



**HAL**  
open science

## Knowledge flows through patent citation data

Hugo Pillu

► **To cite this version:**

Hugo Pillu. Knowledge flows through patent citation data. Economics and Finance. Ecole Centrale Paris, 2009. English. NNT : 2009ECAP0048 . tel-00458678

**HAL Id: tel-00458678**

**<https://theses.hal.science/tel-00458678>**

Submitted on 22 Feb 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**ÉCOLE CENTRALE DES ARTS  
ET MANUFACTURES  
« ÉCOLE CENTRALE PARIS »**

## **THÈSE**

présentée et soutenue par :

**Hugo PILLU**

**Le 18/12/2009**

pour l'obtention du grade de

## **DOCTEUR**

Spécialité : **Sciences Economiques**

Laboratoire d'accueil : **ERASME**

TITRE :

### **Knowledge Flows through Patent Citation Data**

---

Directeur de Thèse :

**Paul ZAGAME**, Professeur à l'Université Paris 1 – Panthéon Sorbonne

---

Membre du Jury :

**Dominique FORAY**, Professeur à l'Ecole Polytechnique Fédérale de Lausanne, rapporteur

**Jean-Luc GAFFARD**, Professeur à l'Université de Nice - Sophia Antipolis, rapporteur

**Gilles KOLEDA**, Directeur des Etudes à Coe-Rexecode

**Bruno van POTTELSBERGHE**, Professeur à l'Université Libre de Bruxelles, président

**Numéro d'ordre : 2009ECAP0048**



# Knowledge flows through patent citation data

Hugo Pillu



---

ERASME Ecole Centrale Paris, Grande voie des vignes, 92295 Châtenay Malabry cedex, FRANCE

Phone: +0033 (0)1 41 13 11 79 Fax: +0033 (0)1 41 13 16 67 Email: hugo.pillu@ecp.fr

---

## Acknowledgement

First I would like to express my deep and sincere gratitude to my supervisor, Professor Paul Zagame, Head of ERASME Laboratory and professor at the University Paris 1 - Panthéon Sorbonne, for his supervision, advises, and guidance from the early stages of this research. His vast knowledge and his logical way of thinking have been of great value to me. Above all and most needed, he gave me unwavering encouragement and support in various ways. His scientist intuitions made him a constant oasis of ideas and passions in science, which exceptionally inspired me and enriched my improvement as the student, the researcher and the scientist I want to be. I am indebted to him more than he knows.

A special thank goes to Gilles Koleda, who is most responsible for the completion of my first paper and the challenging research that lies behind it. He taught me how to write academic papers, helped me become a researcher, and highlighted my most significant ideas. Without his encouragement and constant guidance in the first period of my Ph.D., I could not have finished this dissertation. His involvement and his originality triggered and nourished my intellectual maturity. A maturity that I will enjoy for a long time to come.

I gratefully thank Professor Dominique Foray and Professor Jean-Luc Gaffard for agreeing to be the official referees of this thesis and for their detailed review. I also wish to thank Professor Bruno van Pottelsberghe, I am honored he has agreed to participate in the jury of this thesis.

During this work in the ERASME Laboratory at Ecole Centrale Paris, I collaborated with many colleagues for whom I have great regard, and I wish to extend my warmest thanks to everyone who helped me in my work. I thank Pascal Da Costa and Florent Pratlong for their intelligent comments on my work, I also thank Arnaud Fougeyrollas and Pierre Le Mouel for their useful comments and the fact they introduced me to one of the most challenging science that is modeling ; and Danielle Schirmann Duclos for her energy and for her help during these years. Many thanks also go to Baptiste Boitier, Gabriel Galland, Oualid Gharbi, and Boris Le Hir for giving me a good time when working with them. A special thanks goes to Boris, who taught me how to fly a helicopter.

The research of this thesis was sponsored by the Ecole Centrale Paris for three years (2006 to 2009). I am also grateful to the Research D.G. of the European Commission for their additional financial support of my thesis and of several conferences.

It is also a pleasure to pay tribute to Marion Senant, Marine Rabreau and Rudy Bezir for reading over my manuscript and for their comments on the latter, it was a great help for me.

---

I thank my family: my mother Nathalie Leconte, my father Pascal Pillu and my stepfather François Coutin, who taught me the joy of intellectual pursuit, provided me moral advises and supported me ever since I was a child.

I have been fortunate to come across many funny and good friends, without whom life would be bleak. It would be difficult to list them all because it would be dangerous to forget some.

Last but not least, I wish to thank Charlotte for accompanying me during these years, for putting up with me and just be present.

# Contents

<b>1</b>	<b>Introduction</b>	<b>17</b>
1.1	Introduction . . . . .	17
1.2	Motivations . . . . .	20
1.3	Overview . . . . .	24
<b>2</b>	<b>Knowledge Spillovers: a literature review</b>	<b>29</b>
2.1	Introduction . . . . .	29
2.2	An overview of knowledge externalities . . . . .	31
2.2.1	Knowledge as an economic good . . . . .	31
2.2.2	Knowledge transfers mechanisms . . . . .	32
2.2.3	The genesis . . . . .	33
2.3	Two kinds of spillovers . . . . .	34
2.3.1	Difference between private and social return created by knowledge spillovers . . .	36
2.3.2	Difference between private and social return created by a rent spillover . . . . .	38
2.4	Different channels used in the measure of knowledge externalities . . . . .	40
2.4.1	Economic transaction flows . . . . .	42
2.4.1.1	I-O matrices . . . . .	42
2.4.1.2	Import flows . . . . .	43
2.4.1.3	Which spillovers aspects are measured through economic good flows? . . . . .	44
2.4.2	Foreign direct investment: FDI . . . . .	44
2.4.3	Technological flow matrices . . . . .	46
2.4.4	Immaterial transmissions channels . . . . .	48
2.4.4.1	Technological proximity matrices . . . . .	48

2.4.4.2	matrices based on technological classification of patents . . . . .	50
2.4.5	Patent citations . . . . .	50
2.4.5.1	The mechanism of patent citations . . . . .	50
2.4.5.2	Are Citations a good knowledge flows measure? . . . . .	51
2.4.5.3	The use of patent citation data to measure knowledge flows . . . . .	53
2.5	Geographical distance . . . . .	55
2.5.1	The different measures used . . . . .	55
2.5.2	What is the geographical component of knowledge spillover? . . . . .	56
2.6	Conclusion . . . . .	58
2.7	Annexes . . . . .	59
2.7.1	Note on the adoption capacity of technology . . . . .	59
2.7.2	Production Function . . . . .	60
<b>3</b>	<b>Knowledge Spillovers: A Meta-Analysis</b>	<b>63</b>
3.1	Introduction . . . . .	63
3.2	On the measurement of knowledge spillovers effects . . . . .	65
3.2.1	Specifications . . . . .	65
3.2.2	Aggregation level . . . . .	69
3.2.3	Spillovers channels . . . . .	69
3.2.4	Others explanatory variables . . . . .	71
3.3	Meta-Analysis . . . . .	71
3.3.1	Sample selection . . . . .	72
3.3.2	Dealing with multiple results per study . . . . .	73
3.3.3	Publication Bias . . . . .	76
3.3.4	Meta Regression Analysis . . . . .	78
3.4	Econometric results . . . . .	79
3.4.1	FTA and PET tests . . . . .	79
3.4.2	MRA econometrics results . . . . .	81
3.4.2.1	Definition of the moderators variables . . . . .	81
3.4.2.2	MRA of Spillovers at a firm level within a same sector . . . . .	82
3.4.2.3	MRA of intersectoral spillovers within a same country . . . . .	82
3.4.2.4	MRA of Spillovers between countries . . . . .	82

---

## CONTENTS

---

3.4.2.5	MRA of Spillovers within sectors and between countries . . . . .	83
3.4.2.6	MRA of Spillovers between sectors and countries . . . . .	84
3.4.2.7	General remarks . . . . .	85
3.5	Conclusion . . . . .	89
3.6	Appendix . . . . .	89
3.6.1	Methodological issues . . . . .	89
3.6.2	Tables . . . . .	92
<b>4</b>	<b>Patent Citation Data and Knowledge Flows</b>	<b>101</b>
4.1	Introduction . . . . .	101
4.2	Patent and Patent Citation as indicator of knowledge and knowledge flows . . . . .	103
4.2.1	What is a patent ? . . . . .	104
4.2.2	Patent Procedures . . . . .	107
4.2.2.1	EPO. . . . .	109
4.2.2.2	JPO . . . . .	111
4.2.2.3	USPTO . . . . .	112
4.2.2.4	National European patent offices . . . . .	114
4.2.2.5	Cost for obtaining a patent . . . . .	114
4.2.3	Patent as an indicator of innovative activities . . . . .	115
4.2.3.1	Advantages . . . . .	116
4.2.3.2	Drawbacks . . . . .	116
4.2.4	Citations and knowledge flows . . . . .	119
4.2.4.1	Concepts . . . . .	119
4.2.4.2	Practical differences between offices . . . . .	120
4.2.4.3	Patent citations and knowledge flows . . . . .	121
4.3	The citation function . . . . .	124
4.3.1	The origin of the patent citation function . . . . .	124
4.3.2	Rationalization of the patent citation function. . . . .	128
4.3.2.1	Jaffe and Trajtenberg . . . . .	128
4.3.2.2	Popp and the return on the original idea . . . . .	131
4.4	Underlying assumptions and properties of the citation function . . . . .	132
4.4.1	Underlying assumptions . . . . .	133

---

4.4.2	Properties of the function . . . . .	133
4.4.2.1	The modal lag . . . . .	133
4.4.2.2	The cumulative citation probability . . . . .	133
4.4.2.3	Median lag . . . . .	134
4.5	The input based patent citation indicator . . . . .	134
4.5.1	Citations and technological opportunities embodied in cited patents at an inter-national level . . . . .	135
4.5.2	Construction of our input-based patent citation indicator . . . . .	136
4.5.3	Dealing with different database . . . . .	138
4.5.3.1	Advantages of such approach are multiple . . . . .	139
4.5.3.2	Some difficulties have to be challenged . . . . .	139
4.6	Conclusion . . . . .	140
4.7	Appendix . . . . .	141
4.7.1	The modal lab calculi . . . . .	141
4.7.2	The cumulative citation probability calculi . . . . .	142
4.7.3	Median lag calculi . . . . .	143
4.7.4	PIM calculi . . . . .	144
4.7.5	Patents figures . . . . .	147
<b>5</b>	<b>Statistical Lessons from the PATSTAT database</b>	<b>153</b>
5.1	Introduction . . . . .	153
5.2	What is PATSTAT ? . . . . .	155
5.3	The sub-database construction . . . . .	156
5.3.1	The Citing Table . . . . .	157
5.3.2	The Citing publication Table . . . . .	160
5.3.3	The Citing citation Table . . . . .	162
5.3.4	The Cited Table . . . . .	165
5.4	Empirical investigations of patent citations among the different offices . . . . .	168
5.4.1	Analysis & statistical overview . . . . .	168
5.4.2	Number of citations per patent . . . . .	168
5.4.3	Origin of references . . . . .	169
5.4.4	Similarity of citation between EPO and European national offices . . . . .	173

