmcliquind23.L1	1.03093400	0.00987170	3.180	0.001	***	1.01176700	1.05046500
d01Laval	1.03803900	0.03957620	0.980	0.327		0.96329790	1.11857800
d02McGill	1.00591100	0.03677040	0.160	0.872		0.93636360	1.08062400
d07Concordia	0.93250430	0.07847760	-0.830	0.406		0.79070680	1.09973000
d08SherbrookeG	1.01493800	0.05680210	0.260	0.791		0.90949650	1.13260300
d09UQAMG	0.93350720	0.04128340	-1.560	0.120		0.85600070	1.01803200
d1989							
d1990	1.05543700	0.07422260	0.770	0.443		0.91954380	1.21141300
d1991	1.04656300	0.07189920	0.660	0.508		0.91471850	1.19741100
d1992	1.12765900	0.07574890	1.790	0.074	*	0.98855220	1.28634100
d1993	1.19669700	0.07883390	2.730	0.006	***	1.05174400	1.36162600
d1994	1.35121300	0.08622120	4.720	0.000	****	1.19236300	1.53122600
d1995	1.34524300	0.08516340	4.680	0.000	****	1.18826600	1.52295800
d1996	1.81622200	0.11049280	9.810	0.000	****	1.61207300	2.04622400
d1997	1.82531800	0.11097000	9.900	0.000	****	1.62027900	2.05630400
d1998	1.65824800	0.10106910	8.300	0.000	****	1.47153000	1.86865700
d1999	1.66041700	0.10146890	8.300	0.000	****	1.47299000	1.87169200
d2000	1.41894800	0.08808680	5.640	0.000	****	1.25639100	1.60253700
d2001	1.60132900	0.09780850	7.710	0.000	****	1.42065800	1.80497700
d2002	1.41853200	0.08773950	5.650	0.000	****	1.25658000	1.60135600
d2003	1.62305600	0.09899580	7.940	0.000	****	1.44017700	1.82915800
d2004	1.55348900	0.09549360	7.170	0.000	****	1.37716000	1.75239400
d2005	1.81997200	0.11044860	9.870	0.000	****	1.61587500	2.04984700
				*	$p < 0.1$		
				**	$p < 0.05$		
				***	$p < 0.01$		
				****	$p < 0.001$		

Tableau 4.23 Résultats de régressions pour le modèle B11_3

nbarticleB	*IRR*	*Std. Err.*	*z*	*P>|z|*	*p-value level*	*[95% Conf. Interval]*	
nbsubv3.L1	1.00678400	0.00113360	6.010	0.000	****	1.00456500	1.00900800
ltotsubvmoyan3.L1	1.00989500	0.00344610	2.890	0.004	***	1.00316400	1.01667200
nbcont3.L1	1.00523800	0.00197670	2.660	0.008	***	1.00137100	1.00911900
ltotcontmoyan3.L1	1.00433600	0.00225540	1.930	0.054	*	0.99992520	1.00876600
mdegcentindn5.L1	1.50782000	0.04698660	13.180	0.000	****	1.41848400	1.60278300
mbetcentind3.L1	1.00714100	0.00802430	0.890	0.372		0.99153600	1.02299200
lmcliquind23.L1	1.02817900	0.00939230	3.040	0.002	***	1.00993400	1.04675400
d01Laval	1.03277500	0.03917790	0.850	0.395		0.95877330	1.11248900
d02McGill	0.99793820	0.03621720	-0.060	0.955		0.92941950	1.07150800
d07Concordia	0.94837300	0.07886470	-0.640	0.524		0.80574030	1.11625500
d08SherbrookeG	0.99366990	0.05558160	-0.110	0.910		0.89049110	1.10880400
d09UQAMG	0.92625330	0.04084060	-1.740	0.082	*	0.84956850	1.00986000
d1989							
d1990							
d1991	0.98493680	0.06256740	-0.240	0.811		0.86963380	1.11552800
d1992	1.02845800	0.06372690	0.450	0.651		0.91084230	1.16126200
d1993	1.07507900	0.06513410	1.190	0.232		0.95470620	1.21062800
d1994	1.22450900	0.07138390	3.470	0.001	***	1.09229600	1.37272500
d1995	1.21684900	0.07032390	3.400	0.001	***	1.08653600	1.36279000
d1996	1.64970900	0.09097440	9.080	0.000	****	1.48070000	1.83800800
d1997	1.65324700	0.09105560	9.130	0.000	****	1.48407700	1.84170200
d1998	1.51136700	0.08343570	7.480	0.000	****	1.35637200	1.68407300
d1999	1.50976400	0.08361100	7.440	0.000	****	1.35447000	1.68286300
d2000	1.29449800	0.07299980	4.580	0.000	****	1.15904500	1.44578200
d2001	1.45472300	0.08053030	6.770	0.000	****	1.30514800	1.62144000
d2002	1.29370600	0.07267280	4.580	0.000	****	1.15883100	1.44427900
d2003	1.49175100	0.08233740	7.250	0.000	****	1.33879500	1.66218100
d2004	1.42072700	0.07913500	6.300	0.000	****	1.27379200	1.58461200
d2005	1.65501000	0.09079490	9.180	0.000	****	1.48628800	1.84288400
				*	$p < 0.1$		
				**	$p < 0.05$		
				***	$p < 0.01$		
				****	$p < 0.001$		

Tableau 4.24 Résultats de régressions pour le modèle B11_4

nbarticleB	*IRR*	*Std. Err.*	*z*	*P>|z|*	*p-value level*	*[95% Conf. Interval]*	
nbsubv3.L1	1.00653900	0.00112900	5.810	0.000	****	1.00432900	1.00875500
ltotsubvmoyan5.L1	1.01323700	0.00373070	3.570	0.000	****	1.00595100	1.02057500
nbcont3.L1	1.00520500	0.00197640	2.640	0.008	***	1.00133900	1.00908600
ltotcontmoyan3.L1	1.00408800	0.00225690	1.810	0.070	*	0.99967410	1.00852100
mdegcentindn5.L1	1.50160900	0.04677930	13.050	0.000	****	1.41266600	1.59615100
mbetcentind3.L1	1.00772500	0.00802190	0.970	0.334		0.99212470	1.02357100
lmcliquind23.L1	1.02788400	0.00938770	3.010	0.003	***	1.00964800	1.04644900
d01Laval	1.03409700	0.03925450	0.880	0.377		0.95995250	1.11396900
d02McGill	0.99699520	0.03621130	-0.080	0.934		0.92848960	1.07055500
d07Concordia	0.95144490	0.07928140	-0.600	0.550		0.80808170	1.12024200
d08SherbrookeG	0.99544350	0.05571010	-0.080	0.935		0.89202910	1.11084700
d09UQAMG	0.92841940	0.04093790	-1.680	0.092	*	0.85155190	1.01222600
d1989							
d1990	1.01506100	0.06445600	0.240	0.814		0.89627480	1.14959000
d1991							
d1992	1.04378600	0.06209250	0.720	0.471		0.92891370	1.17286400
d1993	1.09142700	0.06343170	1.510	0.132		0.97392310	1.22310900
d1994	1.24207000	0.06933390	3.880	0.000	****	1.11334800	1.38567400
d1995	1.23479100	0.06827750	3.810	0.000	****	1.10796600	1.37613300
d1996	1.67301300	0.08780450	9.810	0.000	****	1.50947500	1.85427000
d1997	1.67857400	0.08808630	9.870	0.000	****	1.51451000	1.86041100
d1998	1.52923200	0.08040310	8.080	0.000	****	1.37949200	1.69522500
d1999	1.52857000	0.08070600	8.040	0.000	****	1.37829900	1.69522500
d2000	1.30979400	0.07054280	5.010	0.000	****	1.17858000	1.45561700
d2001	1.47215800	0.07767630	7.330	0.000	****	1.32752300	1.63255100
d2002	1.30813200	0.07016100	5.010	0.000	****	1.17760000	1.45313300
d2003	1.50349900	0.07914690	7.750	0.000	****	1.35610800	1.66690900
d2004	1.43417100	0.07627340	6.780	0.000	****	1.29220500	1.59173300
d2005	1.67053300	0.08735130	9.810	0.000	****	1.50780800	1.85081900
				*	$p < 0.1$		
				**	$p < 0.05$		
				***	$p < 0.01$		
				****	$p < 0.001$		

Tableau 4.25 Résultats de régressions pour le modèle B11_5

nbarticleB	*IRR*	*Std. Err.*	*z*	*P>|z|*	*p-value level*	*[95% Conf. Interval]*	
nbsubv5.L1	1.00515400	0.00079630	6.490	0.000	****	1.00359400	1.00671600
ltotsubvmoyan5.L1	1.01312100	0.00372510	3.550	0.000	****	1.00584600	1.02044900
nbcont3.L1	1.00451000	0.00196740	2.300	0.022	**	1.00066200	1.00837400
ltotcontmoyan3.L1	1.00463800	0.00224910	2.070	0.039	**	1.00024000	1.00905600
mdegcentindn3.L1	1.39456700	0.04711930	9.840	0.000	****	1.30520700	1.49004600
mbetcentind3.L1	1.01605700	0.00880550	1.840	0.066	*	0.99894470	1.03346300
lmcliquind23.L1	1.03144600	0.00987810	3.230	0.001	***	1.01226600	1.05098900
d01Laval	1.04020600	0.03968130	1.030	0.301		0.96526830	1.12096100
d02McGill	1.00713000	0.03682920	0.190	0.846		0.93747250	1.08196400
d07Concordia	0.93348110	0.07864170	-0.820	0.414		0.79139910	1.10107100
d08SherbrookeG	1.01449100	0.05680170	0.260	0.797		0.90905260	1.13215800
d09UQAMG	0.93160480	0.04120860	-1.600	0.109		0.85423950	1.01597700
d1989							
d1990	1.04551000	0.07351560	0.630	0.527		0.91091050	1.20000000
d1991	1.03667900	0.07122480	0.520	0.600		0.90607250	1.18611300
d1992	1.11597400	0.07499070	1.630	0.102		0.97826270	1.27307200
d1993	1.18163900	0.07788760	2.530	0.011	**	1.03843200	1.34459600
d1994	1.33056900	0.08499010	4.470	0.000	****	1.17399700	1.50802300
d1995	1.32470700	0.08395140	4.440	0.000	****	1.16997400	1.49990500
d1996	1.78566400	0.10876890	9.520	0.000	****	1.58471400	2.01209400
d1997	1.79291600	0.10921950	9.580	0.000	****	1.59113500	2.02028600
d1998	1.62982800	0.09946910	8.000	0.000	****	1.44608100	1.83692300
d1999	1.63183100	0.09988290	8.000	0.000	****	1.44735100	1.83982500
d2000	1.39503100	0.08675750	5.350	0.000	****	1.23494400	1.57587000
d2001	1.57125900	0.09621880	7.380	0.000	****	1.39355200	1.77162800
d2002	1.38933600	0.08625350	5.300	0.000	****	1.23016300	1.56910500
d2003	1.58895900	0.09733860	7.560	0.000	****	1.40918700	1.79166500
d2004	1.51948100	0.09390410	6.770	0.000	****	1.34614200	1.71514000
d2005	1.77312900	0.10836280	9.370	0	****	1.57296900	1.99876000
				*	$p < 0.1$		
				**	$p < 0.05$		
				***	$p < 0.01$		
				****	$p < 0.001$		

Tableau 4.26 Résultats de régressions pour le modèle B11_6

nbarticleB	*IRR*	*Std. Err.*	*z*	*P>|z|*	*p-value level*	*[95% Conf. Interval]*	
nbsubv5.L1	1.00489200	0.00079880	6.140	0.000	****	1.00332700	1.00645900
ltotsubvmoyan3.L1	1.01003100	0.00342650	2.940	0.003	***	1.00333800	1.01676900
nbcont3.L1	1.00488500	0.00197950	2.470	0.013	**	1.00101200	1.00877200
ltotcontmoyan3.L1	1.00416900	0.00225690	1.850	0.064	*	0.99975570	1.00860200
mdegcentindn5.L1	1.49809400	0.04681120	12.940	0.000	****	1.40909800	1.59271000
mbetcentind3.L1	1.00819200	0.00804040	1.020	0.306		0.99255570	1.02407500
lmcliquind23.L1	1.02888500	0.00940450	3.120	0.002	***	1.01061700	1.04748400
d01Laval	1.03477000	0.03929820	0.900	0.368		0.96054400	1.11473200
d02McGill	0.99915280	0.03629720	-0.020	0.981		0.93048520	1.07288800
d07Concordia	0.95006490	0.07914990	-0.610	0.539		0.80693700	1.11858000
d08SherbrookeG	0.99311070	0.05560440	-0.120	0.902		0.88989500	1.10829800
d09UQAMG	0.92402160	0.04076990	-1.790	0.073	*	0.84747170	1.00748600
d1989							
d1990	1.01535200	0.06446280	0.240	0.810		0.89655210	1.14989400
d1991							
d1992	1.04353000	0.06207580	0.720	0.474		0.92868840	1.17257300
d1993	1.08841900	0.06324720	1.460	0.145		0.97125500	1.21971600
d1994	1.23669600	0.06903480	3.810	0.000	****	1.10852900	1.37968100
d1995	1.22896400	0.06795490	3.730	0.000	****	1.10273800	1.36963800
d1996	1.66317700	0.08728220	9.690	0.000	****	1.50061100	1.84335400
d1997	1.66662300	0.08753110	9.730	0.000	****	1.50360000	1.84732200
d1998	1.52310100	0.08006490	8.000	0.000	****	1.37399000	1.68839500
d1999	1.52143400	0.08033000	7.950	0.000	****	1.37186300	1.68731300
d2000	1.30477100	0.07028250	4.940	0.000	****	1.17404200	1.45005600
d2001	1.46380700	0.07728690	7.220	0.000	****	1.31990200	1.62340200
d2002	1.29940000	0.06979020	4.880	0.000	****	1.16956700	1.44364500
d2003	1.49767100	0.07896700	7.660	0.000	****	1.35062800	1.66072400
d2004	1.42523500	0.07596160	6.650	0.000	****	1.28386500	1.58217100
d2005	1.65421600	0.08681250	9.590	0.000	****	1.49252500	1.83342400
				*	$p < 0.1$		
				**	$p < 0.05$		
				***	$p < 0.01$		
				****	$p < 0.001$		

Tableau 4.27 Résultats de régressions pour le modèle B11_7

nbarticleB	*IRR*	*Std. Err.*	*z*	*P>|z|*	*p-value level*	*[95% Conf. Interval]*	
nbsubv5.L1	1.00470000	0.00080690	5.840	0.000	****	1.00312000	1.00628300
ltotsubvmoyan5.L1	1.01272500	0.00375340	3.410	0.001	***	1.00539600	1.02010900
nbcont3.L1	1.00487400	0.00197910	2.470	0.014	**	1.00100200	1.00876000
ltotcontmoyan3.L1	1.00399900	0.00225780	1.770	0.076	*	0.99958380	1.00843400
mdegcentindn5.L1	1.49293400	0.04662790	12.830	0.000	****	1.40428600	1.58717800
mbetcentind3.L1	1.00872700	0.00803800	1.090	0.276		0.99309520	1.02460500
lmcliquind23.L1	1.02853400	0.00939880	3.080	0.002	***	1.01027700	1.04712200
d01Laval	1.03605100	0.03937200	0.930	0.351		0.96168730	1.11616600
d02McGill	0.99826740	0.03629210	-0.050	0.962		0.92961130	1.07199400
d07Concordia	0.95224790	0.07947380	-0.590	0.558		0.80855470	1.12147800
d08SherbrookeG	0.99484680	0.05572710	-0.090	0.927		0.89140590	1.11029100
d09UQAMG	0.92610780	0.04086130	-1.740	0.082	*	0.84938610	1.00975900
d1989							
d1990							
d1991	0.98513670	0.06252510	-0.240	0.813		0.86990540	1.11563200
d1992	1.02768800	0.06362990	0.440	0.659		0.91024550	1.16028300
d1993	1.07233600	0.06491090	1.150	0.249		0.95236960	1.20741300
d1994	1.21739100	0.07090810	3.380	0.001	***	1.08605300	1.36461200
d1995	1.21022600	0.06988990	3.300	0.001	***	1.08071200	1.35526100
d1996	1.63741600	0.09022460	8.950	0.000	****	1.46979300	1.82415500
d1997	1.64176100	0.09036500	9.010	0.000	****	1.47386800	1.82877900
d1998	1.49617800	0.08255360	7.300	0.000	****	1.34281800	1.66705300
d1999	1.49561400	0.08280480	7.270	0.000	****	1.34181500	1.66704100
d2000	1.28188400	0.07228540	4.400	0.000	****	1.14775600	1.43168700
d2001	1.43859100	0.07965860	6.570	0.000	****	1.29063700	1.60350700
d2002	1.27625500	0.07177720	4.340	0.000	****	1.14305000	1.42498200
d2003	1.46640300	0.08109970	6.920	0.000	****	1.31576200	1.63429000
d2004	1.39794600	0.07809130	6.000	0.000	****	1.25297100	1.55969500
d2005	1.62276200	0.08938970	8.790	0.000	****	1.45668700	1.80777000
				*	$p < 0.1$		
				**	$p < 0.05$		
				***	$p < 0.01$		
				****	$p < 0.001$		