## CONTENTS

---

5.4.5	Similarity of citation between EPO and JPO . . . . .	176
5.5	Conclusion . . . . .	176
5.6	Appendix . . . . .	178
5.6.1	Tables and figures . . . . .	178
<b>6</b>	<b>A measure of International Knowledge Flows through our Input-based Patent Ci- tation Indicator</b>	<b>183</b>
6.1	Introduction . . . . .	183
6.2	The input-based patent citation indicator . . . . .	186
6.3	Data . . . . .	188
6.3.1	Technology classification . . . . .	188
6.3.2	Patents . . . . .	190
6.4	Estimation of the patent citation function . . . . .	193
6.4.1	The citation function . . . . .	193
6.4.2	Econometric issues and interpretations . . . . .	195
6.4.3	Results . . . . .	196
6.5	Construction of our knowledge stocks . . . . .	202
6.6	Empirical validations . . . . .	204
6.6.1	Correlation with R&D expenditures data . . . . .	206
6.6.2	Explanatory power on productivity growth . . . . .	207
6.7	Conclusion . . . . .	211
6.8	Appendix . . . . .	212
6.8.1	Multifactor productivity variables . . . . .	212
6.8.2	R&D capital stocks . . . . .	213
6.8.3	Tables and figures . . . . .	215
6.8.4	Productivity estimates . . . . .	219
<b>7</b>	<b>Demand-Pull and Technological-Push mechanisms in Energy-Efficient Technologies</b>	<b>229</b>
7.1	Introduction . . . . .	229
7.2	Literature Review . . . . .	231
7.2.1	Demand pull theory . . . . .	231
7.2.2	Technology push theory . . . . .	233



7.3	Modeling . . . . .	234
7.4	Data . . . . .	236
7.4.1	Energy patent data set . . . . .	236
7.4.2	Energy prices . . . . .	239
7.4.3	Policies . . . . .	241
7.4.4	Knowledge Stocks . . . . .	242
7.5	Econometrical results . . . . .	247
7.6	Conclusion . . . . .	251
7.7	Appendix . . . . .	251
7.7.1	Definitions of energy technologies . . . . .	251
7.7.2	Tables . . . . .	255
<b>8</b>	<b>Conclusions</b>	<b>259</b>
8.1	Main conclusions . . . . .	260
8.2	Limits . . . . .	261
8.3	Future researches . . . . .	262
	<b>Bibliography</b>	<b>264</b>

# List of Figures

2.1	Rent spillovers return . . . . .	38
3.1	Distribution of the knowledge spillovers effects ( $\gamma$ and $\rho$ ) . . . . .	79
3.2	Funnel graph . . . . .	80
4.1	Number of first patent filing at selected patent offices, 1971-2006 . . . . .	108
4.2	Number of first patent filing at selected patent offices, 1971-2006 . . . . .	108
4.3	EPO and JPO patent procedures . . . . .	110
4.4	USPTO patent procedure . . . . .	113
4.5	Patent Citations . . . . .	120
4.6	Patent Citation between Regions . . . . .	136
4.7	P1; A first patent application at the INPI (FR) . . . . .	147
4.8	P1 search report ; Citations made by the patent P1 . . . . .	148
4.9	P12; A second application of P1 at the EPO . . . . .	149
4.10	P12 search report ; Citations made by the patent P12 . . . . .	150
4.11	P2; A first application at the JPO . . . . .	151
4.12	P22; A second application of P2 at the EPO . . . . .	152
5.1	Summary of the conceptual model . . . . .	157
5.2	An example extracted from the PASTATS database . . . . .	158
5.3	The Citing Table . . . . .	160
5.4	Inventors routes . . . . .	163
5.5	The Citing Publication and the Citing Citation Tables . . . . .	164
5.6	Citing cohort statistics summary . . . . .	166

5.7 The Cited Table . . . . .	167
5.8 Number of citations per patent application . . . . .	170
5.9 Direct and Corrected references by prior application . . . . .	171
5.11 The matching citations structure . . . . .	174
5.12 The PATSTAT physical model . . . . .	180
5.13 Inventors first applications according to the selected offices . . . . .	181
5.10 Origin of references . . . . .	182
6.1 Counting citations process . . . . .	192
6.2 Estimated parameters from the citation function . . . . .	195
6.3 Productivity parameters for Chemical technology . . . . .	203
6.4 Detailed knowledge stocks for Chemical technology . . . . .	205
6.5 Germany, United Kingdom and France estimated citation functions . . . . .	217
6.6 Japan and US estimated citation functions . . . . .	218
6.7 German productivity parameters - sector 1-8 . . . . .	219
6.8 German productivity parameters - sector 9-13 . . . . .	220
6.9 French productivity parameters - sector 1-8 . . . . .	221
6.10 French productivity parameters - sector 9-13 . . . . .	222
6.11 UK productivity parameters - sector 1-8 . . . . .	223
6.12 UK productivity parameters - sector 9-13 . . . . .	224
6.13 Japan productivity parameters - sector 1-8 . . . . .	225
6.14 Japan productivity parameters - sector 9-13 . . . . .	226
6.15 US productivity parameters - sector 1-8 . . . . .	227
6.16 US productivity parameters - sector 9-13 . . . . .	228
7.1 First patent application in energy efficiency technologies . . . . .	237
7.2 First patent application in energy efficiency technologies (continued) . . . . .	238
7.3 Distribution of technologies between countries . . . . .	240
7.4 Energy prices . . . . .	241
7.5 Productivity estimates for Solar energy technology . . . . .	245
7.6 Productivity estimates for Heat Exchange technology . . . . .	246
7.7 1982 Solar knowledge flow for German and US inventors . . . . .	247

---

## LIST OF FIGURES

---

7.8	Solar Knowledge stocks . . . . .	248
7.9	Heat Exchange Knowledge Stocks . . . . .	248
7.10	US knowledge stocks in energy efficiency technologies . . . . .	255
7.11	Ahamd's induced innovation model . . . . .	257



# List of Tables

2.1	Channels and geographic dependencies . . . . .	56
3.1	Sample characteristics . . . . .	74
3.2	Sensitivity of meta-estimates to the choice of “preferred” estimates . . . . .	76
3.3	Meta-analysis of spillovers effects on productivity output according to their nature . . . . .	77
3.4	FTA and PET test for the complete sample . . . . .	80
3.5	MRA of between firms spillovers . . . . .	83
3.6	MRA of inter-sectoral spillovers . . . . .	84
3.7	MRA of international spillovers . . . . .	85
3.8	MRA of international - intra-sectoral spillovers . . . . .	86
3.9	MRA of international - inter-sectoral spillovers . . . . .	87
3.10	Authentic spillovers effects . . . . .	88
3.11	Papers included in the database . . . . .	92
3.12	Papers included in the database (continued) . . . . .	93
3.13	Papers included in the database (continued) . . . . .	94
3.14	Papers included in the database (continued) . . . . .	95
3.15	Definition of the moderators used in the MRA . . . . .	96
3.16	Sensitivity of meta-estimates to the choice of “preferred” estimates . . . . .	97
3.17	Sensitivity of meta-estimates to the choice of “preferred” estimates (continued) . . . . .	98
3.18	Sensitivity of meta-estimates to the choice of “preferred” estimates (continued) . . . . .	99
4.1	First filing in major patent offices (1978-2006) . . . . .	107
4.2	Information available in patent . . . . .	117
4.3	Patent office breadth coefficients . . . . .	140

5.1	Citing table statistics - First patent application (1978-2006) . . . . .	159
5.2	Citing publication table statistics (1978-2006) . . . . .	162
5.3	Citing citation table statistics (1978-2006) . . . . .	164
5.4	Cited table statistics (1976-2006) . . . . .	167
5.5	Average number of patent citations per application . . . . .	169
5.6	Matching rate of citations . . . . .	175
5.7	PATSTAT data coverage . . . . .	178
5.8	PATSTAT elements . . . . .	179
6.1	Technological class based on IPC classification . . . . .	189
6.2	Sample characteristics . . . . .	191
6.3	Regression results . . . . .	197
6.4	Regression results (continued) . . . . .	198
6.5	Modal Lag of citations . . . . .	201
6.6	Pattern of domestic localization over time . . . . .	202
6.7	Correlation between R&D stocks and Patent stocks controlled by their productivities . . . . .	207
6.8	Log Level estimation results . . . . .	209
6.9	Error Correction Model estimations results . . . . .	210
6.10	Schmoch concordance table . . . . .	215
6.11	Definition of the 44 sectoral field based on ISIC codes . . . . .	216
7.1	Patents in Energy Efficient Technologies by Innovating country, 1978-2003 . . . . .	239
7.2	Date of introduction of different public policies for renewable technologies . . . . .	242
7.3	All technologies estimation results . . . . .	249
7.4	Product technologies estimation results . . . . .	250
7.5	Process technologies estimation results . . . . .	250
7.6	Patent citation function estimation results . . . . .	256

# Chapter 1

## Introduction

“Give a man a fish, and you feed him for a day. Teach a man how to fish, and you feed him for a lifetime. To which it now must be added, invent a better method of fishing, or farming fish, selling fish, changing fish (through genetic engineering), or preventing overfishing in the sea, and you feed a great many people, because these methods can be copied virtually without cost and spread around the world.” David Warsh (2006) [187].

### 1.1 Introduction

New ideas more than saving or investment are the keys to prosperity and to the wealth of nations. It is knowledge - not labor, machines, land or natural resources - that is the key economic asset that drives long-run economic performance and the growth process. For more than two centuries and specifically for the last sixty years, the western world economies have grown at a pace that greatly exceeds anything previously known in the long sweep of human history.

It is commonplace in modern times to claim that we live in a knowledge economy, an economy where the creation, the distribution and the use of knowledge become decisive factors. The heart of this knowledge-based economy is innovation and technological change process. They are the sources of growth, but their underlying mechanisms are complex and multiple, making these notions elusive, difficult to conceptualize and even harder to measure. Nevertheless, the immense complexity of this subject has not quelled scholarly interest in this topic, but rather has increased it.

Since its early days the economic discipline has been interested in issues concerning economic



growth. Adam Smith (1776) [165] analyzed the theoretical link between innovation and economic growth. Not only did he articulate the productivity gains from specialization through the productivity of labor as well as from technological improvements to capital equipment and processes, Smith even recognized an early version of technology transfer from suppliers to users and the role of a distinct R&D function operating in the economy. Although the relation between innovation and growth had been investigated at an intuitive level afterward, technical change was not introduced into formal economic growth models until Solow in the late 1950s. One year after he had showed that long-run growth depends fundamentally on continued technical change (1956) [166], Solow (1957) [167] gave us a way to measure it. Indeed, the growth accounting he proposes, allows us to decompose the sources of growth into technical change and the growth of physical inputs. While these generation of models still treated technological change as exogenous to the growth process, some early attempts to develop endogenous technical change models took place during the 1960s (see Arrow (1962) [11]). However, the endogenous growth models were put at the core of growth theory during the 1980s by Romer (1986 [149]; 1990 [150]) and Lucas (1988) [112]. In particular, Romer emphasizes the necessary role of knowledge spillovers in escaping the fate of diminishing returns. Thus, without the presence of knowledge externalities, it is unlikely that economic growth can proceed at a constant or undiminished rate into the future. The concept of knowledge externalities was first introduced by Marshall (1920) [120]. Since his seminal work, economics have established the special characteristics of knowledge in generating positive externalities. Knowledge is non-rivalrous (the use by one agent does not preclude the use by another) and non-excludable (it is difficult to prevent the others from using it). Therefore, knowledge spillovers arise when an agent developing a new idea or process, cannot fully appropriate the results of its innovation. For three decades, empirical works have attempted to measure the extent and the effect of such spillovers at different aggregation levels (inter firm, inter sectoral, inter country) but some confusing points remain.

Understanding mechanisms of innovation, technical change and knowledge life is fundamental for policy makers, since these mechanisms are at the source of growth. In March 2000, at the European social and economic summit held in Lisbon, the EU set itself the strategic objective for 2010 to “become the most dynamic and competitive knowledge-based economy, capable of sustainable economic growth, accompanied by quantitative and qualitative improvement in employment and a more cohesive society”.

But some practical difficulties limit our ability to study these phenomena empirically. Since the visionary paper of Kuznet (1956) [102], standing the difficulties of measuring the result of the inventive process, we are aware that one of the major challenges and needs was the development of new

## Chapter 1. Introduction

---

methodologies and indicators to grasp the different facets of the knowledge-based economy. According to that, numbers of efforts were carried out to overcome measurement problems and gather data that can be used for empirical analysis of technical change.

One solution to the lack of such data was brought by Jaffe and Trajtenberg during the 1990s in using patent as indicator of newly created knowledge and patent citation as an indicator of knowledge diffusion. Patents are one of the richest sources of data on inventions and certainly the one with the widest coverage providing an unique insight into the processes, the diffusion and the outcomes of inventive activities. Their collaborative works open a “window on the knowledge economy” through the massive use of detailed USPTO patent data<sup>1</sup>, this set of researchs is grouped into Jaffe and Trajtenberg (2002) [87]. Their work does not only open a window on the knowledge economy, it also lays out new conceptual and methodological foundations opening ways for the empirical analysis of the importance of innovations and to the measure of knowledge flows.

Nonetheless, their studies use only US data that are strongly biased toward US inventors, limiting the analysis of global patterns of knowledge diffusion. The central goal of this thesis is to evaluate how knowledge flows across agents thanks to international patent citation data. Data used in this thesis are taken from a new global patent database, PATSTAT<sup>2</sup>, that includes all patent citations from most of patent offices in the world. For this purpose, we will attempt to shed some lights about the literature on knowledge spillovers and patent citations. These reviews are followed by the empirical investigation of patent and patent citation data that were never studied before (taken from the Patstat database for the G5 patent offices plus the WIPO and the EPO<sup>3</sup>) and by the investigation of knowledge diffusion for the manufacturing industries of the five largest industrial countries (G5). These explorations lay the groundwork for the creation of a new input-based indicator of knowledge that takes into account international knowledge spillovers and that can be very helpful because traditional indicators of technical change are not always available. And finally, this indicator will be applied in a case study related to the analysis of innovation determinants for efficient energy technologies.

---

<sup>1</sup>USPTO is the United States Patent and Trademark Office

<sup>2</sup>PATSTAT is the EPO Worldwide Patent Statistical Database and was developed by patent information experts at the EPO's Vienna sub-office. It includes patent data from 73 offices world-wide and post-grant data from about 40 offices. It was developed specifically with the needs of policymakers, academics, analysts and IP institutions in mind. Researchers working in this field have previously had to assemble data sets from various and disparate sources and were obliged to perform extensive “cleaning” of the data at considerable cost and time. The PATSTAT dataset addresses these issues, efficiently harmonizing data, resolving issues over family members and addressing such problems as applications from one applicant appearing under several different names. The database also contains related information on citations, procedural information and legal status, which are all of interest to statisticians. We actually use the September 2008 version that included information on about 75 millions of entry.

<sup>3</sup>EPO is the European Patent Office and the WIPO is the World Intellectual Property Organization

## 1.2 Motivations

The aims of this dissertation are multiple, but have the common point to be concerned with the understanding of technological change. A major limitation to grasp these economic phenomena results from the lack of empirical data available. We will attempt to contribute to the development of alternative methods for the measurement of innovative activity, specifically about the most intriguing aspect of technical innovation that is knowledge spillovers. The development of new methods and indicators takes part in the wide-ranging debate on the use of patent as an economic indicator.

Knowledge spillovers among different economic units are of great importance for public policy making. They are the major sources of economic growth and on their magnitudes depend on the escape of the fate of diminishing return and future growth rates. The presence of knowledge externalities induces a sub-investment in private R&D expenditures, because agents do not take into account the positive effect of their own R&D expenditures on the productivity of other agents. The gap between social and private return (spillovers) can be resolved by public policy interventions. Thereby, the appropriate management of R&D incentives and public funding researches require a clear comprehension of how spillovers flow and their underlying mechanism.

Numerous studies have analyzed the patterns and the effects of knowledge spillovers, at both microeconomic and macroeconomic levels. These studies report very different estimates since they differ greatly in data sets, sample sizes and independent variables used in the analysis. Traditional studies analyze the effect of one agent activity on the outcome of another. Knowledge created by the innovative activity of the first agent is weighted according to a specific channel that is supposed to reflect the proximity between these two agents. But, a wide range of possible channels exists in the literature (Input-Output matrices, import, Foreign Direct Investment, technology flows, technological proximities, patent citations...), all of them lay on different assumptions and lead to different results. Although their implications in term of estimates variability is already proved, no measures have been conducted. This situation raises interrogations about the way we should interpret these results and the real impact of knowledge spillovers on innovative activity. Such interrogations are important since it is not uncommon to see authors willing to measure knowledge spillovers while they measure rent spillovers.

One way to deal with these issues is to conduct a traditional literature review on knowledge spillovers. The qualitative review unfolds the concept, the basis and the development of the empirical methodology used by the literature. Such a review sheds some lights on the different practices

## Chapter 1. Introduction

---

used and the implications of using specific channels for measuring spillovers. Although this approach is necessary, it is not sufficient because it suffers from a methodological selective bias to discount evidence unfavorable to the reviewer prior theoretical beliefs.

To avoid such bias, a complementary method was developed in medicine: Meta-Analysis. A Meta-Analysis is a set of quantitative techniques for evaluating and combining empirical results from different studies that distinguish themselves not only by the precise specification they adopt, but also by the underlying assumption and the type of information they use (Rose & Standley 2005 [151]). Thus, meta-analysis is essentially an analysis of analyzes, a combination of data from several studies to produce a single estimate and the explanation of variation between studies estimates by the moderator variables<sup>4</sup>. For our purpose, M.A. will give an idea of the true effect of knowledge spillovers, but also of the influence of moderator variables. What is the influence of the channels associated with spillovers? Are the effects of spillovers growing in time? What is the difference in the spillovers effect through the different level of aggregation?

While traditional studies investigate the effect of one innovative activity to another with some assumptions about the channel by which knowledge flows, a new literature using patent citations has proliferated in recent years. Patent citations provide a more direct measure of diffusion than has typically been used in prior researchs. The development of this literature was partly due, to the availability of patent citation data in a readable form for the US patents during the 1990s. Patent provides a unique range of details on the innovation. If an inventor chooses to protect his invention through a patent, he has the counterpart to disclose information concerning himself and the development of his invention. Specially, patent includes references or citations to previous patent and scientific literature. Using patent citations as a predicated view by which each inventor benefits from the work of those before, researchers are able to have a direct view on knowledge flows. In addition of the direct view that it offers, patent conveys important information on the invention that it protects and therefore allows detailed analysis of knowledge flows.

Explorations of knowledge flows with such data were first implemented by Caballero and Jaffe in 1993 [25], closely followed by Jaffe et al. (1993) [87]. After them, a large literature, led by Jaffe and Trajtenberg, has exploited this database in order to have a clearer look on knowledge diffusion. Using a structural framework for the analysis of citation patterns<sup>5</sup>, they track the influence of past inventions

---

<sup>4</sup>A moderator variable is a qualitative (e.g., gender, class...) or quantitative (e.g., level of reward) variable that affects the direction and/or strength of the relation between dependent and independent variables.

<sup>5</sup>The structural framework provided a model that estimates the different contributions to the citations process. This

across time, space, technological fields and institutions. But, due to the lack of European, and more widely global, counterpart database, empirical studies were almost exclusively conducted on US data (with the exception of few studies using the EPO database). One goal of this dissertation is to restore the balance in favor of European countries and of the Japan, and that thanks to the advent of the PATSTAT database.

The current work is as much relevant as patent data, taken from these offices (EPO and USPTO), are strongly biased toward domestic inventors inducing the well-known “home advantage effect”, Criscuolo (2006) [37]. Fortunately, the EPO has spent some efforts on the construction of a worldwide database, PATSTAT, that regroups all patent and patent citation data from almost all patent offices dispersed around the world. The utilization of citation data taken from different offices, assumes that all of them are a good indicator of knowledge flows. But the use of patent citations, not coming from the USPTO, into the analysis of knowledge flows, still raises some doubts and critics. One part of this dissertation is directed to investigate, explore and highlight the technical characteristics of patent citations for the seven major patent offices in the world, opening widely the window on the process of knowledge flows. The qualitative review of patent procedures will, once again, goes with a quantitative review of the principal characteristics of patent citations among these different offices.

Once the validity of using such data is assessed, we will, for the first time, explore the citation patterns for the G5 according to data taken from both the DPMA, the EPO, the INPI, the IPO, the JPO, the USPTO and the WIPO<sup>6</sup>. This exercise gives a much richer picture of the knowledge diffusion for the countries selected, allowing a measure of the speed of diffusion within and among the combination of these countries.

The richness in terms of information of patent data also allows patent to measure other facets of the innovative activity. Especially, it can be used as an input based indicator of technical change. One drawback of patent is its large heterogeneity, a lot worth nothing and a few worth a lot both in a monetary and in a technological perspective. Using citations as a control for technological opportunities they embody, through the structural framework mentioned above (Jaffe and Trajtenberg 2002), we are able to build knowledge stocks that are very helpful when R&D data are missing. This intuition was first developed by Popp (2002) [146] for the US energy technologies. In his formulation, patent

---

model, also called the citation function, identifies separately the contribution of citations lag distribution (obsolesce and diffusion of knowledge), fertility (for our purpose : productivity of knowledge embodied) and proportion to cite in the citation process. A detail review of this model is done in this thesis, but the reader can also refers to Hall et al.(2001) [71].

<sup>6</sup>The DPMA is the German Patent and Trade Mark Office; the INPI is the French National Industrial Property Institute; the IPO is the UK Intellectual Property Office and the JPO is the Japan Patent Office.

## Chapter 1. Introduction

---

productivities depend on the number of citations received. But, by developing an indicator based only on domestic knowledge, the author withdraws advances made by the international knowledge spillovers literature that set up the importance of foreign knowledge into the national productivity growth. In this dissertation, we extend Popp formulation to all industrial sectors of the economy, in a way that takes into account international knowledge spillovers.

Finally, all our previous developments would be useless if we do not illustrate the advances undertaken through the dissertation into a concrete study case. One of a fundamental issue nowadays for our economy is to deal with global warming, search for energy security and increasing world energy consumption. For all these issues, technological change stands to play a crucial role from both an environmental and an economical point of view. Reducing carbon emissions without dramatic reductions in output and consumption requires the use of new technologies. These may be as simple as improvements in energy efficiency, or in advanced technologies for generating electricity from solar power or capturing and storing carbon emissions from coal combustion. Thereby, the potential of technical change is related to concerns for energy supply and to the complexity of energy systems. Understanding the process of technological change is central to evaluate public policies intended to increase the overall potential solutions to environmental and energy challenges without compromising economic growth.

Literature identifies two main determinants in the process of technical change: demand-side and supply-side factors. Demand-side factors are derived from demand-pull theories and explain new inventions by the market demand for them. The most relevant demand pull theory for energy saving technologies is the induced innovation theory. Supply-side factors are derived from technology push theories and explain inventions by the technological opportunities that inventors could benefit from, i.e. focusing on the role of existing knowledge (expressed in term of stocks of knowledge). Unfortunately, data on supply-side are not readily available (R&D expenditures are not available at a technological level for private sector). To deal with this issue, we make use of patent citation data in order to estimate the technological opportunities available to inventors as explained previously. The construction of such input-based indicator allows us to investigate the determinants of innovation process for energy efficiency technologies together with the extent of international knowledge spillovers.

### 1.3 Overview

This dissertation presents three main parts, each of them are composed by two chapters. The first part is dedicated to an extensive review of the knowledge spillovers literature both with a traditional and analytical viewpoint. This part serves as the conceptual basis for the rest of the dissertation. The second part is dedicated to the investigation of patents and patent citations as indicators of technological change. Once again the analysis is double and unfolds both the qualitative and quantitative aspects. In parallel, this part proposes the development of the structural framework of the citation function in order to encompass international knowledge diffusion and to lay the groundwork for the construction of a knowledge input based indicator. Finally, the third part presents the empirical implementations of the structural framework presented in the second part. In one hand, we investigate how knowledge diffuses among the G5 countries for 13 manufacturing sectors and we evaluate its effects on national productivities, in the other hand, we will implement our indicator, in a case study, in order to estimate the different determinants of innovation in energy field.