Tableau 4.28 Résultats de régressions pour le modèle B11_8

nbarticleB	*IRR*	*Std. Err.*	*z*	*P>|z|*	*p-value level*	*[95% Conf. Interval]*	
nbsubv5.L1	1.00470000	0.00080690	5.840	0.000	****	1.00312000	1.00628300
ltotsubvmoyan5.L1	1.01272500	0.00375340	3.410	0.001	***	1.00539600	1.02010900
nbcont3.L1	1.00487400	0.00197910	2.470	0.014	**	1.00100200	1.00876000
ltotcontmoyan3.L1	1.00399900	0.00225780	1.770	0.076	*	0.99958380	1.00843400
mdegcentindn5.L1	1.49293400	0.04662790	12.830	0.000	****	1.40428600	1.58717800
mbetcentind3.L1	1.00872700	0.00803800	1.090	0.276		0.99309520	1.02460500
lmcliquind23.L1	1.02853400	0.00939880	3.080	0.002	***	1.01027700	1.04712200
d01Laval	1.03605100	0.03937200	0.930	0.351		0.96168730	1.11616600
d02McGill	0.99826740	0.03629210	-0.050	0.962		0.92961130	1.07199400
d07Concordia	0.95224790	0.07947380	-0.590	0.558		0.80855470	1.12147800
d08SherbrookeG	0.99484680	0.05572710	-0.090	0.927		0.89140590	1.11029100
d09UQAMG	0.92610780	0.04086130	-1.740	0.082	*	0.84938610	1.00975900
d1989							
d1990							
d1991	0.98513670	0.06252510	-0.240	0.813		0.86990540	1.11563200
d1992	1.02768800	0.06362990	0.440	0.659		0.91024550	1.16028300
d1993	1.07233600	0.06491090	1.150	0.249		0.95236960	1.20741300
d1994	1.21739100	0.07090810	3.380	0.001	***	1.08605300	1.36461200
d1995	1.21022600	0.06988990	3.300	0.001	***	1.08071200	1.35526100
d1996	1.63741600	0.09022460	8.950	0.000	****	1.46979300	1.82415500
d1997	1.64176100	0.09036500	9.010	0.000	****	1.47386800	1.82877900
d1998	1.49617800	0.08255360	7.300	0.000	****	1.34281800	1.66705300
d1999	1.49561400	0.08280480	7.270	0.000	****	1.34181500	1.66704100
d2000	1.28188400	0.07228540	4.400	0.000	****	1.14775600	1.43168700
d2001	1.43859100	0.07965860	6.570	0.000	****	1.29063700	1.60350700
d2002	1.27625500	0.07177720	4.340	0.000	****	1.14305000	1.42498200
d2003	1.46640300	0.08109970	6.920	0.000	****	1.31576200	1.63429000
d2004	1.39794600	0.07809130	6.000	0.000	****	1.25297100	1.55969500
d2005	1.62276200	0.08938970	8.790	0.000	****	1.45668700	1.80777000
				*	$p < 0.1$		
				**	$p < 0.05$		
				***	$p < 0.01$		
				****	$p < 0.001$		

En analysant les résultats présentés dans les 8 derniers tableaux au niveau des *variables statistiques continues* l'introduction de la variable pour la mesure du degré de centralité (« degree centrality ») provenant des réseaux avec un horizon de 5 ans a un impact négatif sur la signifiance de la variable de centralité de relation d'intermédiaire (« betweenness centrality ») d'un horizon de 3 ans. Comme la mesure de centralité de relation d'intermédiaire est basée sur la

capacité d'un chercheur à contrôler le flux d'informations en fonction de sa position dans le réseau c'est l'importance d'un chercheur pour la transmission des informations d'un réseau alors que la centralité est l'importance des contacts d'un chercheur avec des sources d'informations (par exemple, avec d'autres chercheurs). Comme la position dans le réseau est basée sur le nombre de liens vers les voisins (contacts), il est prévisible qu'en prenant une variable avec un historique plus grand, soit avec un horizon de 5 ans, celle-ci va inclure des données de la variable de l'horizon de 3 ans et donc pouvoir s'exprimer en fonction de celle-ci ce qui peut expliquer la présence de colinéarité dans les variables nominales.

En analysant les mêmes tableaux de résultats pour les *variables nominales* on observe qu'évidemment la variable nominale pour l'année 1989 n'est pas incluse car celle-ci étant l'année de référence, par contre, pour les modèles B11_3, B11_4, B11_7 et B11_8 l'introduction de la variable pour la mesure du degré de centralité provenant des réseaux avec un horizon de 5 ans ajoute de la **colinéarité** corrigée par Stata en enlevant une autre variable nominale. Pour le modèle B11_3 et B11_8 on remarque que l'année 1990 a été éliminée tandis que pour le modèle B11_4 et B11_7 c'est l'année 1991. Ce comportement explique donc la perte d'un degré de liberté des deux derniers modèles observés dans les résultats des mesures AIC/BIC au Tableau 4.20. Les modèles B11_3, B11_4, B11_7 et B11_8 sont donc omis du reste de l'analyse. Il revient donc à comparer les modèles B11_1, B11_2, B11_5 et B11_6 qui sont similaires en termes de résultats. Ils ont le même degré de liberté (31) et ont des mesures AIC/BIC sont très comparables soit moins de 20 points d'écart. C'est l'introduction de la mesure du total des subventions sur une base annuelle provenant des réseaux avec un horizon de 5 ans qui fait changer la catégorie du « *p*-value » de *** à **** et fait augmenter positivement la valeur du test statistique z. Le choix du modèle est donc entre B11_5 et B11_6.

Bien que les deux modèles soient valides, l'élément qui fait pencher le choix du modèle final pour **B11_6** provient de la caractéristique de *durée* des subventions dans un contexte des réseaux sociaux présentées à la section 1.3. Les subventions sont généralement d'une durée plus longue que les contrats et donc un modèle qui capture les variables liées aux subventions sur un horizon de 5 ans alors que les variables liées aux contrats seraient calculées avec un horizon de 3 ans.

La robustesse du modèle a été validée par les faibles différences présentes dans les résultats des différentes variations du modèle général. En substituant une variable pour une autre on obtient

très peu de variation. La section suivante traitera de la nécessité de catégoriser le financement pour mieux l'analyser.

4.2.1.2 Catégorisation du financement

La section 2.3.1.2, traitant des données provenant de SIRU, a défini dans le Tableau 2.4 les catégories du financement sous la forme de contrats et de subventions. Jusqu'à présent, l'analyse a considéré le financement total, soit la somme des catégories, mais qu'advient'il si on sépare ces catégories dans le modèle final soit B11_6 ? L'analyse suivante répondra à cette question en regroupant selon deux catégories soit « appareillage-infrastructure » et « fonctionnement ».

Le Tableau 4.29 et le Tableau 4.30 montrent la catégorisation des variables liées aux subventions et aux contrats.

Tableau 4.29 Catégorisation des subventions pour le modèle B11_6

Modèles	***Mesures de subventions***			
B11_6	nbsubv5		ltotsubvmoyan5	
B11_6AIF	nbsubvAI5	nbsubvF5	ltotsubvAImoyan5	ltotsubvFmoyan5

Tableau 4.30 Catégorisation des contrats pour le modèle B11_6

Modèles	***Mesures de contrats***			
B11_6	nbcont3		ltotcontmoyan3	
B11_6AIF	nbcontAI3	nbcontF3	ltotcontAImoyan3	ltotcontFmoyan3

Tel que présenté dans le Tableau 4.29 et le Tableau 4.30 la variation provient uniquement des variables liées au financement. Voyons maintenant quel sera l'impact sur les résultats de régressions.

Tableau 4.31 Mesures AIC/BIC pour le modèle B11_6

Model	***Obs***	***ll(null)***	***ll(model)***	***df***	***AIC***	***BIC***
B11_6	17432	.	-24897.7	31	49857.310	50098.060
B11_6AIF	17420	.	-24875.4	35	49820.740	50092.530

Tableau 4.32 Régressions pour le modèle B11_6

nbarticleB	IRR	Std. Err.	z	P>|z|	*p*-value level	[95% Conf. Interval]	
nbsubv5.L1	1.00515400	0.00079630	6.490	0.000	****	1.00359400	1.00671600
ltotsubvmoyan5.L1	1.01312100	0.00372510	3.550	0.000	****	1.00584600	1.02044900
nbcont3.L1	1.00451000	0.00196740	2.300	0.022	**	1.00066200	1.00837400
ltotcontmoyan3.L1	1.00463800	0.00224910	2.070	0.039	**	1.00024000	1.00905600
mdegcentindn3.L1	1.39456700	0.04711930	9.840	0.000	****	1.30520700	1.49004600
mbetcentind3.L1	1.01605700	0.00880550	1.840	0.066	*	0.99894470	1.03346300
lmcliquind23.L1	1.03144600	0.00987810	3.230	0.001	***	1.01226600	1.05098900
d01Laval	1.04020600	0.03968130	1.030	0.301		0.96526830	1.12096100
d02McGill	1.00713000	0.03682920	0.190	0.846		0.93747250	1.08196400
d07Concordia	0.93348110	0.07864170	-0.820	0.414		0.79139910	1.10107100
d08SherbrookeG	1.01449100	0.05680170	0.260	0.797		0.90905260	1.13215800
d09UQAMG	0.93160480	0.04120860	-1.600	0.109		0.85423950	1.01597700
d1990	1.04551000	0.07351560	0.630	0.527		0.91091050	1.20000000
d1991	1.03667900	0.07122480	0.520	0.600		0.90607250	1.18611300
d1992	1.11597400	0.07499070	1.630	0.102		0.97826270	1.27307200
d1993	1.18163900	0.07788760	2.530	0.011	**	1.03843200	1.34459600
d1994	1.33056900	0.08499010	4.470	0.000	****	1.17399700	1.50802300
d1995	1.32470700	0.08395140	4.440	0.000	****	1.16997400	1.49990500
d1996	1.78566400	0.10876890	9.520	0.000	****	1.58471400	2.01209400
d1997	1.79291600	0.10921950	9.580	0.000	****	1.59113500	2.02028600
d1998	1.62982800	0.09946910	8.000	0.000	****	1.44608100	1.83692300
d1999	1.63183100	0.09988290	8.000	0.000	****	1.44735100	1.83982500
d2000	1.39503100	0.08675750	5.350	0.000	****	1.23494400	1.57587000
d2001	1.57125900	0.09621880	7.380	0.000	****	1.39355200	1.77162800
d2002	1.38933600	0.08625350	5.300	0.000	****	1.23016300	1.56910500
d2003	1.58895900	0.09733860	7.560	0.000	****	1.40918700	1.79166500
d2004	1.51948100	0.09390410	6.770	0.000	****	1.34614200	1.71514000
d2005	1.77312900	0.10836280	9.370	0.000	****	1.57296900	1.99876000
				*	$p < 0.1$		
				**	$p < 0.05$		
				***	$p < 0.01$		
				****	$p < 0.001$		

Tableau 4.33 Régressions pour le modèle B11_6AIF

nbarticleB	IRR	Std. Err.	z	P>|z|	*p*-value level	[95% Conf. Interval]	
nbsubvAI5.L1	1.00295800	0.00363280	0.820	0.4150		0.99586340	1.01010400
nbsubvF5.L1	1.00448300	0.00089010	5.050	0.0000	****	1.00273900	1.00622900
ltotsubvAImoyan5.L1	1.00666700	0.00267260	2.500	0.0120	**	1.00144200	1.01191900
ltotsubvFmoyan5.L1	1.01405500	0.00378710	3.740	0.0000	****	1.00666000	1.02150500
nbcontAI3.L1	1.01480700	0.02243480	0.660	0.5060		0.97177490	1.05974500
nbcontF3.L1	1.00436200	0.00197320	2.220	0.0270	**	1.00050200	1.00823600
ltotcontAImoyan3.L1	1.00758000	0.01379900	0.550	0.5810		0.98089460	1.03499200
ltotcontFmoyan3.L1	1.00467400	0.00225390	2.080	0.0380	**	1.00026600	1.00910100
mdegcentindn3.L1	1.39516500	0.04713840	9.860	0.0000	****	1.30576800	1.49068200
mbetcentind3.L1	1.01585700	0.00879790	1.820	0.0690	*	0.99875850	1.03324700
lmcliquind23.L1	1.03174300	0.00988830	3.260	0.0010	***	1.01254400	1.05130700
d01Laval	1.03431300	0.03956160	0.880	0.3780		0.95960900	1.11483300
d02McGill	1.01205000	0.03710650	0.330	0.7440		0.94187460	1.08745500
d07Concordia	0.93610840	0.07890470	-0.780	0.4330		0.79355720	1.10426700
d08SherbrookeG	1.01156600	0.05672840	0.210	0.8380		0.90627300	1.12909200
d09UQAMG	0.93081840	0.04127250	-1.620	0.1060		0.85334110	1.01533000
d1990	1.04660100	0.07352960	0.650	0.5170		0.91196810	1.20111100
d1991	1.03784800	0.07124920	0.540	0.5880		0.90718920	1.18732400
d1992	1.11790700	0.07507440	1.660	0.0970	*	0.98003650	1.27517300
d1993	1.18304400	0.07794980	2.550	0.0110	**	1.03971900	1.34612700
d1994	1.33408500	0.08519480	4.510	0.0000	****	1.17713300	1.51196300
d1995	1.32872900	0.08418320	4.490	0.0000	****	1.17356600	1.50440700
d1996	1.78991600	0.10903950	9.560	0.0000	****	1.58846800	2.01691100
d1997	1.79828100	0.10954280	9.630	0.0000	****	1.59590300	2.02632400
d1998	1.64037500	0.10010180	8.110	0.0000	****	1.45545800	1.84878600
d1999	1.63998100	0.10037820	8.080	0.0000	****	1.45458600	1.84900600
d2000	1.40014100	0.08706680	5.410	0.0000	****	1.23948200	1.58162400
d2001	1.57577800	0.09645420	7.430	0.0000	****	1.39763100	1.77663200
d2002	1.39195800	0.08636410	5.330	0.0000	****	1.23257500	1.57195100
d2003	1.59157700	0.09743800	7.590	0.0000	****	1.41161500	1.79448200
d2004	1.52342400	0.09410080	6.820	0.0000	****	1.34971700	1.71948700
d2005	1.77976800	0.10877320	9.430	0.0000	****	1.57885000	2.00625400
				*	$p < 0.1$		
				**	$p < 0.05$		
				***	$p < 0.01$		
				****	$p < 0.001$		

L'analyse des valeurs du « *p*-value » pour les variables du Tableau 4.33 confirme l'importance de catégoriser le financement de type « fonctionnement » par rapport au groupe « appareillage-

infrastructure ». Les résultats présentés à la section 3.3 quant à l'importance de la catégorie « fonctionnement » sont validés par la signifiance (au moins $p < 0.05$) des variables liées au financement.