**Chapter 2** Chapter 2 explores the notion of knowledge spillovers through a critical survey of the literature. It first studies the characteristics of knowledge as an economic good that is source of externalities. The second part identifies the two main externalities coming from research activities, i.e. the rent and knowledge spillovers. The third part, the most important one, presents the different channels used to measure knowledge externalities. There will be particular focus on the underlying assumptions related to their constructions and what their utilization implies. Channels presented in this part are: input-output transaction flows, the foreign direct investment, technological flow matrices, technological proximity matrices and patent citations. Finally, the last section is concerned about the geographical dependence of knowledge externalities.

**Chapter 3** Chapter 3 conducts a Meta-Analysis (M.A.) of the literature on knowledge spillovers. Because the rate of return of external knowledge differs greatly in function of the channel and the characteristics of studies, only a M.A. can emphasize the contribution of these moderator variables and identify a genuine value or authentic effect. We survey an expansive body of empirical literature (544 estimations from 69 papers). Knowledge spillovers effect presents considerable heterogeneity that is mainly due to the channel involved. If we focus on the most used channels, the elasticity of knowledge

## Chapter 1. Introduction

---

spillovers on productivity lies between 0.207 to 0.276 for inter-firm spillovers<sup>7</sup>, 0.263 to 0.323 for inter-sectoral spillovers<sup>8</sup>, 0.048 to 0.050 for international spillovers, 0.116 for intra-sectoral international spillovers<sup>9</sup> and 0.178 for inter-sectoral international spillovers<sup>10</sup>. The use of a Meta Regression Analysis (M.R.A.) allows, in addition of the identification of the genuine value, to highlight the presence of a strong publication effect that tends to affect positively the value of reported estimates through the literature. The estimated value of simple international spillovers is two times more important when we do not take into account the presence of publication effect. Finally, we find a significant increase in spillovers effect estimates over time, studies referring explicitly to data covering period before 1990 report systematically lower estimates.

**Chapter 4** Chapter 4 intends to give a first insight into the definition and specification of patents and how we can use them for different purposes for the measurement of economic knowledge based values. Emphasis is placed on the differences between the six major patent offices in the world. First, we review the patent procedures from the DPMA, EPO, INPI, IPO, JPO, and USPTO (respectively the German, the European, the French, the UK, the Japan and the US patent office) and second, we discuss the opportunities offered by patent data, arising from these offices, as a measure of the creation of new ideas and knowledge flows. Third, we present the patent citation function used in the analysis of patent citation determinants. This presentation comes back on the development of this function and presents its technical properties. We then transform this function in order to identify the productivity of knowledge embodied in specific patent cohorts and the usefulness of this knowledge for subsequent inventors. The last point details the construction of our indicator, i.e. a stock of knowledge, based on international patent and patent citation data, that takes into account international knowledge spillovers.

**Chapter 5** Chapter 5 presents in detail the PATSTAT database and its data management in order to extract the database that we use thereafter. We retain all patents application from 1970-2006 filed either at the DPMA, EPO, INPI, IPO, JPO, USPTO and the WIPO. The resulting sample embodies 6 204 213 patents with their citations. The construction of such sub-database allows us to fill literature gaps about practices in patent offices, other than the EPO and the USPTO. To be specific, we analyze

---

<sup>7</sup>without and with technology proximity weighted scheme

<sup>8</sup>Input-Output and technology flow

<sup>9</sup>Trade

<sup>10</sup>trade with Input-Output



different aspects of patent citations : (i) how many citations contain a search report, (ii) what is the origin of the cited patents and (iii) to what extent the search reports are similar between the offices considered (specifically how the citations are similar between the national patent application, filed at the DPMA, INPI, IPO or the JPO, and the EPO subsequent ones). First, the results show that the number of citations per patent is increasing over time for national offices whereas it is constant for regional offices. Second, patent citations suffer from a home bias toward patents belonging to the same office. And finally, patent search reports are highly similar between INPI, JPO and EPO (respectively they present a matching rate of 87 % and 60 %), meaning that these citations can be useful into the measurement of international knowledge flows. Concerning DPMA and IPO, the matching rate is quite lower with a respective rate of 44 % and 40 %.

**Chapter 6** Chapter 6 estimates the determinants of knowledge diffusion through the structural approach developed in chapter 3 according to data managed in chapter 4. We investigate knowledge diffusion within and among the G5 countries for 13 industrial sectors. Main results show that when we control knowledge by its productivity, the effect of geographic distance seems to decrease, for instance, knowledge coming from Japan diffuses comparably to knowledge coming from Germany for a French inventor point of view. Nonetheless, domestic knowledge is predominant, it diffuses almost twice as fast as foreign knowledge although the probability of domestic citations compared with other countries decreases with time.

Parameters estimated allow the construction of our input based indicator and the measurement of the contribution of foreign knowledge, i.e. international spillovers. In order to check the robustness of our indicator, we test its similarities with traditional indicators (R&D based indicators) and its explanatory power on economic growth. Our indicator and our measures of foreign stocks have passed both tests successfully. This validation process gives two important results. First it validates the use of patent citations, from the different offices selected, as a measure of international knowledge spillovers. Second, it validates the use of the model developed by Popp (2002) [146] for the creation of knowledge stocks that could compensate the lack of R&D data generally not available at the patent detail range.

**Chapter 7** Chapter 7 deals with a case study in which our indicator is very helpful. In this chapter we attempt to identify and estimate the two main determinants of energy technology innovation for four countries and eleven technologies. We begin by a literature review about the demand pull and technology push theories on innovation. Mechanisms linked to the Demand pull theory are easily

## Chapter 1. Introduction

---

identified and take the form of induction innovation mechanisms that can be approximated by energy prices. Technologies push mechanisms, representing the technology opportunities, are identified by the stocks of national and international knowledge that inventors could benefit from. Since data are not available at this level of analysis, we implement our indicator. The empirical analysis is composed of two parts. The first one develops the knowledge input based indicator which represents the knowledge available to inventors. The second part estimates the contribution of the two main determinants on the innovation process (i.e. linked on induce innovation and technological opportunities). Results show that high energy prices induce innovation with an elasticity close to unity. But the mechanism is not automatic and leaves space for public policies aimed at promoting innovation since we also find a strong respond to technological opportunities with an elasticity of 0.39 for domestic knowledge and 0.45 for foreign knowledge.



## Chapter 2

# Knowledge Spillovers: a literature review

### 2.1 Introduction

Since the seminal work of Robert Solow (1957) [167] on the decomposition of growth, many studies have investigated the role of knowledge as a factor of production. But the special characteristics of knowledge as an economic good give it special properties that are different from traditional inputs. These characteristics involve an incomplete appropriation of its return, which is the cause of strong knowledge externalities. Today, this notion is widely used to describe every transfer or diffusion of knowledge or technology between firms, industries or countries. Economists specialized in the endogenous growth theory, geographical economy or more generally those of the knowledge economy use this notion a lot (Godin 2006). Knowledge externality is quite an ancient notion. It appeared with A. Marshall in 1920 [120]. In *Principles of Economics*, he described for the first time knowledge externalities. He defined them under the term of external economy, because they come from “the general improvement of the industrial environment, in contrast with “internal economies” that come from “individual firms’ resources [...] from their organisation and their management”. What Marshall implies by “general improvement of the industrial environment” are the effects “tied to knowledge and technical progress’ growth” and which mainly depend on “on the aggregate volume of production in

the whole civilized world”.<sup>1</sup>

Thus, the existence of “external economies ” explains the geographical concentration of firms. Krugman used this intuition to establish the bases of the New geographical economy (1992) [101]. But the main role of knowledge externalities has been introduced by Romer (1986, 1990) [149, 150]. He completely renewed the vision of technical progress and growth by making growth endogenous. The assumption of diminishing returns, necessary for endogenous growth, is, generally, not compatible with the micro-economics equilibrium. To reconcile micro economic diminishing returns and non diminishing returns at a macro economic level, the main idea has been to introduce knowledge externalities in between.

After this theoretical renewal, empirical literature, initiated by Griliches’ works (1958,1964) [60,61], experienced a revival of interest. Thus, many works have been conducted on knowledge externalities, to confirm their existences and quantify their effects. Basically, most of these works measure the effect of an external knowledge stock on the firm’s, sector’s or country’s productivity. However, the importance of the knowledge’s transfer between these agents depends on their proximity. Economic literature used many measures of proximity (economical, geographical, technological...) and associated channels (for instance: investment goods for economical proximity, distance for geographical proximity, patent application for technological proximity), without really thinking of what these choices implied (see Cincera & van Pottelsberghe 2001 [28]). Comprehension and control of these channels is all the more necessary so their choices have a significant impact on the estimation of effects that concern us (van Pottelsberghe (1997) [183]). Excellent reviews of the measure of knowledge externalities literature have already been done. The most influential ones are, among others, Griliches’ 1992 [66] and van Pottelsberghe’s works [28, 183] (without forgetting Mohnen’s 1996 [123], Verspagen’s 1997 [184] and van Mij’s 1997 [182] contribution ). The most recent of these reviews is eight years old, and since that time, new channels have been developed, in particular those based on patent’s citations. The present work then offers a review of knowledge externalities literature with a special attention on the characteristics of associated channels.

The first part of this review focuses on the study of knowledge as an economic good, on how it flows and on its early studies. The second part identifies the two main sources of externalities

---

<sup>1</sup>“Again, it is to his interest also that the secrecy of business is on the whole diminishing, and that the most important improvements in method seldom remain secret for long after they have passed from the experimental stage. It is to his advantage that changes in manufacture depend less on mere rules of thumb and more on broad developments of scientific principle; and that many of these are made by students in the pursuit of knowledge for its own sake, and are promptly published in the general interest.”

coming from research activities, i.e the rent and knowledge spillovers. The third part, which is the most important one, presents the different channels used to measure knowledge externalities. There will be particular focus on the underlying assumptions related to their constructions and what their utilisation implies. Channels presented in this part, are: input-output transaction flows, foreign direct investment, technological flow matrices, technological proximity matrices and patent citations. Finally, the last section focuses on the geographical dependence of knowledge externalities.

## 2.2 An overview of knowledge externalities

### 2.2.1 Knowledge as an economic good

Knowledge as an economic good has three characteristics which are the cause of important externalities. We can consider economic goods under two attributes: their excludability degree and their non-rivalrous degree. The non-rivalrous is a purely technological characteristic. It allows the simultaneous use of a good by different agents. The excludability is both a technical and legal characteristic. A good is excludable if you can regulate its use in a discriminative way. The good “knowledge” has the unusual characteristic of being, in the same time, non rivalrous and imperfectly excludable.

- Knowledge is a imperfectly excludable good. Being fluid and easy to diffuse, it is difficult to fully control it. Knowledge can escape from entities that produced it and benefit to other agents without having to bear the cost of R&D. Thus agents who created it can't fully appropriate it and it can benefit to others for free. Literature uses the generic term of positive externality to designate this positive impact on others. In other terms, knowledge externalities designate the fact that an agent can benefit from a knowledge created by others without financial compensation (non-pecuniary externality) or with a compensation inferior to the costs of production of the knowledge (pecuniary externality). There are several channels for knowledge to flow: the informal ones (technical cooperation between engineers, mobility of the scientists...) and the formal ones (studies on new products launched on the market, technology forecasting through patent). Economic literature have shown that information about R&D decisions are known by rivals within six months, whereas technical details are within a year (Mansfield 1985 [118]).
- Knowledge is a non-rivalrous good. First, it can be used simultaneously by several agents. Second, using it does not destroy it. In other terms, an agent can use a knowledge an endless amount

of time without having to bear the cost of reproducing it. An endless number of agents can use this knowledge without depriving others to use it. So agents are not rivals in the consumption of knowledge. This characteristic comes from the fact that knowledge, as a resource, is inexhaustible and that the use of the knowledge by an other agent does not imply the production of it. This has an important impact on costs and prices. Since the marginal cost is null, the market can't follow the rules of price fixing on the basis of marginal costs.

- The third characteristic of the good “knowledge” reinforces externalities. It is tied to the fact that it is a cumulative good, because every knowledge can be the main factor for the production of new ones. This is illustrated by Newton’s famous metaphor: “If I have seen far, it is by standing on the shoulders of giants”. Knowledge being a cumulative good reinforces externalities, it is not just a consumption good that can be used an endless number of time. It is mainly a production good that allows production of new knowledge, which can be in turn production factors themselves.

Because of these three characteristics, knowledge is partially a public good. These are the cause of the importance of the differential between the social and private returns of R&D activity. Nevertheless those properties are also the cause of what is known as the knowledge dilemma. When producing knowledge leads to profit, retrieving all this profit is almost impossible : one can't keep the entire control of the knowledge. Some of the benefits are captured by others, which means that there are externalised. Thus, the private marginal return an inventor will receive from its work will be inferior to the social return. This is a typical lack of incentives situation, which leads to a insufficient level of private investment, from a social point of view. One solution is to give a temporary monopoly to the inventor to exploit his invention, through patent. A patent protects the inventor from the consumption of his knowledge in a classical production scheme, i.e as a merchant goods. Conversely, a patent allows knowledge to be used as a production good for the creation of new knowledge. In exchange of the protection patent offers, knowledge characteristics have to be made public.

### 2.2.2 Knowledge transfers mechanisms

Literature discusses extensively the way knowledge diffuses from one agent to an other. Firstly, distinction has to be made between tacit knowledge and explicit one (Cowan & Foray 1997 [34]). Knowledge is explicit if it is codified, accessible and understandable by everybody. In other term if it is stored or conveyed as symbols (scheme, article...) or incorporated in a tangible form (machine tool). In

contrast, tacit knowledge is not codified and can be diffused only by social interactions. Tacit knowledge is then much more dependent of distance, therefore it is localised. This can be summarized by grouping transfer mechanisms into five categories.

- By purchasing inputs in which knowledge is incorporated, which means the purchase of machine tools, equipments. . . to other firms (by reverse engineering<sup>2</sup> for instance)
- By knowledge which is publicly accessible: patents, press clippings. . .
- By the mobility of people from one entity to another, more precisely by people moving between firms (human capital turnover), through educational systems (from newly graduated to firms), from public researches entity to private ones. . .
- By innovation projects, realised in cooperation with different people with complementary knowledge and competences.
- By formal and informal contacts between agents implied in the innovation process.

It is obvious that knowledge diffused by the first two categories refers to explicit knowledge and the one of the last two to tacit knowledge. The first two categories are the more commonly used to measure knowledge externalities. These mechanisms assume that the more a firm is implicated in those two mechanisms, the more it can benefit from another firm's R&D. The third category is tied to geographical distances and is more close to the last two categories (Brenner 2007 [22]). For all of them three, the profit resulting from innovation process of another firm will be dependent of distance between the two of them. Since the development of communication technologies, knowledge cost of transport tends to be null. This knowledge is, most of the time, incorporated to human capital or merchant goods which bear a positive cost of transport.

### 2.2.3 The genesis

First works on spillovers tried to measure the social yield of one innovation. They focused more particularly on the farming sector because there were a lot of data and resources available<sup>3</sup>. First

---

<sup>2</sup>Reverse engineering is the process of discovering the technological principles of a device, object or system through analysis of its structure, function and operation.

<sup>3</sup>In an interview to the Journal of Economic Perspectives in 2000, Griliches explains, talking about the agricultural field: "That's where the data were; that's where the money was. And there was not much money for empirical research coming into the economics establishment, except for a little bit of money from the Rockefeller Foundation, until the Ford Foundation came in and started to put in money. Then came the NSF. There were very few resources for serious empirical research in mainstream economics. Then the computer came and changed everything."



researches that can be assimilated to the search for externalities measured social investment returns in agricultural research in the United States. Griliches (1958) [60], in his work, measured social return linked to the use of hybrid seeds. The methodology developed is based on the measure of the consumer surplus, which will be largely resumed in the following years. Thus, Peterson (1967) [143] measured the return of the researches on poultry farm, Schmitz & Seckler (1970) [160] on tomato harvests, Ardito-Barletta (1971) on corn in Mexico. . . We had to wait 1977 and the work of Mansfield et al. [119] to find such studies at an industrial level. They studied 17 innovations returns and found out a social return twice as high as the private return.

By measuring externalities from specific innovations, those studies only took the knowledge related to the successful innovations into account. They neglected a part of the R&D investments ; the ones that did not concluded to successfully results. Those cases studies don't take research expenditures as an externality source, but the innovations realised from these expenditures, i.e a sub-group. This can be problematic since most research expenditures led to knowledge creations, which do not always become innovation.

The following researches, wanting to be more inclusive, focused on the impact of the R&D expenditures on production growth or on total factor productivity (TFP). Again, first works have been conducted on the farming sector. The first one who worked on it was Schultz's (1954) [164]. He calculated farming TFP growth rate to measure the impact of public R&D investment on efficiency improvements and found a positive relation. Griliches (1964) [61] estimated public R&D elasticity on agricultural farming production and found out positive relation. Evenson's (1968) [47] and Huffman & Evenson's (1991) [78] studies followed. They brought up the question of geographical spillovers, i.e they assumed research expenditures made in a specific region can also have an impact on another region's productivity.

### 2.3 Two kinds of spillovers

The literature identified two different concepts in the notion of externality. There are two mechanisms by which R&D expenditures generate externalities. The first one allows to an agent to benefit from R&D expenditures through the purchase of a good which is R&D intensive, whereas the second one allows an agent to benefit from R&D expenditures without financial compensation.

With Griliches (1971,1979) [62,64], this distinction between these two aspects has been theoretically

well established. Rent externality or market externality is tied to the purchase of a good which is R&D intensive to a price inferior to the one reflecting its real value. It comes from the fact that producers cannot appropriate all the rent coming from innovation. So the consumer benefits from a good with a price not fully adjusted for quality improvements. Those externalities exclusively manifest themselves through economic transactions. They come from market forces and not from knowledge externalities. For Griliches, this mechanism is due to a measurement problem in economic transactions. If every price was hedonic, rent externalities would not exist. This is why Griliches does not consider this phenomenon as a “real knowledge spillover”, but rather as a measurement error. Therefore, he does not use the word “spillover” to describe this mechanism, even if it has been studied and modeled by many authors under the name of “rent spillovers” or “market spillovers”. The ICT industry (information and communication technologies) is a good example of such a sector, generating important market externalities. Consumers can acquire R&D intensive goods, such as computers, at prices which are not fully adjusted to the increase of quality. The power and the quality of computers increase regularly whereas their prices are stable or decreasing. Productivity gains are not carried forward into the prices. With constant productivity, these goods’ prices are inferior to those of the previous period : it is referred to the decrease of efficiency prices. More generally speaking, the increase of TFP of a specific sector depends on the increase of internal productivity of this sector and also on factors prices measurement issue, if prices don’t fully report the quality increase.

The second notion represents the “pure knowledge spillovers” as Griliches described them. It is the possibility for agents to benefit from knowledge produced by others without any financial compensation, because of the “non-appropriability” of ideas. This occurs when knowledge produced by a research team named  $i$  are borrowed and then contribute to the innovation process of team  $j$ . Those knowledge externalities are independent from economic transactions (such as intermediate good flows), because they have their own logic which is not necessarily the market one. In certain circumstances, the inventor can willingly create such externalities. Thus he can decide to publish his researches or file a patent<sup>4</sup> so they become public. But most of the time, externalities creation is unintentional and comes from practices such as technology sourcing<sup>5</sup>, reverse engineering, turnover... Generally speaking,

---

<sup>4</sup>In order to avoid any confusion, some precisions are necessary. In exchange of its invention publication, the inventor protect the commercial exploitation of his invention, he describes its characteristics in a way that a professional can reproduce it. This is pure knowledge spillovers. Everybody can consult the patent granted for free. On the other side, the patent applicant can license its patent. If he does it at a price that does not fully reflect the invention quality, this is a rent spillovers.

<sup>5</sup>Technology sourcing activity regroups all the learning and acquisition mechanisms of resources and skills from external knowledge. Thus we can explain the fact of a firm entering a new country, which is on the technological frontier,

commercial development and the use of new invention tend to diffuse the newly created knowledge, or at least to reveal several of its characteristics.

Theoretically, distinction between these two notions is easy, but it is not as simple from an empirical point of view, because of the difficulty to distinguish market externalities from knowledge ones. Van Pottelsberghe (1997) [183] sees two main reasons for that. The first one is that market externalities are often measured by economic transactions, which can be closely linked to knowledge externalities : “An example of similarity is given by the drug industry which buys a substantial part of its intermediate inputs, and most likely borrows ideas from the chemical industry”, van Pottelsberghe (1997) [183]. The second one is related to strong co linearity problems, which can appear in econometric estimations because of the possible similarity between two types of externalities. Griliches (1979) [64] even said that this problem is the main obstacle to the measurement of spillovers. In the following times, the first aspect of externalities/spillovers will be named market or rent externality and the second one knowledge externalities/spillovers.

### 2.3.1 Difference between private and social return created by knowledge spillovers

First mathematical formulations appear with Simon (1947), Grunfeld and Levhari (1962) and Chipman (1970) [26]. In these works, a firm’s productivity no longer depends only on its private R&D efforts, but also on efforts made by the sector. We pick up the Griliches (1992) [66] formulation, which describes a simple intra-industrial spillovers model. The production function is represented by a Cobb-Douglas with constant return on internal firm  $i$  inputs.

$$Y_i = C \cdot X_i^\alpha \cdot R_i^{1-\alpha} \cdot S^\kappa \quad (2.1)$$

$Y_i$  represents firm  $i$  output, which depends on its conventional inputs  $X_i$ , its own knowledge  $R_i$  and the aggregated knowledge of industry  $S$ .

Assuming that the knowledge level of the industry is defined by  $S = \sum_i R_i$  and that all firms face the same relative price of their factors.

It follows that  $\frac{R_i}{X_i} = \frac{1-\alpha}{\alpha} \cdot \frac{P_X}{P_K} = r$  where  $\frac{R}{X}$  does not depend on  $i$ . Then individual production functions can be aggregated by that way:

---

by its willingness to take advantage of the technological basis of the hosting country.

$$\sum_i Y_i = \sum_i C \cdot X_i \cdot (R_i/X_i)^{1-\alpha} \cdot S^\kappa = \sum_i C \cdot X_i \cdot r^{1-\alpha} \cdot S^\kappa = C \cdot r^{1-\alpha} \cdot S^\kappa \cdot \sum_i X_i$$

Knowing that  $\sum_i R_i / \sum_i X_i = r$ , we can deduce:

$$\sum_i Y_i = C \cdot \left( \sum_i R_i / \sum_i X_i \right)^{1-\alpha} \cdot S^\kappa \cdot \sum_i X_i$$

with  $\sum_i X_i = X_a$  and  $\sum_i R_i = S$ , it results :

$$Y_a = C \cdot X_a^\alpha \cdot S^{\kappa+1-\alpha}$$

The model easily shows that the aggregated coefficient of knowledge  $\kappa + 1 - \alpha$  is higher than the individual one  $1 - \alpha$ . The social return of knowledge is thus larger than its private return. This is because there are intra sectoral spillovers allowing firm to benefit from R&D efforts of other firms belonging to the same sector. This model can easily be extended to other kind of spillovers. For instance, it is possible to consider several sectors  $i$  into a same country. Then spillovers are inter sectoral and knowledge spill-overs between different industries. We can also consider that  $i$  represents a region or a country in a globalized economy, spillovers then are inter-regional/national.