Pour les subventions, la variable capturant le nombre de subventions liées au « fonctionnement » est très significative par rapport à celle capturant l'« appareillage-infrastructure » qui ne l'est pas. Cette différence peut s'expliquer par l'écart moyen de 68% présent entre le nombre de contrats et de subventions annuels qui fait en sorte que la « force » et le « sens » de cette variable est contrôlée par l'aspect « fonctionnement » tel que le témoigne la valeur du test statistique z qui est de 5.050 pour nbsubvF5 et de 0.820 pour nbsubvAI5. Pour la variable capturant le total annuel de subventions, on trouve que les deux variables (AI & F) sont significatives et ont un sens positif. Bien que la variable liée au « fonctionnement » soit plus significative, on conclut que les deux catégories doivent être présentes dans le modèle de façon à bien modéliser le financement par subventions. Il serait important de suivre l'évolution de l'écart entre les variables du modèle ltotsubvAImoyan5 et ltotsubvFmoyan5 pour les années futures car la tendance illustrée à la Figure 3.22 démontre une croissance des montants alloués pour la catégorie « appareillage-infrastructure ».

Pour les contrats, au niveau de la variable capturant le nombre de contrats on trouve le même résultat mais la signifiance et la « force » de la variable sont moins importantes, ce qui confirme nos observations précédentes quant au plus faible nombre de cas de financement par contrats. Pour les variables capturant le total du financement annuel par contrats, la variable liée au « fonctionnement » est significative tandis que celle liée à l' « appareillage-infrastructure » ne l'est pas. Ce résultat s'explique toujours par la grande différence entre les montants de certaines des catégories présentée à la section 3.3.2.

Un autre résultat intéressant est la valeur du coefficient affiché sous la forme IRR qui nous indique que l'<u>ensemble</u> des variables statistiques continues ont une influence *positive* de croissance sur le nombre de publications scientifiques. On remarque une valeur IRR plus grande pour le degré de centralité (1.39) du chercheur indiquant l'importance de la position dans le réseau pour la publication d'articles.

Au niveau des universités, comme on contrôle par l'ensemble des *autres* universités québécoises que l'École Polytechnique de Montréal/Université de Montréal, on voudrait voir l'impact d'être

affiliée à une université sur la publication d'articles scientifiques dans le domaine de la biotechnologie, par contre, ces résultats ne sont pas significatifs la valeur du « *p*-value » étant supérieure à 0.1. On peut donc en conclure que les autres variables déjà présentes dans le modèle contribuent à bien prendre en compte les variations qui pourraient exister entre les différentes universités.

Au niveau de l'impact de l'historique sur la publication d'articles scientifiques en biotechnologie, la Figure 4.7 illustre l'évolution du coefficient sous forme IRR dans le temps.

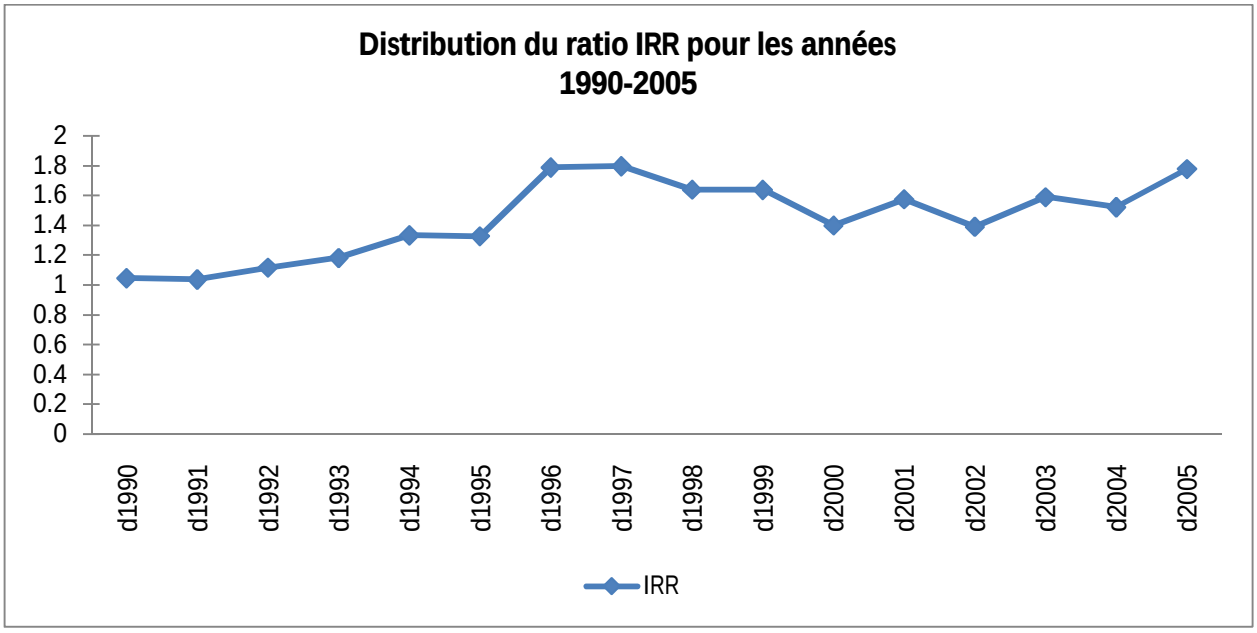


Figure 4.7 Distribution du ratio IRR pour les années

L'analyse de la Figure 4.7 nous confirme l'importance de l'historique de publication en rappelant que le fait d'avoir publié dans un passé proche aurait un impact positif croissant sur la publication car la courbe est globalement croissante. On peut faire le lien entre la variation de l'IRR par année et l'impact du financement tel que présenté à la section 3.3 dont plusieurs années ont un niveau de financement décroissant ou stagnant.

4.2.1.3 Résultats de la validation à postériori du choix du modèle

La statistique Pearson de dispersion pour le modèle final Poisson B11_6 est de 1.65844 indiquant la présence d'hyperdispersion. Ce résultat confirme donc l'approche d'utiliser un modèle binomial négatif pour la modélisation des données.

4.2.2 Résultats de corrélation

La matrice de corrélation pour le modèle B11_6 est présentée au Tableau 4.34

Tableau 4.34 Matrice de corrélation du modèle B11_6

		1	2	3	4	5	6	7	8
1	**L.nbsubv5**	1							
2	**L.ltotsubvmoyan5**	0.5334	1						
3	**L.nbcont3**	0.2747	0.1538	1					
4	**L.ltotcontmoyan3**	0.3555	0.2625	0.5339	1				
5	**L.mdegcentindn3**	0.1516	0.1227	0.0478	0.0938	1			
6	**L.mbetcentind3**	0.1113	0.0941	0.0274	0.0552	0.7638	1		
7	**L.lmcliquind23**	0.0665	0.0516	0.0312	0.0454	0.5629	0.2517	1	
8	**nbarticleB**	0.1829	0.1377	0.0669	0.1169	0.4764	0.4480	0.1880	1

La matrice de corrélation pour le modèle B11_6AIF est présentée au Tableau 4.35.

Tableau 4.35 Matrice de corrélation du modèle B11_6AIF

		1	2	3	4	5	6	7	8	9	10	11	12
1	**L.nbsubvAI5**	1											
2	**L.nbsubvF5**	0.3133	1										
3	**L.ltotsubvAImoyan5**	0.6436	0.3313	1									
4	**L.ltotsubvFmoyan5**	0.2000	0.5287	0.2818	1								
5	**L.nbcontAI3**	0.0351	0.0728	0.0414	0.0321	1							
6	**L.nbcontF3**	0.1073	0.2697	0.0819	0.1518	0.0152	1						
7	**L.ltotcontAImoyan3**	0.0715	0.0841	0.0714	0.0460	0.6061	0.0277	1					
8	**L.ltotcontFmoyan3**	0.1350	0.3549	0.1248	0.2609	0.0527	0.5321	0.0819	1				
9	**L.mdegcentindn3**	0.0179	0.1591	0.0164	0.1269	-0.0140	0.0491	-0.0164	0.0946	1			
10	**L.mbetcentind3**	0.0140	0.1166	0.0305	0.0970	-0.0081	0.0281	-0.0110	0.0559	0.7638	1		
11	**L.lmcliquind23**	-0.0124	0.0735	-0.0268	0.0557	-0.0142	0.0325	-0.0169	0.0458	0.5629	0.2517	1	
12	**nbarticleB**	0.0725	0.1827	0.0761	0.1400	0.0035	0.0669	0.0032	0.1175	0.4764	0.4480	0.1880	1

L'analyse des coefficients de corrélation du Tableau 4.35 nous confirme qu'il y a évidemment une relation entre le nombre de subventions et le total de ces subventions surtout pour la catégorie « fonctionnement ». Les mêmes relations existent pour les contrats. Une corrélation faible (0.2818) existe entre le total des subventions moyennes annuelles (horizon de 5 ans) pour la catégorie « financement » et le total des subventions moyennes annuelles (horizon de 5 ans) pour la catégorie « appareillage-infrastructure ». Ce résultat est très intéressant et confirme la

courbe présentée à la Figure 3.22 indiquant qu'une augmentation des subventions dans les catégories « appareillage » et « infrastructure » doit être accompagnée d'une augmentation de subventions pour la catégorie « financement ». Il faut donc faire fonctionner ces nouveaux appareils et ces nouvelles infrastructures.

Au niveau du degré de centralité d'un chercheur pour un horizon de 3 ans, on trouve une corrélation faible avec le nombre (0.1591) et le total annuel (0.1269) des subventions d'un horizon de 5 ans de la catégorie « fonctionnement ». Au niveau du « betweenness centrality » du chercheur, une faible corrélation existe avec le nombre de subvention « fonctionnement » et une forte corrélation est présente avec le degré de centralité (0.7638) ce qui avait déjà été détecté à section 4.2.1.1. Pour l'interconnexion des chercheurs mesurée par la « cliquishness », on note sans grande surprise qu'il existe un lien avec les deux autres mesures réseau, soit « degree centrality » (0.5629) et « betweenness centrality » (0.2517).

Ces résultats nous permettent confirmer que cette première version du modèle semble bien expliquer les données tout en étant robuste aux variations de variables présentes.

CONCLUSION

En guise de conclusion, commençons par revoir les étapes réalisées pour ce mémoire. Le CHAPITRE 1 nous a permis de faire état de la littérature en lien avec notre sujet d'étude qui est l'« *évaluation de l'incidence du financement public de la recherche universitaire québécoise dans le domaine de la biotechnologie* ». Ce premier chapitre a amené la formulation d'hypothèses présentées au CHAPITRE 2 en même temps que la méthodologie utilisée pour obtenir une matrice d'analyse. Le CHAPITRE 3 a présenté des caractéristiques et statistiques élémentaires permettant d'avoir un aperçu global des données avant de passer à l'analyse et aux résultats présentés au CHAPITRE 4.

Depuis le début, nous connaissions l'ampleur du projet et l'objectif était la construction d'un cadre d'analyse pouvant faciliter les travaux futurs tout en présentant les résultats des analyses et le modèle développé. Des outils d'automatisation ont été développés justement pour répondre à la complexité d'intégration des données et à la génération des mesures.

Des hypothèses avaient été énoncées à la section 2.1, voyons maintenant comment, à la lumière des résultats, nous pouvons les valider.

Pour la première hypothèse, soit :

Les chercheurs financés par les organismes subventionnaires ont une plus grande propension à collaborer par l'intermédiaire de publications scientifiques et ils ont, par le fait même, un réseau d'innovation plus étendu, et contribuent de façon plus active aux flux de connaissance du réseau;

Les données présentées à la section 3.2.2 nous indiquent que la collaboration est à la hausse depuis le début des années 1980 dans le domaine de la biotechnologie. Le nombre de chercheurs isolés du réseau de collaboration est à la baisse tandis que la super composante, représentant l'ensemble des chercheurs connectés, suit une courbe de croissance. Lier le financement à ces observations est évidemment nécessaire pour expliquer les résultats mais très *complexe*. Il est évident que les différentes catégories de financement soit l'appareillage, le fonctionnement et l'infrastructure auront des impacts différents sur le réseau mais contribueront toutes à créer un contexte de recherche favorable à la collaboration. Il faut comprendre que la nature complexe de

la biotechnologie force les rapprochements entre chercheurs et, naturellement, il y a regroupement autour de ceux qui reçoivent le plus de financement dans plusieurs catégories. D'autres variables pourraient aider à expliquer le niveau de publication d'un chercheur. Les revues universitaires en sont un type mais on pourrait facilement penser à la participation des chercheurs à différents colloques favorisant aussi les échanges et la création de liens pouvant aussi mener à des publications. L'importance des universités pour la publication d'articles est une évidence connue justement parce qu'elles sont des plateformes pouvant recevoir et gérer, non seulement, des infrastructures et de l'équipement de recherche mais les faire fonctionner dans un contexte où une multitude d'équipes doivent partager les ressources. Encore une fois, on peut noter que cette collaboration « forcée » des infrastructures et des appareillages peut avoir un impact positif sur la création de liens entre chercheurs et est indirectement liée au financement reçu dans le passé. La collaboration entre chercheurs dépasse les murs de l'établissement où le chercheur est affilié et a une portée nationale et internationale. La notion de réseau étendu a été présentée à travers la géographie des publications. La croissance des collaborations internationales est loin d'être un phénomène éphémère et celles-ci sont devenues une nécessité pour aller chercher une expertise de plus en plus spécialisée.

Pour la deuxième hypothèse, soit :

Les chercheurs financés par les organismes subventionnaires produisent plus d'articles en collaboration avec d'autres chercheurs universitaires, et de par leur position centrale dans les réseaux d'innovation contribuent à transmettre la connaissance à travers le milieu universitaire;

Les données analysées nous permettent d'avancer qu'un nombre croissant d'articles sont publiés annuellement dans le domaine de la biotechnologie. L'augmentation de la publication d'articles scientifiques peut être expliquée, en partie, par l'augmentation du financement reçu comme semble l'indiquer nos résultats. Bien que le laissaient entendre nos résultats initiaux, aucune évidence d'impact négatif ne peut lier le nombre de contrats ou les montants des contrats reçus à une décroissance de la publication d'articles. Nous avons aussi examiné la proportion de contrats par rapport au montant final et n'avons pas trouvé de valeur significative. Il n'y aurait donc pas de frein à la publication due à l'octroi de contrats. Dans des recherches futures, il serait important d'examiner s'il n'y a pas d'effets disciplinaires qui peuvent être décelés.

De plus, à travers le modèle final, un effet positif est présent entre le nombre et le total des contrats sur une période de trois ans ou des subventions sur une période de cinq ans et le nombre d'articles publiés au cours de l'année. Au niveau de sa position dans le réseau, les résultats indiquent que plus un chercheur a des liens directs avec d'autres chercheurs, plus il sera porté à publier des articles scientifiques. Ce qui est cohérent avec le concept de performance lié au « small world ». La position centrale est souvent liée aux « star scientists » qui réussissent à aller chercher beaucoup de contrats et subventions mais redistribuent l'argent en s'assurant d'être co-auteur des articles publiés. Quelques exemples précis de « vedettes » du financement ont été observés dans les réseaux construits.

Les résultats démontrent que les subventions de fonctionnement auront un impact direct fort tandis que les subventions d'infrastructure auront une contribution positive mais pas aussi forte que les subventions de fonctionnement. Les mêmes résultats sont observés pour les contrats mais de façon moins significative.

Les résultats obtenus présentent l'évolution annuelle du financement reçu par les universités et ceux-ci permettent de lier l'importance du nombre de publications à la position plus centrale qu'occupe les chercheurs affiliés aux universités québécoises. Plusieurs questions se posent face à cette situation, par exemple, est-ce qu'on peut parler de financement récursif et toujours lié aux mêmes chercheurs financés par années ? On peut aussi se demander quelle est la proportion du financement de la recherche qui va aux chercheurs qui produisent des résultats de valeur académique moyenne mais de grande valeur pour l'industrie ? Les résultats présentés ne permettent pas de capturer complètement la complexité du lien entre le financement et la publication d'articles scientifiques. On peut donc expliquer partiellement le nombre de publications générées par chercheur. Les résultats démontrent aussi que le taux de publication d'articles scientifique en biotechnologie suit une courbe de croissance beaucoup plus élevée que celle du financement et que donc le financement seulement ne pouvait expliquer le nombre plus élevé de publications. Il faut donc se tourner vers la combinaison du financement mais aussi de l'impact du réseau. Un chercheur ne deviendra pas plus productif du jour au lendemain. Par contre, produire plus d'articles en étant co-auteur dans le cadre de plusieurs projets de recherche est une réalité observée. Cette situation explique, en partie, qu'un nombre croissant de chercheurs publient plus. Notre étude a porté principalement sur la transmission de la connaissance sous forme d'articles scientifiques mais qu'en est-il de l'impact du financement sur la capacité

d'absorption de ces connaissances par le milieu universitaire. Pour que la connaissance ait une certaine valeur il faut être en mesure de la capturer.

La contribution de ce travail de recherche se divise en deux volets. Premièrement, des outils logiciels ont été développés pour permettre l'automatisation de plusieurs étapes complexes de génération de statistiques et d'intégration de plusieurs sources de données. Ces étapes étant critiques à l'analyse des données. Deuxièmement, un premier modèle statistique a été proposé pour mesurer l'impact du financement reçu par un chercheur universitaire sur la publication d'articles scientifiques. Le modèle démontre l'importance de ventiler le financement reçu selon deux catégories soit l' « appareillage-infrastructure » et le fonctionnement.

Ce mémoire s'inscrit dans le cadre d'un projet d'*envergure* et ouvre la porte à des analyses beaucoup plus détaillées. Celles-ci seront possibles en améliorant la qualité des données actuelles et la quantité de celles-ci avec des mots-clés plus complets, en ajoutant des sources de données et en capturant les subtilités du financement, du réseau que de l'environnement. Par exemple, un troisième type de financement nommé « bourses d'excellence » est utilisé par Statistiques Canada dans ses rapports (*Science Statistics*, 2009). Ce type de financement visant exclusivement les étudiants n'est pas présent dans nos données mais devrait être incorporé dans des analyses futures. Lorsqu'on parle d'étudiants, jusqu'à présent c'est uniquement les chercheurs qui étaient présents, l'ajout de données du financement des étudiants permettra d'étendre l'analyse au niveau du Canada. Lorsqu'on parle de nouvelles mesures, l'analyse de la composante liée au chercheur permettrait de mieux saisir l'impact de son environnement mais cela implique évidemment une nouvelle phase de programmation pour l'automatisation du processus. L'ajout des données liées aux brevets enrichira le modèle et permettra d'expliquer une partie de l'utilisation des ressources financières et matérielles.

Pour terminer, voici une liste sommaire de questions qui pourraient être étudiées dans les recherches futures:

- Étudier l'incidence du financement, selon la catégorie du récipiendaire (étudiant, professeur, etc.), sur la publication d'articles scientifiques;
- Étudier l'évolution temporelle des affiliations des chercheurs financés;

- Proposer un modèle évaluant le rendement des chercheurs en fonction du financement reçu.
- Étudier l’effet disciplinaire des sous-domaines de la biotechnologie sur le financement reçu en lien avec les publications générées.

BIBLIOGRAPHIE

Aharonson, B. S., Baum, J. A. C., & Feldman, M. P. (2004). *Industrial clustering and the returns to inventive activity: Canadian biotechnology firms, 1991-2000*: Department of Business Studies, Univ.

Aharonson, B. S., Baum, J. A. C., & Plunket, A. (2008). Inventive and uninventive clusters: The case of Canadian biotechnology. *Research Policy, 37*(6-7), 1108-1131.

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative science quarterly*, 425-455.

Arora, A., David, P. A., & Gambardella, A. (1998). Reputation and competence in publicly funded science: estimating the effects on research group productivity. *Annales d'Economie et de Statistique*, 163-198.

Avenel, E., Favier, A. V., Ma, S., Mangematin, V., & Rieu, C. (2007). Diversification and hybridization in firm knowledge bases in nanotechnologies. *Research Policy, 36*(6), 864-870.

Bavelas, A. (1947). A mathematical model for group structures. *Human organization, 7*(3), 16-30.

Blume-Kohout, M., Kumar, K., & Sood, N. (2009). Federal Life Sciences Funding and University R&D. *NBER Working Paper*.

Bonaccorsi, A., & Thoma, G. (2007). Institutional complementarity and inventive performance in nano science and technology. *Research Policy, 36*(6), 813-831.

Bozeman, B., Laredo, P., & Mangematin, V. (2007). Understanding the emergence and deployment of “nano” S&T. *Research Policy, 36*(6), 807-812.

Burt, R. S. (2001). Bandwidth and echo: trust, information, and gossip in social networks. *Networks and markets*, 30–74.

Cantner, U., & Graf, H. (2006). The network of innovators in Jena: An application of social network analysis. *Research Policy, 35*(4), 463-480.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: a new perspective on learning and innovation. *Administrative science quarterly, 35*(1).

Cohn, B. S., & Marriott, M. K. (1958). Networks and centres in the integration of Indian civilization. *Journal of social research, 1*(1), 1-9.

Cowan, R., & Jonard, N. (2004). Network structure and the diffusion of knowledge. *Journal of economic Dynamics and Control, 28*(8), 1557-1575.

Daniel, H. Z., Hempel, D. J., & Srinivasan, N. (2003). Project selection A process analysis. *Industrial Marketing Management, 32*(1), 39-54.

Darby, M. R., & Zucker, L. G. (2003). Grilichesian breakthroughs: inventions of methods of inventing and firm entry in nanotechnology. *NBER Working Paper*.

Dasgupta, P., & David, P. A. (2002). Toward a new economics of science. *Science bought and sold: essays in the economics of science*, 219.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40*(1), 35-41.

Freeman, L. C. (1979). *The networkers network: A study of the impact of a new communications medium on sociometric structure*: School of Social Sciences Univ. of Calif.

Gittelman, M., & Kogut, B. (2003). Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logic of citation patterns. *MANAGEMENT SCIENCE, 49*(4), 366-382.

Goldfarb, B. (2008). The effect of government contracting on academic research: Does the source of funding affect scientific output? *Research Policy, 37*(1), 41-58.

Guan, J., & Ma, N. (2007). China's emerging presence in nanoscience and nanotechnology A comparative bibliometric study of several nanoscience 'giants'. *Research Policy, 36*(6), 880-886.

Gulati, R., & Gargiulo, M. (1999). Where Do Interorganizational Networks Come From? 1. *American journal of sociology, 104*(5), 1398-1438.

Hagedoorn, J., Link, A. N., & Vonortas, N. S. (2000). Research partnerships. *Research Policy, 29*(4-5), 567-586.

Hall, B. H., Link, A. N., & Scott, J. T. (2003). Universities as research partners. *Review of Economics and Statistics, 85*(2), 485-491.

Hart, D. M. (2001). Antitrust and technological innovation in the US: ideas, institutions, decisions, and impacts, 1890–2000. *Research Policy, 30*(6), 923-936.

Hausman, J., Hall, B. H., & Griliches, Z. (1984). Econometric models for count data with an application to the patents-R & D relationship. *Econometrica, 52*(4), 909-938.

Hilbe, J. (2007). *Negative binomial regression*: Cambridge University Press Cambridge.

Hsiao, C. (2003). *Analysis of panel data*: Cambridge Univ Pr.

Jacob, B., & Lefgren, L. (2007). The impact of research grant funding on scientific productivity: NBER.

Link, A. N., & Scott, J. T. (2004). The Role of Public Research Institutions in a National Innovation System: An Economic Perspective. 53.

Mackenzie, K. D. (1966). Structural centrality in communications networks. *Psychometrika, 31*(1), 17-25.

Mogoutov, A., & Kahane, B. (2007). Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking. *Research Policy, 36*(6), 893-903.

Moxley, R. L., & Moxley, N. F. (1974). Determining point-centrality in uncontrived social networks. *Sociometry*, 122-130.

Munn-Venn, T. (2005). *Biotechnology in Canada A Technology Platform for Growth*: The Conference Board of Canada.

Narula, R., & Duysters, G. (2004). Globalisation and trends in international R&D alliances. *Journal of International management, 10*(2), 199-218.

Newman, M. E. J. (2001). Clustering and preferential attachment in growing networks. *Physical Review E, 64*(2), 25102.

Payne, A. A., & Siow, A. (2003). Does federal research funding increase university research output? *Advances in Economic Analysis & Policy, 3*(1), 1-22.

Porter, A. L., Youtie, J., Shapira, P., & Schoeneck, D. J. (2008). Refining search terms for nanotechnology. *Journal of nanoparticle research, 10*(5), 715-728.

Porter, M. E. (1998). Clusters and the new economics of competition. *Harvard business review, 76*(6), 77-90.

Robinson, D. K. R., Rip, A., & Mangematin, V. (2007). Technological agglomeration and the emergence of clusters and networks in nanotechnology. *Research Policy, 36*(6), 871-879.

Schiffauerova, A., & Beaudry, C. Innovation Networks and Gatekeepers of Canadian Biotechnology Clusters.

Schilling, M. A., & Phelps, C. C. (2007). Interfirm collaboration networks: The impact of large-scale network structure on firm innovation. *MANAGEMENT SCIENCE, 53*(7), 1113-1126.

Science Statistics. (2008). Statistics Canada.

Science Statistics. (2009). Statistics Canada.

Stephan, P., Black, G. C., & Chang, T. (2007). The small size of the small scale market: The early-stage labor market for highly skilled nanotechnology workers. *Research Policy, 36*(6), 887-892.

Stuart, T. E. (2000). Interorganizational alliances and the performance of firms: A study of growth and innovation rates in a high-technology industry. *Strategic management journal*, 791-811.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*: Cambridge Univ Pr.

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

Wouter, d. N., Andrej, M., & Vladimir, B. (2004). *Exploratory Social Network Analysis with Pajek*: Cambridge University Press.

Zucker, L. G., Darby, M. R., & Armstrong, J. S. (2002). Commercializing knowledge: university science, knowledge capture, and firm performance in biotechnology. *MANAGEMENT SCIENCE, 48*(1), 138-153.

Zucker, L. G., Darby, M. R., & Brewer, M. B. (1998). Intellectual human capital and the birth of US biotechnology enterprises. *The American Economic Review, 88*(1), 290-306.

Zucker, L. G., Darby, M. R., Furner, J., Liu, R. C., & Ma, H. (2007). Minerva unbound: Knowledge stocks, knowledge flows and new knowledge production. *Research Policy, 36*(6), 850-863.

ANNEXES

ANNEXE A - AUTOMATISATION

Voici un exemple de fichier de macro Pajek généré dynamiquement par l'outil d'automatisation.

```
NETBEGIN 1
CLUBEGIN 1
PERBEGIN 1
CLSBEGIN 1
HIEBEGIN 1
VECBEGIN 1
N 1 RDN "C:\BIO\Networks\BIO_network.NET"
N 2 TIMEINTXXXXXX 1 [1985,1985,1]
N 3 TIMEINTXXXXXX 1 [1986,1986,1]
N 4 TIMEINTXXXXXX 1 [1987,1987,1]
N 5 TIMEINTXXXXXX 1 [1988,1988,1]
N 6 TIMEINTXXXXXX 1 [1989,1989,1]
N 7 TIMEINTXXXXXX 1 [1990,1990,1]
N 8 TIMEINTXXXXXX 1 [1991,1991,1]
N 9 TIMEINTXXXXXX 1 [1992,1992,1]
N 10 TIMEINTXXXXXX 1 [1993,1993,1]
N 11 TIMEINTXXXXXX 1 [1994,1994,1]
N 12 TIMEINTXXXXXX 1 [1995,1995,1]
N 13 TIMEINTXXXXXX 1 [1996,1996,1]
N 14 TIMEINTXXXXXX 1 [1997,1997,1]
N 15 TIMEINTXXXXXX 1 [1998,1998,1]
N 16 TIMEINTXXXXXX 1 [1999,1999,1]
N 17 TIMEINTXXXXXX 1 [2000,2000,1]
N 18 TIMEINTXXXXXX 1 [2001,2001,1]
N 19 TIMEINTXXXXXX 1 [2002,2002,1]
N 20 TIMEINTXXXXXX 1 [2003,2003,1]
N 21 TIMEINTXXXXXX 1 [2004,2004,1]
N 22 TIMEINTXXXXXX 1 [2005,2005,1]
N 23 AFFIL 2 TRUE FALSE FALSE
N 24 AFFIL 3 TRUE FALSE FALSE
N 25 AFFIL 4 TRUE FALSE FALSE
N 26 AFFIL 5 TRUE FALSE FALSE
N 27 AFFIL 6 TRUE FALSE FALSE
N 28 AFFIL 7 TRUE FALSE FALSE
N 29 AFFIL 8 TRUE FALSE FALSE
N 30 AFFIL 9 TRUE FALSE FALSE
N 31 AFFIL 10 TRUE FALSE FALSE
N 32 AFFIL 11 TRUE FALSE FALSE
N 33 AFFIL 12 TRUE FALSE FALSE
N 34 AFFIL 13 TRUE FALSE FALSE
N 35 AFFIL 14 TRUE FALSE FALSE
N 36 AFFIL 15 TRUE FALSE FALSE
N 37 AFFIL 16 TRUE FALSE FALSE
N 38 AFFIL 17 TRUE FALSE FALSE
N 39 AFFIL 18 TRUE FALSE FALSE
N 40 AFFIL 19 TRUE FALSE FALSE
```

```
N 41 AFFIL 20 TRUE FALSE FALSE
N 42 AFFIL 21 TRUE FALSE FALSE
N 43 AFFIL 22 TRUE FALSE FALSE
Msg 1985 [START]
N 23 WN "C:\BIO\Statistics\betweenness\1_year_1985.net"
V 1 BETWEEN 23
V 1 WV "C:\BIO\Statistics\betweenness\1_year_1985_betweenness_centrality.vec"
Msg 1985 [END]
Msg 1986 [START]
N 24 WN "C:\BIO\Statistics\betweenness\1_year_1986.net"
V 2 BETWEEN 24
V 2 WV "C:\BIO\Statistics\betweenness\1_year_1986_betweenness_centrality.vec"
Msg 1986 [END]
Msg 1987 [START]
N 25 WN "C:\BIO\Statistics\betweenness\1_year_1987.net"
V 3 BETWEEN 25
V 3 WV "C:\BIO\Statistics\betweenness\1_year_1987_betweenness_centrality.vec"
Msg 1987 [END]
Msg 1988 [START]
N 26 WN "C:\BIO\Statistics\betweenness\1_year_1988.net"
V 4 BETWEEN 26
V 4 WV "C:\BIO\Statistics\betweenness\1_year_1988_betweenness_centrality.vec"
Msg 1988 [END]
Msg 1989 [START]
N 27 WN "C:\BIO\Statistics\betweenness\1_year_1989.net"
V 5 BETWEEN 27
V 5 WV "C:\BIO\Statistics\betweenness\1_year_1989_betweenness_centrality.vec"
Msg 1989 [END]
Msg 1990 [START]
N 28 WN "C:\BIO\Statistics\betweenness\1_year_1990.net"
V 6 BETWEEN 28
V 6 WV "C:\BIO\Statistics\betweenness\1_year_1990_betweenness_centrality.vec"
Msg 1990 [END]
Msg 1991 [START]
N 29 WN "C:\BIO\Statistics\betweenness\1_year_1991.net"
V 7 BETWEEN 29
V 7 WV "C:\BIO\Statistics\betweenness\1_year_1991_betweenness_centrality.vec"
Msg 1991 [END]
Msg 1992 [START]
N 30 WN "C:\BIO\Statistics\betweenness\1_year_1992.net"
V 8 BETWEEN 30
V 8 WV "C:\BIO\Statistics\betweenness\1_year_1992_betweenness_centrality.vec"
Msg 1992 [END]
Msg 1993 [START]
N 31 WN "C:\BIO\Statistics\betweenness\1_year_1993.net"
V 9 BETWEEN 31
V 9 WV "C:\BIO\Statistics\betweenness\1_year_1993_betweenness_centrality.vec"
Msg 1993 [END]
Msg 1994 [START]
N 32 WN "C:\BIO\Statistics\betweenness\1_year_1994.net"
V 10 BETWEEN 32
V 10 WV "C:\BIO\Statistics\betweenness\1_year_1994_betweenness_centrality.vec"
Msg 1994 [END]
```

```
Msg 1995 [START]
N 33 WN "C:\BIO\Statistics\betweenness\1_year_1995.net"
V 11 BETWEEN 33
V 11 WV "C:\BIO\Statistics\betweenness\1_year_1995_betweenness_centrality.vec"
Msg 1995 [END]
Msg 1996 [START]
N 34 WN "C:\BIO\Statistics\betweenness\1_year_1996.net"
V 12 BETWEEN 34
V 12 WV "C:\BIO\Statistics\betweenness\1_year_1996_betweenness_centrality.vec"
Msg 1996 [END]
Msg 1997 [START]
N 35 WN "C:\BIO\Statistics\betweenness\1_year_1997.net"
V 13 BETWEEN 35
V 13 WV "C:\BIO\Statistics\betweenness\1_year_1997_betweenness_centrality.vec"
Msg 1997 [END]
Msg 1998 [START]
N 36 WN "C:\BIO\Statistics\betweenness\1_year_1998.net"
V 14 BETWEEN 36
V 14 WV "C:\BIO\Statistics\betweenness\1_year_1998_betweenness_centrality.vec"
Msg 1998 [END]
Msg 1999 [START]
N 37 WN "C:\BIO\Statistics\betweenness\1_year_1999.net"
V 15 BETWEEN 37
V 15 WV "C:\BIO\Statistics\betweenness\1_year_1999_betweenness_centrality.vec"
Msg 1999 [END]
Msg 2000 [START]
N 38 WN "C:\BIO\Statistics\betweenness\1_year_2000.net"
V 16 BETWEEN 38
V 16 WV "C:\BIO\Statistics\betweenness\1_year_2000_betweenness_centrality.vec"
Msg 2000 [END]
Msg 2001 [START]
N 39 WN "C:\BIO\Statistics\betweenness\1_year_2001.net"
V 17 BETWEEN 39
V 17 WV "C:\BIO\Statistics\betweenness\1_year_2001_betweenness_centrality.vec"
Msg 2001 [END]
Msg 2002 [START]
N 40 WN "C:\BIO\Statistics\betweenness\1_year_2002.net"
V 18 BETWEEN 40
V 18 WV "C:\BIO\Statistics\betweenness\1_year_2002_betweenness_centrality.vec"
Msg 2002 [END]
Msg 2003 [START]
N 41 WN "C:\BIO\Statistics\betweenness\1_year_2003.net"
V 19 BETWEEN 41
V 19 WV "C:\BIO\Statistics\betweenness\1_year_2003_betweenness_centrality.vec"
Msg 2003 [END]
Msg 2004 [START]
N 42 WN "C:\BIO\Statistics\betweenness\1_year_2004.net"
V 20 BETWEEN 42
V 20 WV "C:\BIO\Statistics\betweenness\1_year_2004_betweenness_centrality.vec"
Msg 2004 [END]
Msg 2005 [START]
N 43 WN "C:\BIO\Statistics\betweenness\1_year_2005.net"
V 21 BETWEEN 43
```

```
V 21 WV "C:\BIO\Statistics\betweenness\1_year_2005_betweenness_centrality.vec"
```