In the previous example, we considered that all firms contributed in an identical way to the aggregated stock of knowledge. We can easily relax this assumption by including a weighted scheme  $w_i$  to the construction of the stocks. Then we have

$$S = \sum_i w_i R_i$$

$w_i$  is the weighted function, that can be interpreted as being as the part of  $i$  knowledge, which can be used by the industry. In the same way, we can imagine that all firms do not have the same access to knowledge. Then the industry stock of knowledge will have a different impact in function of the firm receiving the knowledge. Consider  $j$  the receiving firm:

$$S_j = \sum_i w_{ij} K_i$$

Where  $w_{ij}$  is the bilateral weighted function of knowledge spillovers, it represents the firm  $i$  knowledge part accessible to firm  $j$ . There are many ways to build this weighted function, which we will study

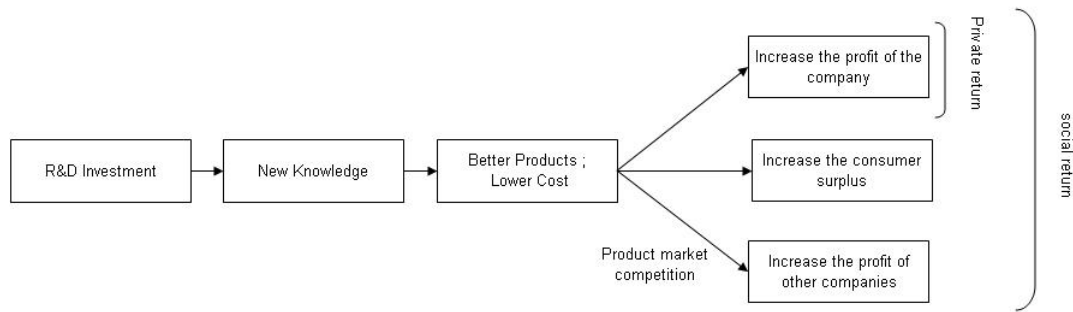


Figure 2.1: Rent spillovers return

later.

### 2.3.2 Difference between private and social return created by a rent spillover

Rent spillovers are generally treated in literature as a simple source of productivity measurement error (Griliches 1979 [64], Griliches & Lichtenberg 1984 [67]). Even if they are not considered as real “knowledge spillovers” by Griliches, there are the cause of a differential between private return and social return of the R&D (Jaffe 1998 [82]).

Rent externalities occur when a market transaction, linked to new goods or processes, has positive effects on firms which are not at the origin of this innovation. Rent externalities are the consequences of market forces and not of knowledge flows. They occur when the innovative firm is unable to appropriate all the returns of its innovation, i.e adjusting correctly and durably its sale price. If the firm realises a product innovation, it tends to sell its new product to a price which does not reflect totally the technological superiority incorporated in its product, because of competition pressure. . . Process innovations reduce the production cost of goods and allow the firm to sell them to a lower price. If those products are intermediate goods, the firms that buy them also benefit from the R&D expenditures realised by other firms. Again, it is function of the competition pressure, see figure 2.1.

In both cases, part of the surplus created is transferred to the buyers. These transfers are going to have a positive impact on firm productivity. They allow the use of relatively more productive or relatively less expensive inputs. This effect is only an accounting one, because, contrary to knowledge externalities, market externalities are just financial ones. The knowledge flows are not concerned. Market externalities are only going to have an impact on the profit or the productivity of the beneficiary firm, but not on its capacity to create new knowledge and to innovate. This is why some

authors recommend to study knowledge externalities through a “knowledge production function”<sup>6</sup> on which rent externalities have no impact. On the other hand, when we try to measure the impact of knowledge externalities through the “production function” approach, the choice of the channel is fundamental. A channel too close to market transactions will tend to be “contaminated” by the effect of rent externalities.

**Measurement problem** As well as being the cause of a pecuniary externality which manifests itself by an impact on productivity, rent externality is also at the origin of a measurement bias on productivity. The main problem implied by rent spillovers (i.e the fact prices don’t adjust to quality) is the introduction of a measurement bias in the price deflator. It has an impact on the productivity measurement. We come back to the production function of firm  $i$  that we used previously but we only take into account the internal input:

$$Y_i = A_i \cdot K_i^\alpha L_i^\beta$$

Where  $A$  is the global productivity of factors, which both depends on firm’s R&D expenditures  $R_i$  and on knowledge spillovers  $S$ .  $K$  is the capital and  $L$  the labor. Under the assumption of perfect market competition, the exponent  $\alpha$  (resp.  $\beta$ ) is equal to the share of capital input (resp. Labor) in the total production cost. Then TFP is given by:

$$A_i = \frac{Y_i}{K_i^\alpha L_i^\beta}$$

With  $\alpha_i = \frac{r_i K_i}{C_i}$ ,  $\beta = \frac{w_i L_i}{C_i}$  and  $C_i = r_i K_i + w_i L_i$ . As we focus on the effect of rent externalities on factors prices and on the productivity measurement, it is easier to work with growth rates, thus:

$$\Delta \ln A_i = \Delta \ln Y_i - \tilde{\alpha} \Delta \ln K_{K_i} - \tilde{\beta} \Delta \ln L$$

Where  $\Delta$  is the first lag operator, then  $\Delta \ln A_i = \ln A_{i,t} - \ln A_{i,t-1}$  represents the growth rate of  $A_i$ ;  $\tilde{\alpha}$  is the arithmetical average of  $\alpha_i$ ,  $\tilde{\alpha} = \frac{1}{2}(\alpha_t - \alpha_{t-1})$ . Let’s Consider in a first time that the firm  $i$  is a firm where innovative efforts lead to a product innovation. If so, the quality of the firm production increases but its goods prices are not fully adjusted to the improvement of incorporated

---

<sup>6</sup>See the next chapter for more details on the different approaches of R&D externalities measures.

quality because of the market forces. Thus if  $i$  produces a good twice as much efficient (or with a twice higher quality) and if price remains constant, the efficient price (price adjusted to the quality) is divided by two. However, most of the statistic institutes do not fully adjust sectoral prices to the quality improvement of new goods. So prices are overestimated which leads to the under estimation of the firm productivity. Consider  $P_i$ , price deflator of the sector in which firm  $i$  belong and  $Q_i$  its value added, constant price production is:

$$\Delta \ln Y_i = \Delta \ln Q_i - \Delta \ln P_i$$

Then the productivity is :

$$\Delta \ln A_i = (\Delta \ln Q_i - \Delta \ln P_i) - \tilde{\alpha} \Delta \ln K_i - \tilde{\beta} \Delta \ln L_i$$

We can easily see that an over-estimation of the industry price leads to an under-estimation of the production in volume, and consequently of the productivity. Conversely, a firm buying these goods of superior quality have an over-estimated growth of their productivity, because of the under-estimation of the factors deflator index. Consider firm  $j$  buying input with rent externalities:

$$\Delta \ln A_j = \Delta \ln Y_i - \tilde{\alpha} (\Delta \ln k_j - \Delta \ln P_i) - \tilde{\beta} \Delta \ln L_j$$

Where  $k_j$  is the capital in value, the input deflator  $P_i$  is always over-estimated. This over-estimation leads to the over-estimation of the buying firm's productivity  $j$ .

## 2.4 Different channels used in the measure of knowledge externalities

Generally speaking, studies on spillovers try to evaluate the effect of an activity of one agent on another. More precisely, they evaluate the effect of knowledge or ideas created by an agent on the innovative activity of the second agent. The presence and the importance of these effects depend on the agents characteristics and the distance (geographical and/or technological) that separate them. Two agents which are geographically close and belong to a similar technological space are more likely to benefit from their respective knowledge. Different studies on spillovers vary in function of the definition of

## Chapter 2. Knowledge Spillovers: a literature review

---

the agent (it can either be a firm, an industry, a sector, a region or a country), of the measure of the innovative activity  $IC$  (i.e. of the knowledge creation : R&D expenditures, human capital, patent application...), of the weighted scheme  $w$  (Input-Output matrix, technological distance. . . ) and of the innovative activity outcome measure  $OC$  (number of innovation, productivity. . . ). So we can sum up this relation by :

$$OC = f(Z, w.IC) + u$$

with  $Z$  is a vector of control variables<sup>7</sup>.

The main issue is about the choice of the parameter  $w$ . To measure the distance between agents, we use a weighted function  $w$ , which represents the characteristics of the channel through which knowledge transfers are supposed to happen. The choice of the channel depends on the level of aggregation considered (firm, sector. . . ) and on the kind of spillovers that we want to measure (inter/intra sectoral, and/or national/international). The easiest way to measure spillovers is to do it without weights, i.e.  $w = 1$ . Thus, Berstein & Nadiri (1989) [16] suppose that spillovers come from a common knowledge stock of the firm's sector. This knowledge stock is thus built on the R&D stocks of each firm belonging to the sector. That case corresponds to the specification presented in section 2.3.1; i.e.  $S = \sum_i w_i R_i$  with  $w = 1$  for firms belonging to the same sector and  $w = 0$  elsewhere. This specification takes only intra sectoral spillovers into account. Other methods use different weights in function of the origin and the destination of knowledge. Each flow of knowledge is then treated uniquely and the weights depend on both agents characteristics, i.e. of the agent  $i$  at the origin of the knowledge and of the agent  $j$  that benefits from. This weight  $w_{ij}$  can either be symmetric ( $w_{ij} = w_{ji}$ ) or not ( $w_{ij} \neq w_{ji}$ ). Early work, sought to introduce the concept of distance between agents, were made by Brown & Conrad (1967) [24]. They use Input-Output matrices to balance the flows of knowledge from one industry to another.

Seeing the multiplicity of existing weighted scheme approaches, it is necessary to define an analytical framework to classify them. We will use the classification defined by Mohnen (1996) [123], Verspagen (1997) [185] and van Pottelsberghe (1997) [183]. Mohnen makes the distinction between two kinds of approaches to measure distances. The first one uses the flows between agents (of goods, services,

---

<sup>7</sup>For a concrete example of the specification used in the study of knowledge spillovers see Appendix (Section 6.2) where the approach is presented by the production function.



patents, capital...). The more the flows are intensive, the more the distance between those two agents is small. The second one is about the construction of vectors which characterise the agents technology in a given space. This technological space can be constructed in different ways (patent citations for Jaffe (1986) [80], R&D expenditures for Goto & Suzuki (1989) [56] or the qualification of the R&D workforce for Adam (1990) [1]). As Verspagen underlines it (1997) [185], the first approach is based on economic transactions between agents, i.e under a user-producer principle, whereas the second one adopts a technological space perspective. Finally, van Pottelsberghe (1997) [183] extends this classification to three kinds of weights : the economic transactions matrices, the technological flows matrices and the technological proximity matrices. In the present work, the classification retained is based on the previous ones and proposes to extend them in a way that takes recent advancements into account.

### 2.4.1 Economic transaction flows

Literature generally uses three types of transaction flows matrices to weight externalities. On a national level, matrices used are the ones of the input flows between sectors (I-O, equipment goods, investment...). On an international level, two kind of matrices are used: is one related to trade flows (import, export...) and the other is based on FDI (foreign direct investment).