ANNEXE B - TYPE DE FINANCEMENT SIRU

SIRU_Financing_Type		
SIRU_Type_Financement	**SIRU_Type_Financement_Detail**	**Categorie**
Contrat	Contrats pour appareillage (incluant les dons d'équipement)	contrat-appareillage
Contrat	Contrats pour dépenses courantes	contrat-fonctionnement
Contrat	Frais directs de recherche	contrat-fonctionnement
Contrat	Frais indirects de recherche (fonctionnement)	contrat-fonctionnement
Contrat	Frais indirects de recherche (immobilisation)	contrat-infrastructure
Contrat	Gros équipements (plus de 7000 $)	contrat-appareillage
Contrat	Soins aux patients	contrat-fonctionnement
Contrat	Sous-traitance à l'extérieur (au-delà de 25000 $)	contrat-fonctionnement
Contrats	01-Contrats pour les dépenses courantes	contrat-fonctionnement
Contrats	02-Contrats d'appareillage	contrat-appareillage
Fonds internes de recherche	Frais directs de recherche	subvention-fonctionnement
Fonds internes de recherche	Frais indirects de recherche (fonctionnement)	subvention-fonctionnement
Fonds internes de	Frais indirects de recherche (immobilisation)	subvention-

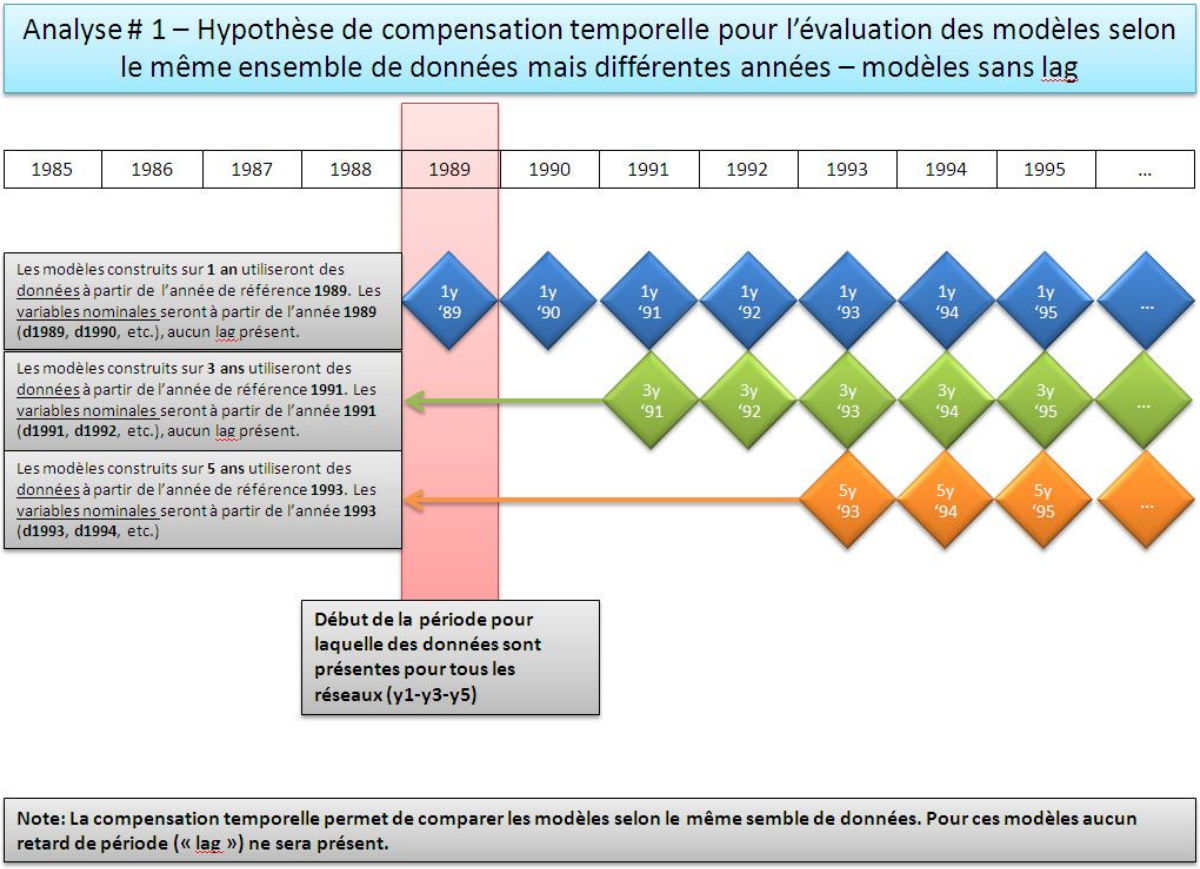
SIRU_Financing_Type		
SIRU_Type_Financement	**SIRU_Type_Financement_Detail**	**Categorie**
recherche		infrastructure
Fonds internes de recherche	Gros équipements (plus de 7000 $)	subvention-appareillage
Subvention	Autres subventions	subvention-fonctionnement
Subvention	Fonds interne de recherche (subvention générale)	subvention-fonctionnement
Subvention	Frais directs de recherche	subvention-fonctionnement
Subvention	Frais indirects de recherche (fonctionnement)	subvention-fonctionnement
Subvention	Frais indirects de recherche (immobilisation)	subvention-infrastructure
Subvention	Gros équipements (plus de 7000 $)	subvention-appareillage
Subvention	Infrastructure	subvention-infrastructure
Subvention	Sous-traitance à l'extérieur (au-delà de 25000 $)	subvention-fonctionnement
Subvention	Subventions de partenariat	subvention-fonctionnement

SIRU_Financing_Type		
SIRU_Type_Financement	**SIRU_Type_Financement_Detail**	**Categorie**
Subvention	Subventions du programme SYNERGIE	subvention-infrastructure
Subvention	Subventions FCI	subvention-infrastructure
Subvention	Subventions pour appareillage	subvention-appareillage
Subvention	Subventions pour dépenses courantes	subvention-fonctionnement
Subvention	Subventions pour frais indirects	subvention-fonctionnement
Subvention	Valorisation et commercialisation	subvention-fonctionnement
Subvention de partenariat	Frais directs de recherche	subvention-fonctionnement
Subvention de partenariat	Frais indirects de recherche (fonctionnement)	subvention-fonctionnement
Subvention de partenariat	Frais indirects de recherche (immobilisation)	subvention-infrastructure
Subvention de partenariat	Gros équipements (plus de 7000 $)	subvention-appareillage
Subvention de partenariat	Sous-traitance à l'extérieur (au-delà de 25000 $)	subvention-fonctionnement

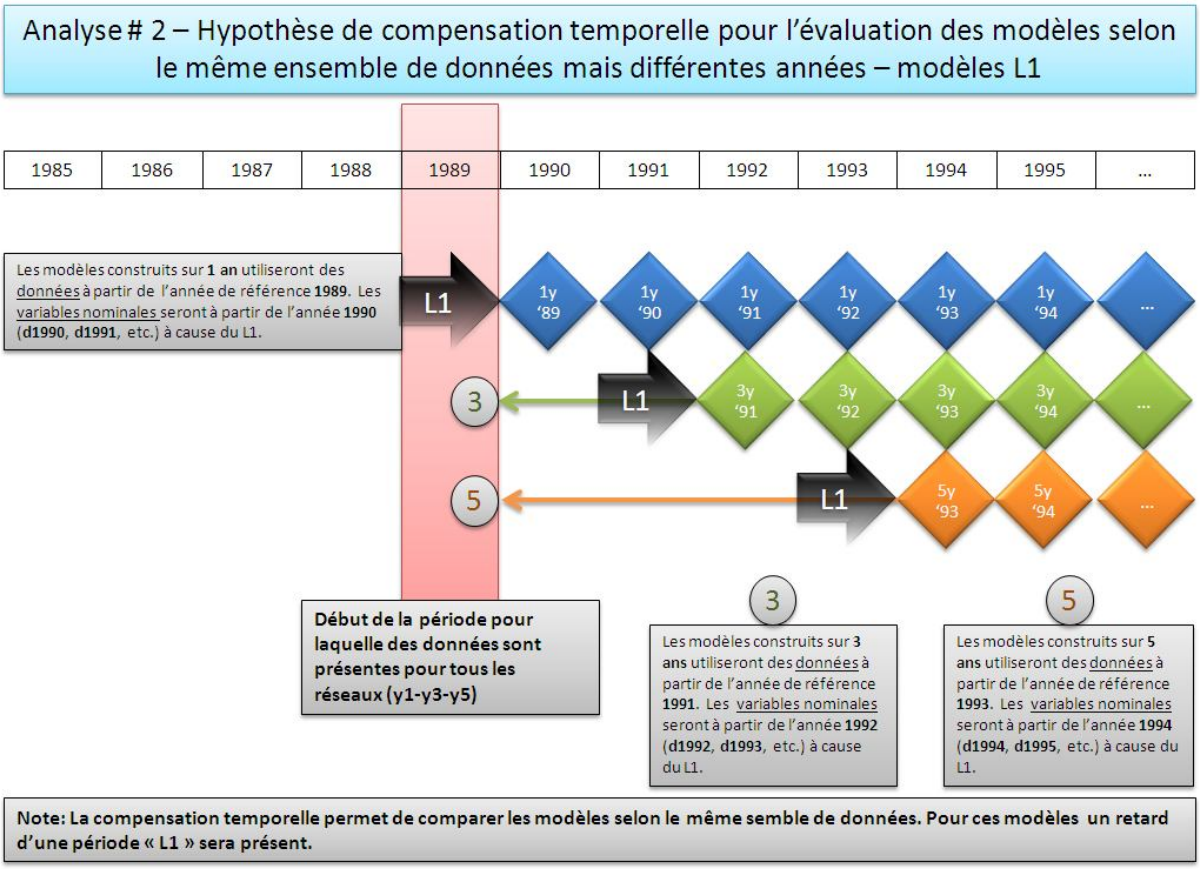
SIRU_Financing_Type		
SIRU_Type_Financement	**SIRU_Type_Financement_Detail**	**Categorie**
Subvention reliée à la FCI	Frais directs de recherche	subvention-infrastructure
Subvention reliée à la FCI	Frais indirects de recherche (fonctionnement)	subvention-fonctionnement
Subvention reliée à la FCI	Frais indirects de recherche (immobilisation)	subvention-infrastructure
Subvention reliée à la FCI	Gros équipements (plus de 7000 $)	subvention-appareillage
Subventions	03-Subventions pour les dépenses courantes	subvention-fonctionnement
Subventions	04-Subventions d'appareillage	subvention-appareillage
Subventions	05-Subventions de partenariat	subvention-fonctionnement
Subventions	06-Subventions reliées à la FCI	subvention-infrastructure
Subventions	07-Subventions pour les frais indirects	subvention-fonctionnement
Subventions	10-Subventions générales	subvention-fonctionnement
Subventions	11-Autres subventions	subvention-fonctionnement

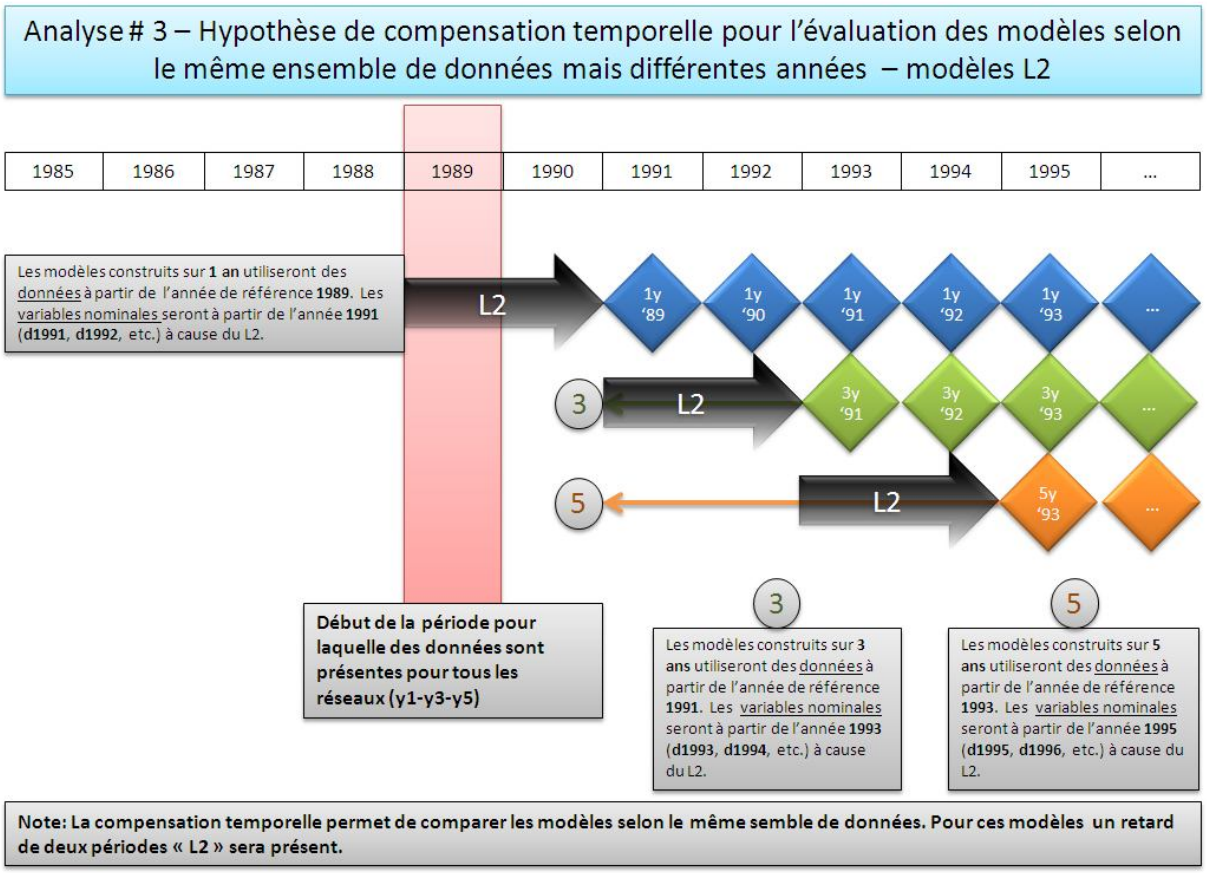
SIRU_Financing_Type		
SIRU_Type_Financement	**SIRU_Type_Financement_Detail**	**Categorie**
Subventions	12-Subventions d'infrastructure	subvention-infrastructure

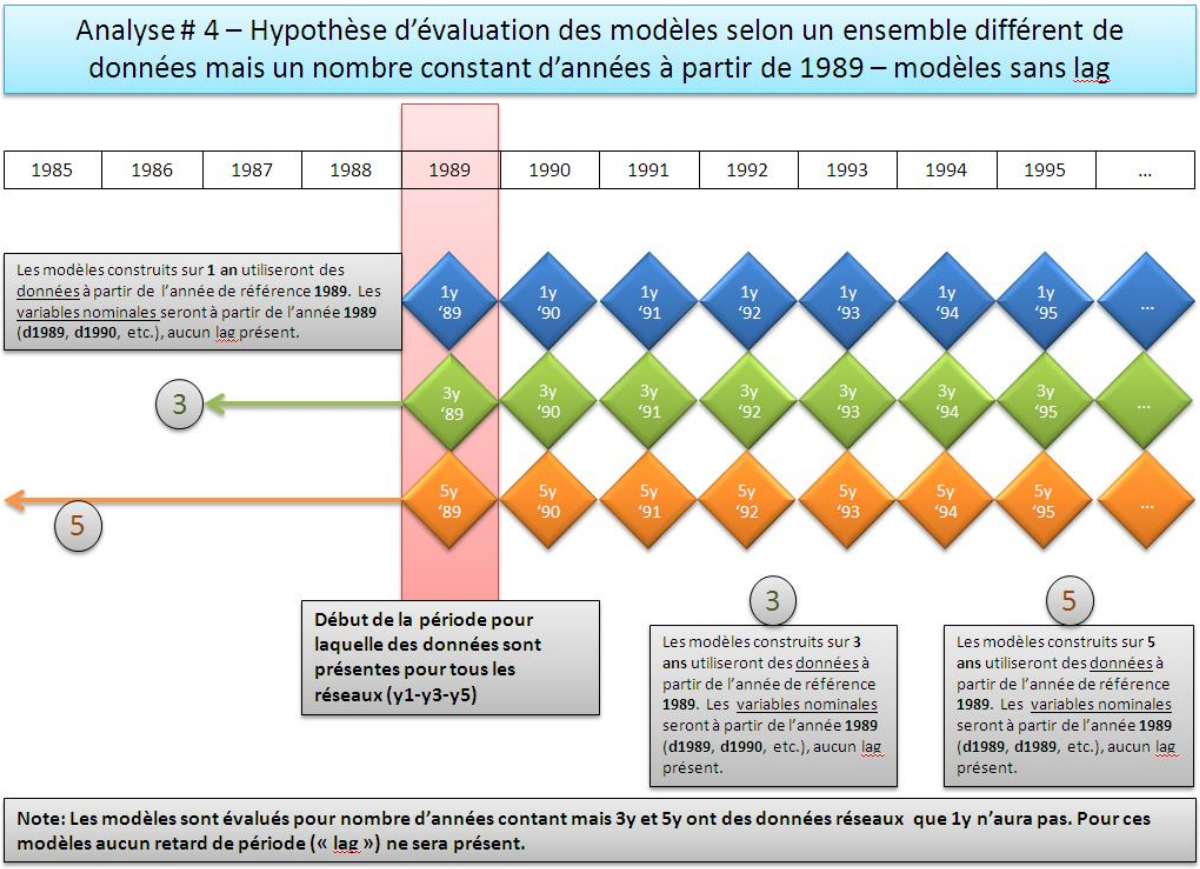
ANNEXE C - MODÈLES D'ANALYSE DE RÉGRESSION

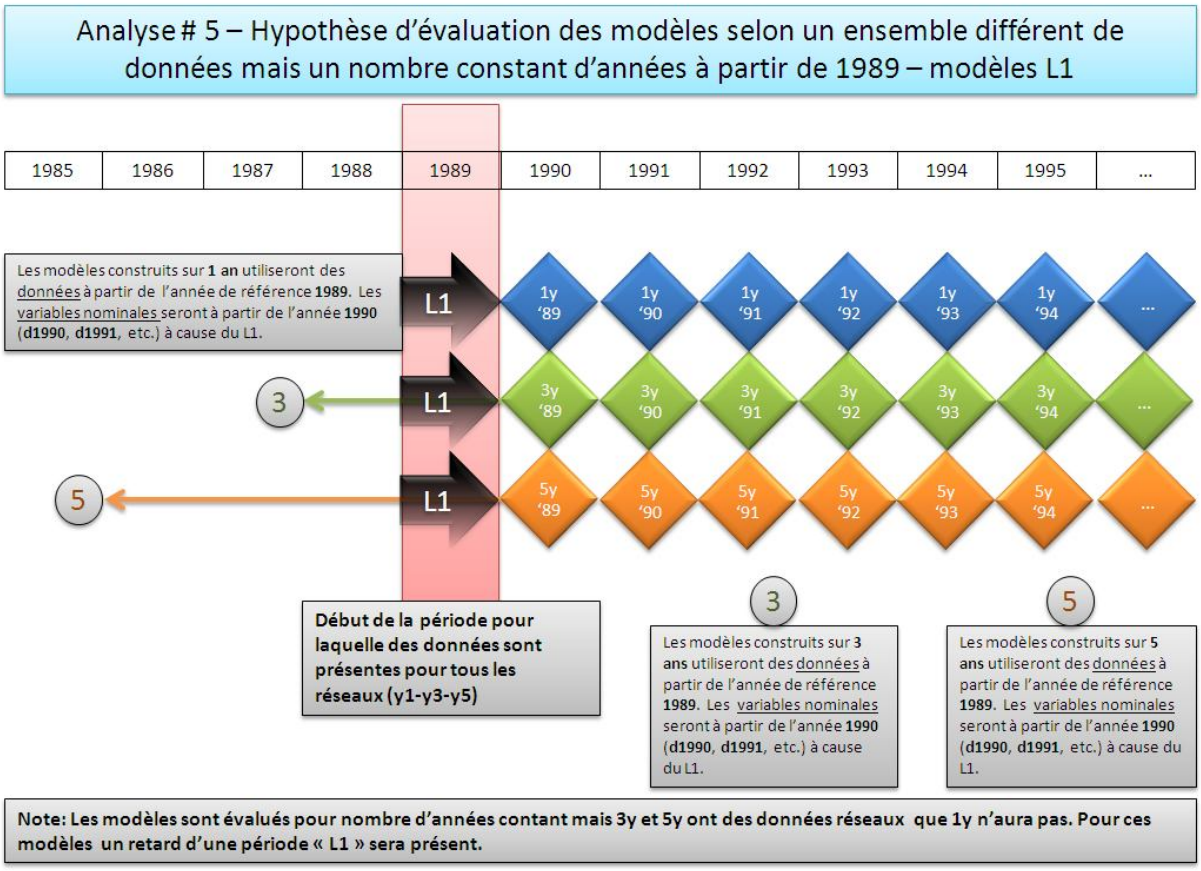


Note: La compensation temporelle permet de comparer les modèles selon le même semble de données. Pour ces modèles aucun retard de période (« lag ») ne sera présent.

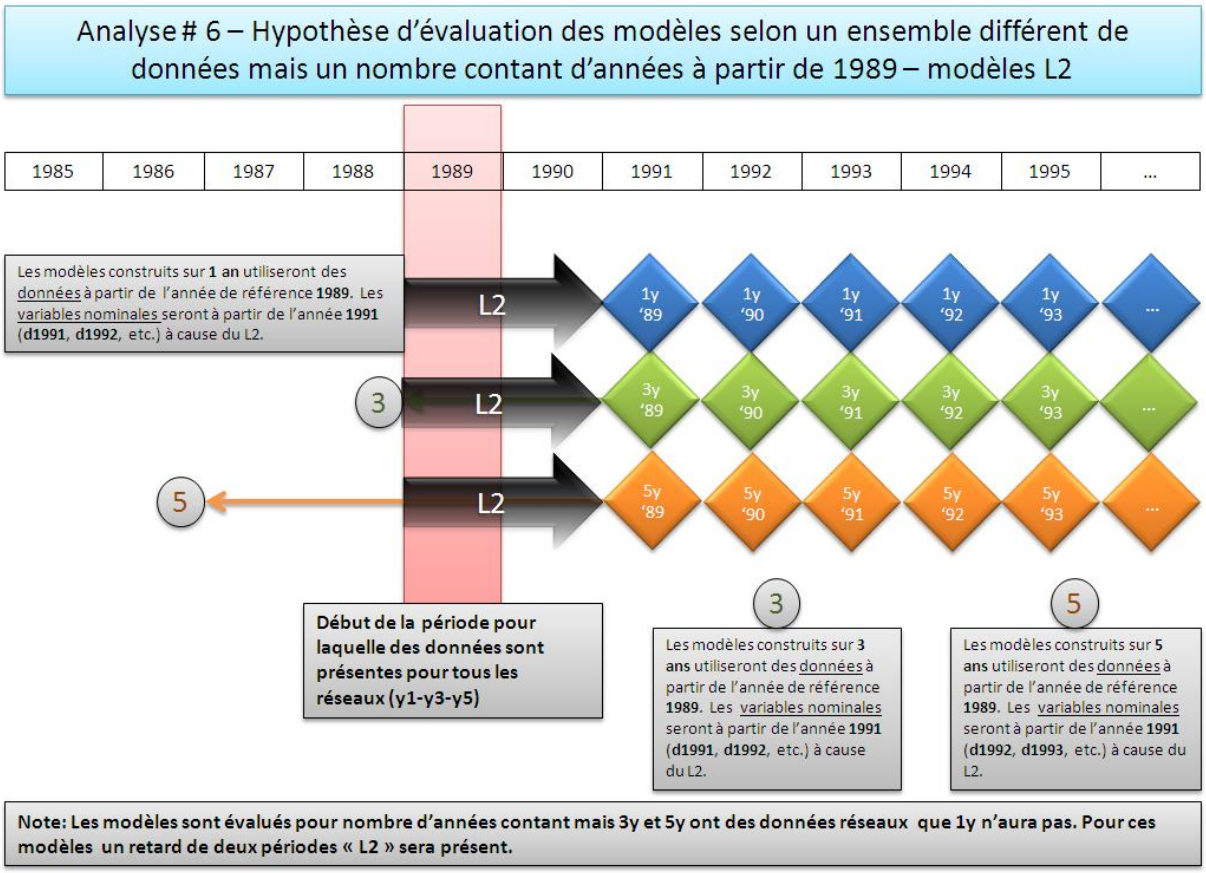

Analyse # 2 – Hypothèse de compensation temporelle pour l'évaluation des modèles selon le même ensemble de données mais différentes années – modèles L1
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
...
Les modèles construits sur 1 an utiliseront des données à partir de l'année de référence 1989. Les variables nominales seront à partir de l'année 1990 (d1990, d1991, etc.) à cause du L1.
L1
1y '89
1y '90
1y '91
1y '92
1y '93
1y '94
...
3
L1
3y '91
3y '92
3y '93
3y '94
...
5
L1
5y '93
5y '94
...
Début de la période pour laquelle des données sont présentes pour tous les réseaux (y1-y3-y5)
3
Les modèles construits sur 3 ans utiliseront des données à partir de l'année de référence 1991. Les variables nominales seront à partir de l'année 1992 (d1992, d1993, etc.) à cause du L1.
5
Les modèles construits sur 5 ans utiliseront des données à partir de l'année de référence 1993. Les variables nominales seront à partir de l'année 1994 (d1994, d1995, etc.) à cause du L1.
Note: La compensation temporelle permet de comparer les modèles selon le même semble de données. Pour ces modèles un retard d'une période « L1 » sera présent.

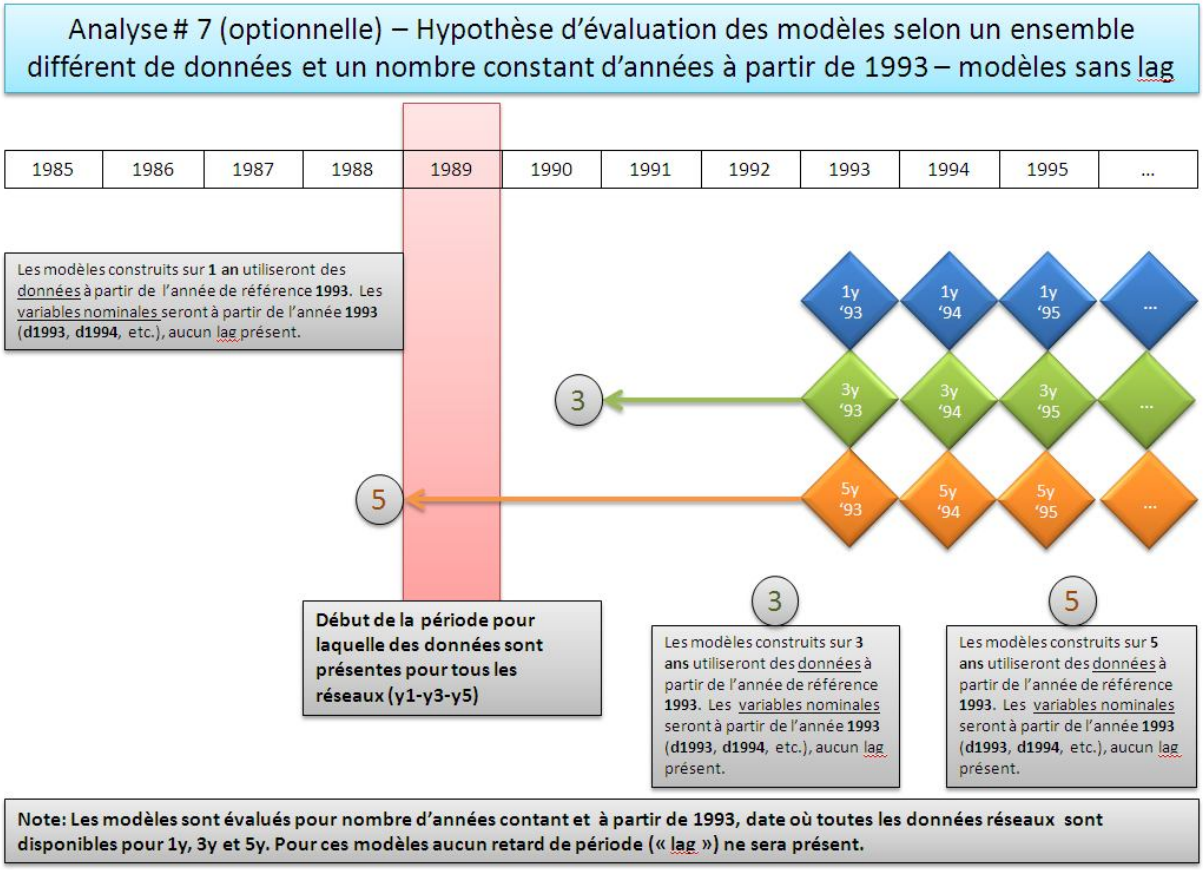

Analyse # 3 – Hypothèse de compensation temporelle pour l'évaluation des modèles selon le même ensemble de données mais différentes années – modèles L2
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
...
Les modèles construits sur 1 an utiliseront des données à partir de l'année de référence 1989. Les variables nominales seront à partir de l'année 1991 (d1991, d1992, etc.) à cause du L2.
L2
1y '89
1y '90
1y '91
1y '92
1y '93
...
3
L2
3y '91
3y '92
3y '93
...
5
L2
5y '93
...
Début de la période pour laquelle des données sont présentes pour tous les réseaux (y1-y3-y5)
3
Les modèles construits sur 3 ans utiliseront des données à partir de l'année de référence 1991. Les variables nominales seront à partir de l'année 1993 (d1993, d1994, etc.) à cause du L2.
5
Les modèles construits sur 5 ans utiliseront des données à partir de l'année de référence 1993. Les variables nominales seront à partir de l'année 1995 (d1995, d1996, etc.) à cause du L2.
Note: La compensation temporelle permet de comparer les modèles selon le même semble de données. Pour ces modèles un retard de deux périodes « L2 » sera présent.