#### 2.4.1.1 I-O matrices

First spillovers measures were based on the hypothesis that knowledge diffusion is proportionate to the economic distance between agents. The more agents trade, the more they are economically close and the more the probability for knowledge transfers being realized is high. Thus, first studies assume that R&D expenditures diffuse proportionately to intermediate input flows. These flows can be measured in different ways, i.e. the weight parameter  $w_{ij}$  can come from different matrices. Brown & Conrad (1967) [24] use input-output matrices (I-O), Raines (1968) [147] uses the NSF (National Science Foundation) horizontal classification, Teleckyj (1974) [174] uses the capital flows and the intermediate input matrices. All these works assume that the importance of knowledge flows are linked to the intensity of economic transaction flows, i.e. to the economic distance. The more common formulation is Terleckyj's (1974) [174].

$$w_{ij} = \frac{M_{ij}}{\sum_j M_{ij}}$$

Where  $M_{ij}$  is the intermediate input flow between  $i$  and  $j$ .

All studies tend to demonstrate a positive and significant effect of these spillovers, but there is important variability in its empirical estimation mainly depending on the weighted scheme method. Nonetheless, these studies show that the estimated effect of these spillovers is higher than the one of the entity's own knowledge; i.e.  $\kappa + 1 - \alpha > 1 - \alpha$  in the equation. For recent application, see Unel (2008) [179], Lopez-Pueyo et al. (2008) [111] or Anon Higon (2007) [10].

### 2.4.1.2 Import flows

International transaction flows have been used in first studies measuring the impact of knowledge spillovers between countries. The principle is the same for economic transaction matrices used at an inter-sectoral level. We assume that knowledge is carried by international trade goods. The seminal study is Coe & Helpman's (1995) [30] (C-H). These authors used the import flows between countries to weight the knowledge flows. It implies that the more the countries are economically dependents, the more they benefit from their respective knowledge. Thus, C-H used the following formula as a way to weight foreign stocks of knowledge.

$$w_{ij} = \frac{M_{ij}}{Y_j}$$

Where  $M_{ij}$  is the import flow of intermediate goods and  $Y_{ij}$  the production output. C-H found a positive and significant impact with a foreign stocks elasticity higher than the national one. Lichtenberg & van Pottelsberghe (1998) [107] (LP) criticised the estimator developed by C-H. To them, it is too dependent from the aggregation degree. It is not the importance of the imports which determines the impact on the domestic productivity anymore, but the distribution of countries from which import goods come from. The more these countries are intensive in R&D, the more they have a strong impact on the domestic productivity. Keller (1998) [97] shows that weighting the R&D stocks by imports for OECD countries is not necessary to obtain similar results to those of Coe & Helpman (1995) [30] (Keller uses random variables to weight foreign R&D expenditures). International transactions flows vary from one study to another. Thus Xu & Wang (1999) [191] show by example that taking capital intermediate goods into account explains productivity gains better and resists to the Keller random

variable (1998) [97]. Lumenga-Neso et al. (2005) [113] show that it is important to take indirect commercial flows into account (C benefits from A without trading with A if A trades with B who in turn trades with C). Nevertheless, all these studies show a significant impact of foreign knowledge stocks on domestic productivity. For recent application, see Lopez-Pueyo, Barcenilla & Sanaù (2008) [111] or Cresp, Foster & Scharler (2008) [36].

### 2.4.1.3 Which spillovers aspects are measured through economic good flows?

By considering the importance of good flows as a measure of proximity, those studies mainly measure the probability for rent spillovers to occur. We remind that rent spillovers take place in the frame of intensive R&D good trades and under the hypothesis that agents who sell their goods cannot appropriate themselves all of the rent produced. So these works cannot clearly separate the effect of the two types of spillovers. By measuring the effects of the research efforts of the industry  $i$  on industry  $j$  productivity through I-O matrices, we measure two things. First, the effect of using more productive input coming from the industry  $i$  in the production process of industry  $j$  increases  $j$ 's TFP (rent spillover). Secondly, the knowledge transfer from  $i$  to  $j$  has a positive impact on  $j$ 's TFP too (knowledge spillover). The second effect (knowledge spillover) depends on the proximity between these two industries, which can partly be reflected by the intensity of their economic goods transactions. A second problem is raised by the hypothesis that knowledge transfers occur with intermediate goods flows. This hypothesis is restrictive since the distance is based on economic transaction and not on the geographical or technological distance. As Griliches highlights it, the photographic equipments industry and the scientific instruments industry do not usually trade together, but they are technologically close, so they can benefit mutually from their respective research progresses.

The problem linked to the use of these matrices is double. Firstly, they can not discriminate between the two types of spillovers. Secondly, they cannot measure knowledge spillovers between two industries that do not trade together but are geographically or technologically close.

### 2.4.2 Foreign direct investment: FDI

FDI (foreign direct investment) has been considered as an important channel of technological diffusion for a long time. Multinational firms, which invest in subsidiary abroad, can make technological transfers to or from their abroad subsidiary more or less intentionally. Thus a multinational company can transfer knowledge from its subsidiary to the hosting country. This point is developed by the theory that one

of the FDI determinant is to enter a market where the firm owns a technological advantage, which cannot be easily used by another way (i.e licenses). In this case, FDI inwards have a positive impact on local firms. Inwards flows (FDI-I) are defined by the investment of foreign firms in the domestic country. We generally define the relating channel in the following way :

$$w_{ij}^{FDI-I} = \frac{N_{ij}}{K_j}$$

With  $N_{ij}$  the FDI inwards and  $K_j$  the fixed capital formation or the production. Their effects have been highly studied. Whereas the first works were skeptical about the effects of FDI-I, more recent studies tend to prove the contrary. They show a positive effect on the productivity of the hosting country firms at an intra-sectoral level (see: Keller & Yeaples (2003) [99] for UK & Liu (2008) [108] for China).

The second channel linked to FDI is in the other direction and is the outwards flows or FDI-O. The underlying idea is that foreign investments are motivated by desire to exploit and take advantage of the technical superiority of foreign firms, which is known as “technology sourcing”. The firm then should use FDI as a “Trojan Horse” to appropriate some of the hosting country’s technology. We generally define the matching channel in the following way:

$$w_{ij}^{FDI-O} = \frac{O_{ij}}{K_j}$$

with  $O_{ij}$  the FDI outwards.

Studies realised with panel data allow to test the FDI-I and the FDI-O channels, but they do not agree on results. Lichtenberg & van Pottelsberghe (2001) [107] show that only FDI-O have a positive and significant impact on productivity, whereas Lee (2006) [106] and Bitzer & Kerekes (2008) [18] find out that only FDI-I have a positive impact on productivity.

One main explanation to these divergent results is that FDI are not only motivated by technological differences. As Driffield et al. (2008) [41] show it, different results are partly due to the fact that authors do not take into account the *ex-ante* motivation of FDI. It can be technological, but FDI can also be motivated by a difference of production costs between countries (even between developed countries, Love 2003 [110]). For example, Driffield et al. (2008) [41] show that in the UK, 75% of FDI-O are motivated by the search of cheaper labor costs and 25% have motivations linked to technology sourcing. Authors find out FDI-I and FDI-O, controlled by their *ex-ante* motivations have a positive impact on

productivity. FDI-I trying to benefit from production costs or to be used as technology sourcing have a negative impact on UK productivity. We should add that most of the studies about FDI estimate the impact of these flows on productivity without taking into account knowledge stocks in foreign and domestic countries. So these studies do not measure the impact of foreign knowledge on productivity, but they only measure the effect of the presence of these flows. In that case, FDI are not considered as a support channel for the R&D diffusion, but as an proper technological flow.

### 2.4.3 Technological flow matrices

First works using the patent data to measure spillovers led to technological flows matrices or technological concordance matrices. Schmookler's (1966) [163] was the first to attempt to create these matrices. They allow to identify sectors at the beginning of the creation of the innovation and sectors which use it. More precisely, these matrices determine the probability for a patent applied in one IPC classification (International Patent Classification) to have a specific IOM-SUE combination (i.e Industry Of Manufacturing, Sector of Use). We talk about probability because technological levels (IPC or USPC<sup>8</sup>) are wide and these patents can either come from and be used respectively in different sectors. We quote Johnson's (2002) [92]:

“For example in the IPC of B05 (sprayers and atomisers), a cosmetic atomiser might have an IOM in the glass container industry or metal valve industry, while a pesticide sprayer might have an IOM in the chemical fertiliser or agricultural machinery industry. Sectors of use (SOUs) would also differ, with the cosmetic atomiser used in the personal hygiene or cosmetic sector, and the pesticide sprayer used in field crop sector.”

These technological flow matrices can be built out of patent data in two different ways. Either out of a representative sample of patents, upon which specialists attribute an IOM-SUE combination. Or, these matrices can be built automatically, out of data distributed by the Canadian Patent Office, which, from 1976 to 1995 has automatically attributed these combinations.

The first technological flow matrices have been built by Scherer (1982, 1984) [156,157]. It has been elaborated with a database of 15.112 patents from 443 US firms. The individual study of these patents allows their assignation of a combination between their original industry and an industry where the invention could be used. The matrices thus defined associates an IOM-SUE combination to a sub-

---

<sup>8</sup>USPC is the USPTO classification

technological class. This matrix was created to estimate R&D flows between industries, i.e to estimate the part of R&D expenditures in industry  $i$  which is used in industry  $j$ . This method relies on the hypothesis that the part of R&D expenditures of  $i$  and used by  $j$  is proportionate to the part of the industry  $i$ 's patents used by industry  $j$ . After that, many works have focused on the creation of such matrices. The most famous and used one is the Yale matrix. It has been built by Evenson, Kortum & Putnam (1989) [48], thanks to Canadian Intellectual Property Office's data. The matrix has been built out of 250.000 patents granted in Canada between 1976 and 1993. It has been improved by Johnson (2002) [92], who renamed it the OECD Technology Concordance (OTC). This new version used patents registered from 1976 to 1995 and transformed the IOM-SUE classification, which was initially built on the Canadian standard, into the ISIC standard. To our knowledge, Verspagen, Moergastel & Slabbers (1994) [186] built the Merit concordance table, Bergeron and Le Bas (1997) [14] a matrix for France and Greif & Potkowiak (1990) [57] a concordance table out of German patents in the national nomenclature.

**Technological flows matrices: is it the same problem as goods flows?** Van Meijl (1995) [181] and Verspagen (1997) [184] note that those matrices are based on users-producers relation. Therefore, they link together the sector producing the knowledge to the sector buying it. So there is an economic transaction. These exchanges are measured thanks to patent licenses, so out of a merchant good which incorporates knowledge. It is then easy to compare technological flows matrices to transaction flows matrices. Knowledge is not carried by intermediate goods but by patents. We note here that knowledge exchanges are willingly and not knowledge spillovers, except if we consider that salesmen cannot sell at a price which do not reflect the real value of the innovation. So it is clear that, in that framework, knowledge transfers do not occur thanks to technological proximity, as it could happen in a case of engineering cooperation for instance. This is why van Meijl (1995) [181] and Verspagen (1997) [185] find that these matrices are a convenient approximation in the measurement of knowledge spillovers, but that they cannot clearly separate the two effects of spillovers.

Main technological flows matrices are built thanks to Canadian data and measure the technological flows in Canada. Thanks to these matrices, many studies focused on the evaluation of inter-industrial spillovers in Canada (Mohnen & Lepin (1991) [124], Hanel (1994) [72], Ducharme & Mohnen (1996) [43]...). Other studies applied those matrices to the study of technological flows in other countries (Englander et al. (1988) [46] on seven countries, Sterlacchini (1989) [172] on the UK, van Meijl (1997) [182] on France, Keller (1997) [96] on eight countries and Verspagen (1997) [184] on nine of them).