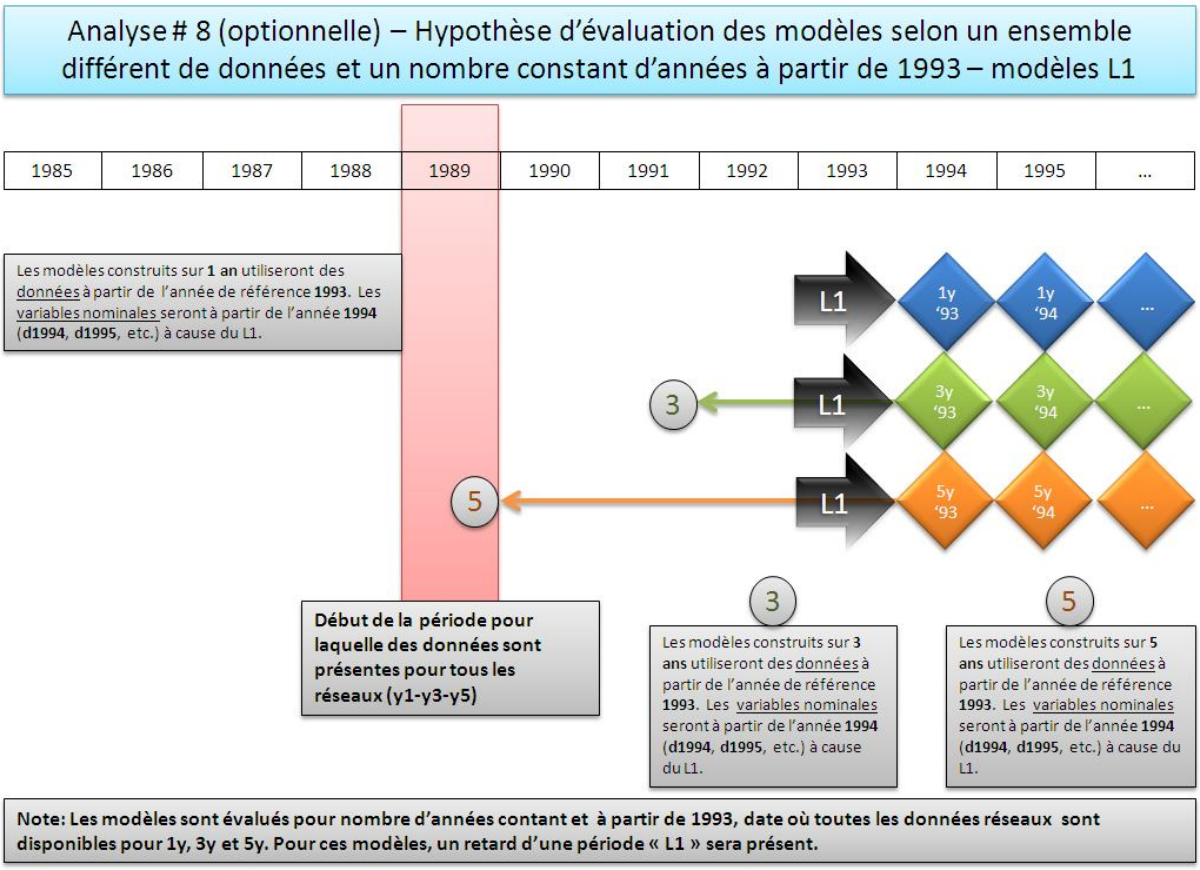

Analyse # 4 – Hypothèse d'évaluation des modèles selon un ensemble différent de données mais un nombre constant d'années à partir de 1989 – modèles sans lag
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
...
Les modèles construits sur 1 an utiliseront des données à partir de l'année de référence 1989. Les variables nominales seront à partir de l'année 1989 (d1989, d1990, etc.), aucun lag présent.
1y '89
1y '90
1y '91
1y '92
1y '93
1y '94
1y '95
...
3
3y '89
3y '90
3y '91
3y '92
3y '93
3y '94
3y '95
...
5
5y '89
5y '90
5y '91
5y '92
5y '93
5y '94
5y '95
...
Début de la période pour laquelle des données sont présentes pour tous les réseaux (y1-y3-y5)
3
Les modèles construits sur 3 ans utiliseront des données à partir de l'année de référence 1989. Les variables nominales seront à partir de l'année 1989 (d1989, d1990, etc.), aucun lag présent.
5
Les modèles construits sur 5 ans utiliseront des données à partir de l'année de référence 1989. Les variables nominales seront à partir de l'année 1989 (d1989, d1989, etc.), aucun lag présent.
Note: Les modèles sont évalués pour nombre d'années contant mais 3y et 5y ont des données réseaux que 1y n'aura pas. Pour ces modèles aucun retard de période (« lag ») ne sera présent.

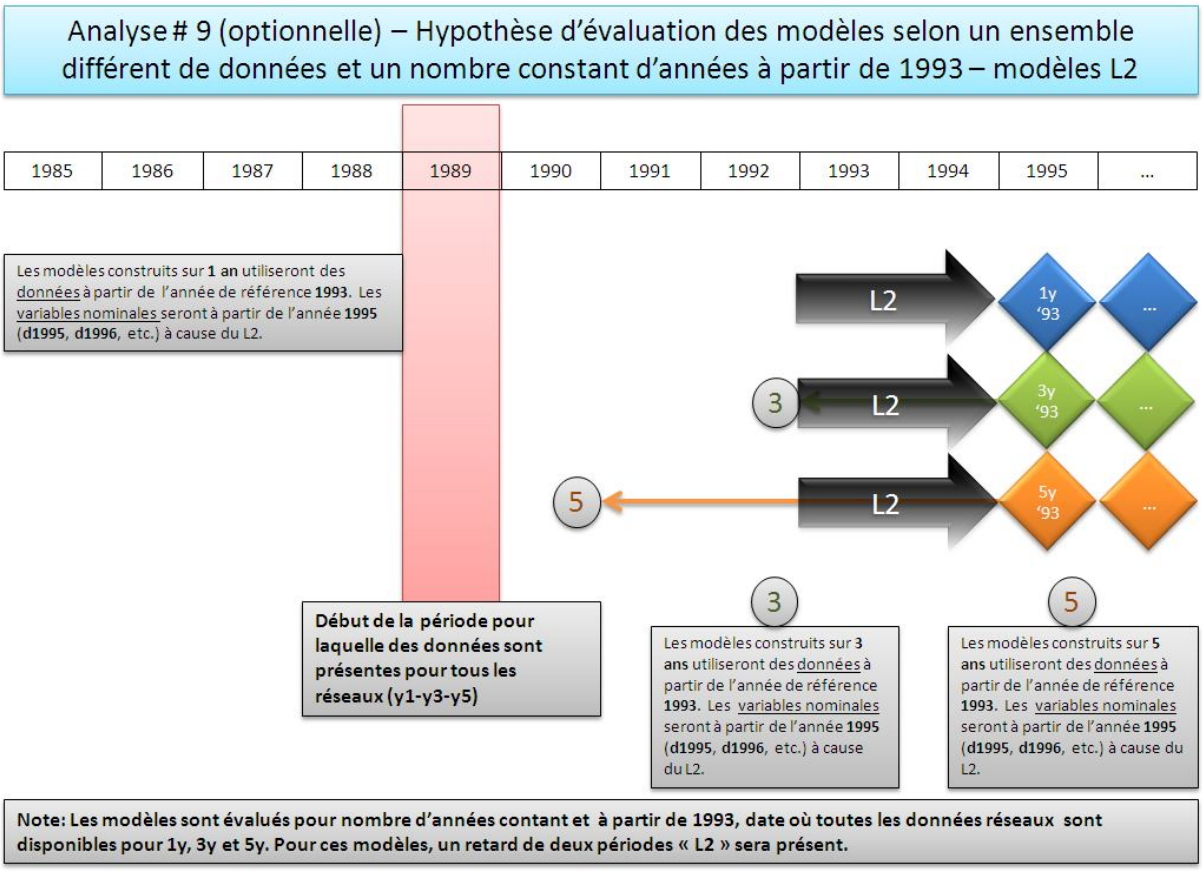


Note: Les modèles sont évalués pour nombre d'années contant mais 3y et 5y ont des données réseaux que 1y n'aura pas. Pour ces modèles un retard d'une période « L1 » sera présent.


Analyse # 6 – Hypothèse d'évaluation des modèles selon un ensemble différent de données mais un nombre contant d'années à partir de 1989 – modèles L2
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
...
Les modèles construits sur 1 an utiliseront des données à partir de l'année de référence 1989. Les variables nominales seront à partir de l'année 1991 (d1991, d1992, etc.) à cause du L2.
L2
1y '89
1y '90
1y '91
1y '92
1y '93
...
3
L2
3y '89
3y '90
3y '91
3y '92
3y '93
...
5
L2
5y '89
5y '90
5y '91
5y '92
5y '93
...
Début de la période pour laquelle des données sont présentes pour tous les réseaux (y1-y3-y5)
3
Les modèles construits sur 3 ans utiliseront des données à partir de l'année de référence 1989. Les variables nominales seront à partir de l'année 1991 (d1991, d1992, etc.) à cause du L2.
5
Les modèles construits sur 5 ans utiliseront des données à partir de l'année de référence 1989. Les variables nominales seront à partir de l'année 1991 (d1992, d1993, etc.) à cause du L2.
Note: Les modèles sont évalués pour nombre d'années contant mais 3y et 5y ont des données réseaux que 1y n'aura pas. Pour ces modèles un retard de deux périodes « L2 » sera présent.


Analyse # 7 (optionnelle) – Hypothèse d'évaluation des modèles selon un ensemble différent de données et un nombre constant d'années à partir de 1993 – modèles sans lag
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
...
Les modèles construits sur 1 an utiliseront des données à partir de l'année de référence 1993. Les variables nominales seront à partir de l'année 1993 (d1993, d1994, etc.), aucun lag présent.
1y '93
1y '94
1y '95
...
3
3y '93
3y '94
3y '95
...
5
5y '93
5y '94
5y '95
...
Début de la période pour laquelle des données sont présentes pour tous les réseaux (y1-y3-y5)
3
Les modèles construits sur 3 ans utiliseront des données à partir de l'année de référence 1993. Les variables nominales seront à partir de l'année 1993 (d1993, d1994, etc.), aucun lag présent.
5
Les modèles construits sur 5 ans utiliseront des données à partir de l'année de référence 1993. Les variables nominales seront à partir de l'année 1993 (d1993, d1994, etc.), aucun lag présent.
Note: Les modèles sont évalués pour nombre d'années contant et à partir de 1993, date où toutes les données réseaux sont disponibles pour 1y, 3y et 5y. Pour ces modèles aucun retard de période (« lag ») ne sera présent.


Analyse # 8 (optionnelle) – Hypothèse d'évaluation des modèles selon un ensemble différent de données et un nombre constant d'années à partir de 1993 – modèles L1
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
...
Les modèles construits sur 1 an utiliseront des données à partir de l'année de référence 1993. Les variables nominales seront à partir de l'année 1994 (d1994, d1995, etc.) à cause du L1.
L1
1y '93
1y '94
...
3
L1
3y '93
3y '94
...
5
L1
5y '93
5y '94
...
Début de la période pour laquelle des données sont présentes pour tous les réseaux (y1-y3-y5)
3
Les modèles construits sur 3 ans utiliseront des données à partir de l'année de référence 1993. Les variables nominales seront à partir de l'année 1994 (d1994, d1995, etc.) à cause du L1.
5
Les modèles construits sur 5 ans utiliseront des données à partir de l'année de référence 1993. Les variables nominales seront à partir de l'année 1994 (d1994, d1995, etc.) à cause du L1.
Note: Les modèles sont évalués pour nombre d'années contant et à partir de 1993, date où toutes les données réseaux sont disponibles pour 1y, 3y et 5y. Pour ces modèles, un retard d'une période « L1 » sera présent.


Analyse # 9 (optionnelle) – Hypothèse d'évaluation des modèles selon un ensemble différent de données et un nombre constant d'années à partir de 1993 – modèles L2
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
...
Les modèles construits sur 1 an utiliseront des données à partir de l'année de référence 1993. Les variables nominales seront à partir de l'année 1995 (d1995, d1996, etc.) à cause du L2.
L2
1y '93
...
3
L2
3y '93
...
5
L2
5y '93
...
Début de la période pour laquelle des données sont présentes pour tous les réseaux (y1-y3-y5)
3
Les modèles construits sur 3 ans utiliseront des données à partir de l'année de référence 1993. Les variables nominales seront à partir de l'année 1995 (d1995, d1996, etc.) à cause du L2.
5
Les modèles construits sur 5 ans utiliseront des données à partir de l'année de référence 1993. Les variables nominales seront à partir de l'année 1995 (d1995, d1996, etc.) à cause du L2.
Note: Les modèles sont évalués pour nombre d'années contant et à partir de 1993, date où toutes les données réseaux sont disponibles pour 1y, 3y et 5y. Pour ces modèles, un retard de deux périodes « L2 » sera présent.

ANNEXE D - RÉSULTATS DES MESURES AIC/BIC

Analyse # 1 – sans retard de phase – 1 an

(not concave)

Analyse # 1 – sans retard de phase – 3 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	16556	.	-22549,78	29	45157,56	45381,28
B2	16556	.	-22550,18	29	45158,35	45382,08
B3	16556	.	-22549,64	29	45157,29	45381,01
B4	16562	.	-22556,38	29	45170,76	45394,49
B5	16556	.	-22555,89	29	45169,79	45393,51
B6	16556	.	-22556,37	29	45170,75	45394,47
B7	16556	.	-22555,74	29	45169,48	45393,2
B8	16562	.	-22563,14	29	45184,29	45408,02
B9	16556	.	-22549,78	29	45157,55	45381,28
B10	16556	.	-22550,17	29	45158,35	45382,07
B11	16556	.	-22549,64	29	45157,28	45381
B12	16562	.	-22556,37	29	45170,75	45394,48

Analyse # 1 – sans retard de phase – 5 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	15192	.	-21252,45	27	42558,89	42764,86
B2	15192	.	-21252,57	27	42559,15	42765,12
B3	15192	.	-21252,35	27	42558,7	42764,67
B4	15194	.	-21254,26	27	42562,52	42768,49
B5	15192	.	-21257,79	27	42569,58	42775,55
B6	15192	.	-21257,99	27	42569,98	42775,95
B7	15192	.	-21257,66	27	42569,32	42775,29
B8	15194	.	-21260,11	27	42574,22	42780,2
B9	15192	.	-21252,44	27	42558,87	42764,84
B10	15192	.	-21252,56	27	42559,13	42765,1
B11	15192	.	-21252,34	27	42558,68	42764,65
B12	15194	.	-21254,25	27	42562,5	42768,47

Analyse # 2 – L1 – 1 an

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	17100	.	-24708,15	31	49478,3	49718,45
B2	17100	.	-24708,37	31	49478,74	49718,89
B3	17104	.	-24717,44	31	49496,89	49737,04
B4	17100	.	-24710,77	31	49483,55	49723,7
B5	17100	.	-24711	31	49484	49724,15
B6	17104	.	-24720,26	31	49502,52	49742,68

Analyse # 2 – L1 – 3 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	16270	.	-23494,45	29	47046,9	47270,12
B2	16270	.	-23495,1	29	47048,19	47271,41
B3	16270	.	-23494,25	29	47046,5	47269,72
B4	16274	.	-23503,71	29	47065,43	47288,65
B5	16270	.	-23495,9	29	47049,81	47273,02
B6	16270	.	-23496,57	29	47051,14	47274,35
B7	16270	.	-23495,7	29	47049,4	47272,61
B8	16274	.	-23505,44	29	47068,87	47292,09
B9	16270	.	-23493,91	29	47045,81	47269,03
B10	16270	.	-23494,55	29	47047,09	47270,31
B11	16270	.	-23493,71	29	47045,41	47268,63
B12	16274	.	-23503,08	29	47064,17	47287,39

Analyse # 2 – L1 – 5 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	14932	.	-21731,58	27	43517,17	43722,67
B2	14932	.	-21731,98	27	43517,95	43723,46
B3	14932	.	-21731,33	27	43516,66	43722,16
B4	14933	.	-21734,23	27	43522,46	43727,96
B5	14932	.	-21732,94	27	43519,88	43725,38
B6	14932	.	-21733,35	27	43520,71	43726,21
B7	14932	.	-21732,67	27	43519,34	43724,85
B8	14933	.	-21735,84	27	43525,68	43731,19
B9	14932	.	-21730,71	27	43515,42	43720,93
B10	14932	.	-21731,1	27	43516,19	43721,7
B11	14932	.	-21730,46	27	43514,93	43720,43
B12	14933	.	-21733,25	27	43520,5	43726,01

Analyse # 3 – L2 – 1 an

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	16696	.	-24190,84	30	48441,68	48673,37
B2	16696	.	-24191,11	30	48442,21	48673,9
B3	16700	.	-24198,26	30	48456,52	48688,21
B4	16696	.	-24193,23	30	48446,46	48678,14
B5	16696	.	-24193,5	30	48446,99	48678,68
B6	16700	.	-24200,93	30	48461,86	48693,55

Analyse # 3 – L2 – 3 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	15853	.	-22986,31	28	46028,62	46243,41
B2	15853	.	-22986,62	28	46029,24	46244,03
B3	15853	.	-22986,14	28	46028,28	46243,07
B4	15856	.	-22991,04	28	46038,08	46252,88
B5	15853	.	-22988,13	28	46032,26	46247,05
B6	15853	.	-22988,45	28	46032,9	46247,7
B7	15853	.	-22987,96	28	46031,91	46246,7
B8	15856	.	-22993,07	28	46042,14	46256,94
B9	15853	.	-22985,73	28	46027,46	46242,25
B10	15853	.	-22986,04	28	46028,08	46242,87
B11	15853	.	-22985,57	28	46027,13	46241,92
B12	15856	.	-22990,41	28	46036,82	46251,61

Analyse # 3 – L2 – 5 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	14580	.	-21304,28	26	42660,56	42857,83
B2	14580	.	-21304,62	26	42661,25	42858,52
B3	14580	.	-21304,05	26	42660,09	42857,36
B4	14581	.	-21306,2	26	42664,39	42861,67
B5	14580	.	-21305,45	26	42662,91	42860,18
B6	14580	.	-21305,82	26	42663,64	42860,91
B7	14580	.	-21305,21	26	42662,42	42859,69
B8	14581	.	-21307,55	26	42667,11	42864,38
B9	14580	.	-21303,64	26	42659,29	42856,56
B10	14580	.	-21303,98	26	42659,97	42857,24
B11	14580	.	-21303,42	26	42658,83	42856,11
B12	14581	.	-21305,49	26	42662,98	42860,25

Analyse # 4 – sans retard de phase – 1 an

(not concave)

Analyse # 4 – sans retard de phase – 3 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	17803	.	-23910,97	31	47883,94	48125,34
B2	17803	.	-23911,33	31	47884,66	48126,06
B3	17803	.	-23910,84	31	47883,69	48125,09
B4	17809	.	-23917,99	31	47897,98	48139,39
B5	17803	.	-23917,56	31	47897,12	48138,52
B6	17803	.	-23917,99	31	47897,99	48139,39
B7	17803	.	-23917,41	31	47896,83	48138,23
B8	17809	.	-23925,19	31	47912,38	48153,79
B9	17803	.	-23910,97	31	47883,93	48125,33
B10	17803	.	-23911,33	31	47884,65	48126,05
B11	17803	.	-23910,84	31	47883,68	48125,08
B12	17809	.	-23917,99	31	47897,97	48139,38

Analyse # 4 – sans retard de phase – 5 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	18019	.	-24533,05	31	49128,11	49369,88
B2	18019	.	-24533,12	31	49128,23	49370,01
B3	18019	.	-24533,01	31	49128,02	49369,8
B4	18021	.	-24534,79	31	49131,58	49373,36
B5	18019	.	-24537,91	31	49137,82	49379,6
B6	18019	.	-24538,05	31	49138,1	49379,88
B7	18019	.	-24537,84	31	49137,67	49379,45
B8	18021	.	-24539,9	31	49141,8	49383,58
B9	18019	.	-24533,04	31	49128,09	49369,86
B10	18019	.	-24533,11	31	49128,21	49369,99
B11	18019	.	-24533	31	49128	49369,78
B12	18021	.	-24534,78	31	49131,56	49373,34

Analyse # 5 – L1 – 1 an

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	17100	.	-24708,15	31	49478,3	49718,45
B2	17100	.	-24708,37	31	49478,74	49718,89
B3	17104	.	-24717,44	31	49496,89	49737,04
B4	17100	.	-24710,77	31	49483,55	49723,7
B5	17100	.	-24711	31	49484	49724,15
B6	17104	.	-24720,26	31	49502,52	49742,68

Analyse # 5 – L1 – 3 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	17426	.	-24890,78	31	49843,56	50084,3
B2	17426	.	-24891,33	31	49844,67	50085,41
B3	17426	.	-24890,65	31	49843,29	50084,03
B4	17430	.	-24899,21	31	49860,41	50101,16
B5	17426	.	-24892,63	31	49847,26	50088
B6	17426	.	-24893,21	31	49848,42	50089,15
B7	17426	.	-24892,5	31	49846,99	50087,73
B8	17430	.	-24901,29	31	49864,58	50105,33
B9	17426	.	-24890,17	31	49842,33	50083,07
B10	17426	.	-24890,72	31	49843,43	50084,17
B11	17426	.	-24890,04	31	49842,07	50082,81
B12	17430	.	-24898,53	31	49859,06	50099,8

Analyse # 5 – L1 – 5 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	17041	.	-24339,13	30	48738,25	48970,55
B2	17041	.	-24339,43	30	48738,86	48971,16
B3	17041	.	-24338,97	30	48737,93	48970,24
B4	17042	.	-24340,82	30	48741,64	48973,94
B5	17041	.	-24341,14	30	48742,28	48974,58
B6	17041	.	-24341,47	30	48742,93	48975,23
B7	17041	.	-24340,97	30	48741,94	48974,24
B8	17042	.	-24343,04	30	48746,08	48978,38
B9	17041	.	-24337,99	30	48735,99	48968,29
B10	17041	.	-24338,29	30	48736,57	48968,88
B11	17041	.	-24337,84	30	48735,68	48967,98
B12	17042	.	-24339,62	30	48739,24	48971,54

Analyse # 6 – L2 – 1 an

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	16696	.	-24190,84	30	48441,68	48673,37
B2	16696	.	-24191,11	30	48442,21	48673,9
B3	16700	.	-24198,26	30	48456,52	48688,21
B4	16696	.	-24193,23	30	48446,46	48678,14
B5	16696	.	-24193,5	30	48446,99	48678,68
B6	16700	.	-24200,93	30	48461,86	48693,55

Analyse # 6 – L2 – 3 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	16909	.	-24287,99	30	48635,98	48868,05
B2	16909	.	-24288,29	30	48636,59	48868,66
B3	16909	.	-24287,88	30	48635,77	48867,84
B4	16912	.	-24292,09	30	48644,18	48876,25
B5	16909	.	-24290,51	30	48641,02	48873,09
B6	16909	.	-24290,83	30	48641,65	48873,72
B7	16909	.	-24290,4	30	48640,79	48872,86
B8	16912	.	-24294,81	30	48649,62	48881,69
B9	16909	.	-24287,31	30	48634,63	48866,7
B10	16909	.	-24287,61	30	48635,23	48867,3
B11	16909	.	-24287,21	30	48634,41	48866,48
B12	16912	.	-24291,37	30	48642,74	48874,81

Analyse # 6 – L2 – 5 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	15958	.	-23042,17	29	46142,33	46364,98
B2	15958	.	-23042,46	29	46142,92	46365,57
B3	15958	.	-23041,97	29	46141,93	46364,59
B4	15959	.	-23043,97	29	46145,94	46368,6
B5	15958	.	-23043,76	29	46145,51	46368,17
B6	15958	.	-23044,07	29	46146,14	46368,79
B7	15958	.	-23043,55	29	46145,09	46367,75
B8	15959	.	-23045,75	29	46149,49	46372,15
B9	15958	.	-23041,39	29	46140,78	46363,43
B10	15958	.	-23041,68	29	46141,36	46364,01
B11	15958	.	-23041,19	29	46140,39	46363,04
B12	15959	.	-23043,13	29	46144,26	46366,91

Analyse # 7 – sans retard de phase – 1 an

(not concave)

Analyse # 7 – sans retard de phase – 3 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	14978	.	-20758,23	27	41570,47	41776,05
B2	14978	.	-20758,58	27	41571,16	41776,74
B3	14978	.	-20758,09	27	41570,18	41775,77
B4	14984	.	-20764,6	27	41583,21	41788,8
B5	14978	.	-20764,58	27	41583,17	41788,76
B6	14978	.	-20765	27	41584	41789,59
B7	14978	.	-20764,42	27	41582,84	41788,43
B8	14984	.	-20771,7	27	41597,4	41802,99
B9	14978	.	-20758,23	27	41570,46	41776,05
B10	14978	.	-20758,57	27	41571,15	41776,73
B11	14978	.	-20758,09	27	41570,17	41775,76
B12	14984	.	-20764,6	27	41583,2	41788,79

Analyse # 7 – sans retard de phase – 5 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	15192	.	-21252,45	27	42558,89	42764,86
B2	15192	.	-21252,57	27	42559,15	42765,12
B3	15192	.	-21252,35	27	42558,7	42764,67
B4	15194	.	-21254,26	27	42562,52	42768,49
B5	15192	.	-21257,79	27	42569,58	42775,55
B6	15192	.	-21257,99	27	42569,98	42775,95
B7	15192	.	-21257,66	27	42569,32	42775,29
B8	15194	.	-21260,11	27	42574,22	42780,2
B9	15192	.	-21252,44	27	42558,87	42764,84
B10	15192	.	-21252,56	27	42559,13	42765,1
B11	15192	.	-21252,34	27	42558,68	42764,65
B12	15194	.	-21254,25	27	42562,5	42768,47

Analyse # 8 – L1 – 1 an

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	14465	.	-21489,08	27	43032,17	43236,81
B2	14465	.	-21489,25	27	43032,5	43237,14
B3	14469	.	-21498,03	27	43050,06	43254,72
B4	14465	.	-21490,5	27	43035	43239,65
B5	14465	.	-21490,67	27	43035,33	43239,98
B6	14469	.	-21499,57	27	43053,15	43257,8

Analyse # 8 – L1 – 3 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	14788	.	-21672,12	27	43398,23	43603,48
B2	14788	.	-21672,69	27	43399,37	43604,62
B3	14788	.	-21671,92	27	43397,83	43603,07
B4	14792	.	-21681,77	27	43417,54	43622,79
B5	14788	.	-21673,31	27	43400,62	43605,86
B6	14788	.	-21673,9	27	43401,79	43607,03
B7	14788	.	-21673,1	27	43400,21	43605,45
B8	14792	.	-21683,22	27	43420,44	43625,69
B9	14788	.	-21671,67	27	43397,34	43602,58
B10	14788	.	-21672,24	27	43398,47	43603,71
B11	14788	.	-21671,47	27	43396,94	43602,18
B12	14792	.	-21681,25	27	43416,49	43621,74

Analyse # 8 – L1 – 5 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	14932	.	-21731,58	27	43517,17	43722,67
B2	14932	.	-21731,98	27	43517,95	43723,46
B3	14932	.	-21731,33	27	43516,66	43722,16
B4	14933	.	-21734,23	27	43522,46	43727,96
B5	14932	.	-21732,94	27	43519,88	43725,38
B6	14932	.	-21733,35	27	43520,71	43726,21
B7	14932	.	-21732,67	27	43519,34	43724,85
B8	14933	.	-21735,84	27	43525,68	43731,19
B9	14932	.	-21730,71	27	43515,42	43720,93
B10	14932	.	-21731,1	27	43516,19	43721,7
B11	14932	.	-21730,46	27	43514,93	43720,43
B12	14933	.	-21733,25	27	43520,5	43726,01

Analyse # 9 – L2 – 1 an

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	14266	.	-21167,48	26	42386,96	42583,67
B2	14266	.	-21167,71	26	42387,43	42584,13
B3	14270	.	-21174,25	26	42400,5	42597,21
B4	14266	.	-21168,8	26	42389,6	42586,3
B5	14266	.	-21169,03	26	42390,07	42586,77
B6	14270	.	-21175,74	26	42403,48	42600,2

Analyse # 9 – L2 – 3 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	14477	.	-21258,19	26	42568,38	42765,47
B2	14477	.	-21258,52	26	42569,03	42766,12
B3	14477	.	-21258,02	26	42568,04	42765,12
B4	14480	.	-21263,06	26	42578,12	42775,21
B5	14477	.	-21259,71	26	42571,41	42768,5
B6	14477	.	-21260,05	26	42572,09	42769,18
B7	14477	.	-21259,53	26	42571,06	42768,15
B8	14480	.	-21264,79	26	42581,58	42778,67
B9	14477	.	-21257,69	26	42567,38	42764,47
B10	14477	.	-21258,01	26	42568,02	42765,11
B11	14477	.	-21257,52	26	42567,04	42764,13
B12	14480	.	-21262,5	26	42577	42774,1

Analyse # 9 – L2 – 5 ans

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
B1	14580	.	-21304,28	26	42660,56	42857,83
B2	14580	.	-21304,62	26	42661,25	42858,52
B3	14580	.	-21304,05	26	42660,09	42857,36
B4	14581	.	-21306,2	26	42664,39	42861,67
B5	14580	.	-21305,45	26	42662,91	42860,18
B6	14580	.	-21305,82	26	42663,64	42860,91
B7	14580	.	-21305,21	26	42662,42	42859,69
B8	14581	.	-21307,55	26	42667,11	42864,38
B9	14580	.	-21303,64	26	42659,29	42856,56
B10	14580	.	-21303,98	26	42659,97	42857,24
B11	14580	.	-21303,42	26	42658,83	42856,11
B12	14581	.	-21305,49	26	42662,98	42860,25