As van Pottelsberghe (1997) [183] points it out, the hypothesis that is made in these cases are indeed very strong, because it assumes that the structure of diffusion of the innovation between countries is identical. Studying three technological flows matrices (USA, Canada, France), van Pottelsberghe revealed their high heterogeneity and the difficulty to reflect technological flows out of their countries of origin (as an example, for the same country, results of R&D returns can vary between 0% with French matrices to 550% with USA's). More recently, Johnson (2002) [92] said, on the contrary, that the Yale's matrices do not impose Canadian inventions structure when applied to other countries. Finally, Scherer (2002) [158] came back over the estimation of his original 1982 matrix and compared it to Yale's. His comparison highlighted a big difference between these two matrices (he found a 0,560 correlation): "Whatever the reason, it seems clear that the Canadian patent classifiers and my late 1970 effort measured somewhat different phenomena"; which seems to be due to the aggregation level and to different "order flow" taken into account.

Finally the works based on patent data comes up against their inherent problems. Patent data have the advantage to be available for many countries, over long period of time and to a high level of technological classification (contrary to R&D data). But these data do not represent the entire output of the innovative activity. When the decision to apply a patent is an economic decision (between a situation where the invention is protected by patent or industrial secret), every inventions are not patented. This exhaustive problem is enlarged by the fact that all technological progress cannot be patented.

### 2.4.4 Immaterial transmissions channels

#### 2.4.4.1 Technological proximity matrices

Jaffe (1986, 1988) [80,81] is the first one who tried to measure proximity between firms in a different way than through transactional flows. His measure is based on the construction of vectors which characterised firm or industry positions in a technological space. Jaffe defines this space in function of the technological classes in which firms filed their patents. Two firms filing their patents in close technological classes will have more chances to benefit from the spillovers created by their respective knowledge. The firm positions are defined through the technological vector  $F$ :

$$F_i = \left[ \frac{P_{i,1}}{\sum_{z=1}^Z P_{i,z}} \dots \frac{P_{i,Z}}{\sum_{z=1}^Z P_{i,z}} \right]$$

Where  $P_{i,n}$  represents patent filed by firm  $i$  in the technological class  $n$ . So distances between firms are just function of the correlation between vectors  $F_i$  and  $F_j$ :

$$w_{ij} = \frac{F_i \cdot F'_j}{[F_i \cdot F'_i]^{1/2} \cdot [F_j \cdot F'_j]^{1/2}}$$

We then talk about angular separation. Jaffe's technological proximity is the first method which clearly distinguishes itself from the user-producer relation of previous approaches. He studies spillovers between firms and finds a positive elasticity between firms which are technologically close. For a recent application of this approach, see Botazzi & Peri (2003) [20], Bloom, Schankerman and van Reenen (2005) [19] or Lee (2005,2006) [105,106].

Goto & Suzuki (1989) [56], and more recently Adam & Jaffe (1996) [2] or Cincera (2005) [27], use the same logic, and consider that the technological distance between industries depends on the kind of R&D investments they make. Adam (1990) [1] uses the scientists part in several scientific fields. Aiello & Cardamone (2007, 2008) [6,7] use the different characteristics of firms such as ICT investments, internal and external R&D investments, the ratio between qualified and unqualified workers...

The inconvenient of this formulation comes from the symmetry in measuring technological proximity, i.e  $w_{ij} = w_{ji}$ . If two agents make the same R&D investments, in the same technological area, the importance of spillovers will be identical for both of them. Moreover, the positive effect found by Jaffe cannot be correlated to knowledge spillovers. This is Griliches' concept of "spatially correlated technological opportunities" (1992) [66]. If new exogenous opportunities come up in a certain technological area, and firms in this area then increase their R&D expenditures and productivity, this is not only being the result of knowledge spillovers.

But, since van Pottelsberghe work (1997) [183], technological proximity is considered as being the only channel allowing the correct measure of knowledge spillovers, because it is not linked to economic transactions. So it does not fit in the user-producer scheme which could generate rent spillovers, as said above. The underlying hypothesis to technological proximities implies that knowledge transfers are unintended. Thus, these transfers are supposed to realise themselves only because two agents are technologically close and not necessarily because they interact (intentionally or not). Of course, since 1997, other channels with similar advantages has been developed (notably thanks to patent citation data) and technological proximity is no longer the only valuable measure of knowledge spillovers.



### 2.4.4.2 matrices based on technological classification of patents

Verspagen (1997) [185] introduces two technological proximity matrices with similar characteristics. They are based on the main and secondary technological classification of patents. Patents are classified by technological class with the IPC classification. The technological areas, in which the patent is connected, is decomposed in two categories: "invention information" and "additional information". The first one represents technological areas claimed by the invention. The second one (non compulsory) is about technological areas that can get close to the invention but which cannot be claimed by the invention. So Verspagen (1997) [185] builds a spillovers matrix where the main classification of the patent represents the knowledge producer sector (thanks to Merit, Verspagen & ali's concordance table IPC-ISIC (1994) [186]) and where the secondary classification of the patent represents the sector that uses the knowledge. Los & Verspagen (2000) [109] find that spillovers associated to this channel have a more important impact on productivity than those associated to technological flow matrices.

### 2.4.5 Patent citations

For a long time, studies on knowledge spillovers explored the impact of the agent's external knowledge on its productivity. In other terms, those works have been made indirectly because :

"Knowledge flows are ... invisible; they leave no paper trail by which they may be measured and tracked" Krugman (1992) [101], but knowledge flows sometimes do leave trails; they can be found in patent citations.

With the development of databases on patents and their computerisation, patent citations became the most informative tool on knowledge flows. Thus, many works have been able to use these data in order to measure knowledge spillovers. Since Caballero & Jaffe's (1993) [25] seminal works, which developed the famous citation function and Jaffe, Trajtenberg & Henderson's (1993) [89] on geographical localisation of citations, an important literature has been developed using citations as an approximation for knowledge flows. All this literature is based on the hypothesis that citations are indeed a link to a pre-existing knowledge, which the inventor uses to develop his invention. The main advantage of these data is that they provide extensive information on agents, who patent their inventions.

#### 2.4.5.1 The mechanism of patent citations

Patent citations are close to bibliographical ones. Each inventor must list the patents or the other documents (NPL: non patent literature) upon which the current patent is built. Cited patents are

therefore a key measure of the knowledge used by the agent to develop his invention. Those citations are a legal statement that delimit the scope of property rights awarded by the patent. In this way, citations are the best way to control and reduce the importance of the novelty created by the inventor. Citations are established during the search stage and are added in the "search report". After receiving this document, the inventor can choose to pursue the patent procedure (which requires extra costs) or to abandon it if the rejection risk is too high. Citations process are different between the EPO (European Patent Office) and USPTO (US Patent and Trademark Office). For EPO patents, approximately 10% of citations are added by the inventor. In USPTO, the proportion climbs to 60%. This pattern can be explained by the legal character tied to patent procedures. In the US, the inventor, or his lawyer, must give the complete list of references describing the state of the art on which the invention is based (duty of candor). On the contrary, no legal disposition of that kind exists at the EPO. As a consequence, US inventors have a tendency to list every references, even if some of them are not directly linked to the patented invention. The USPTO examiners usually do not change the list made by the inventor, even if it exceeds the limits. On the contrary, the EPO is very meticulous in the search report. Examiners only list references that can affect the innovative character of the patent. Thus, by taking up citations made by Breschi & Lissoni (2004) [23]:

"According to the EPO philosophy a good search report contains all the technically relevant information within a minimum number of citations" Michel & Bettels (2001).

"The USPTO examiner purpose is to identify any prior disclosure of technology... which might be similar to the claimed invention and limit the scope of patent protection... or which generally, reveal the state of the technology to which the invention is directed" (OTAF 1976).

In this way the number of citations included in USPTO patents are highly superior to those of EPO patents. As Jaffe et al. (2000) [88] underline it, the recent explosion of patent citations by US firms emphasized this phenomenon, adding another charge on US examiners and reducing the quality of the citations.

### 2.4.5.2 Are Citations a good knowledge flows measure?

Patent citations have been used in several analysis of knowledge flows. The hypothesis that patent citation are good proxies for knowledge flows needs some development (briefly developed here but extensively detailed in chapter 4). First of all, citations can be added by the examiner or the inventor. Moreover, the citations added by the inventor can be made at different stage of the patent procedure

(at the application or during the research phase). So it is fundamental to discriminate which citations reflect knowledge spillovers, between the inventor of the cited patent and the citing patent. Citations which reflect the best knowledge flows appear to be those added by the inventor... but not only.

Jaffe et al. (2000) [88] statistically evaluate the proximity between USPTO citation data and knowledge flows, thanks to a survey realised with R&D managers. Half of the citations are not directly linked to knowledge flows because they mainly come from their addition by the examiner. Nevertheless, they conclude that, in the whole, citations provide a proxy for spillovers, and this all the more so that they are recent and made by the inventor.

"The likelihood of knowledge spillovers (...) is significantly greater (...) than the likelihood without a citation (... but...) a large fraction of citations, perhaps something like half, do not corresponds to any apparent spillovers (...) citations are a noisy signal of the presence of spillovers".

But considering that citations added by examiners are just some kind of noise in the analysis of spillovers is a bit restrictive. Inventors can forget to cite a patent representing a useful knowledge, that an examiner might point out. Moreover, citations realised by inventors can be added during the search stage, i.e after the invention stage, and do not reflect any knowledge flows that made the invention possible, see Thompson (2006) [175].

Breschi & Lissoni (2004) [23] studied the proximity between spillovers and EPO patent citations. For them, the fact that the inventor knows the cited patent is not a necessary condition to the realisation of a knowledge transfer. The inventor could know about the knowledge presented in the patent without knowing the patent itself. So knowledge transfer does not just realise in "face to face" condition, but can take place into a social chain. Thus, in their study of spillovers, authors keep citations made by examiners.

Duguet & Mac Garvie (2005) [44] study EPO patent citations in the eyes of the CIS (community innovation survey) for French firms (without any distinction of the citation origin). They find that patent citations have more chances to realise themselves between firms having joint activities which can lead to knowledge transfers (R&D cooperation, allowing the use of someone else's inventions, firm acquisition, joint venture...). This correlation (citations and activities) varies a lot in function of the source and of the destination of the technological transfer. Nevertheless, the existence of citations in the presence of activities that are knowledge exchanges sources tends to reinforce the idea that they are correlated to knowledge spillovers.

Finally, as Gay & Le Bas (2005) [52] summarises it, the common idea is that patent citations are

a valid but noisy tool to measure knowledge flows.

### 2.4.5.3 The use of patent citation data to measure knowledge flows

There are three ways to use patent citations in the study of knowledge flows. The first one is about the use of the function of citation that allows to measure the probability of citations between patents. The second one is the simple measure of the citations between patents, based on the hypothesis that, the more an agent cites another one, the more he uses the knowledge produced by this other one. Finally, the third one is the use of citations data in order to built matrices close to those of technological proximity. Whereas the last two approaches use patent citations to measure transmission channel of knowledge flows, the first one uses them as a direct proxy. This direct measure allows the study of factors which influence those flows (for complementary approaches see also : Mac Garvie (2005) [114], Peri (2002, 2005) [141, 142]).

**The citation function** First studies using patent citations as a measure for knowledge flows are based on the patent citation function developed by Caballero & Jaffe (1993) [25], improved by Jaffe & Trajtenberg (1996, 1998) [83, 86] and summed up by Hall et al. (2001) [71]. In those models, the probability for a patent  $K$  granted in year  $T$  to cite a patent  $k$  granted in year  $t$  is function of three factors:

$$P(k, K) = \alpha(K, k) \cdot e^{-\beta_1(T-t)} \left(1 - e^{-\beta_2(T-t)}\right)$$

- the diffusion process  $\beta_2$ ,
- the obsolescence rate  $\beta_1$ ,
- a scale parameter  $\alpha(K, k)$  functions of the patent  $K$  and  $k$  characteristics.

Thanks to this specification, we can easily define the variable  $\alpha(K, k)$  to take into account different factors which can influence the probability of citation. The interest of such an approach is that it allows to remove some bias that could affect the citation probabilities (changes in the examiner behaviors, specific year and/or specific diffusion speed...). Accordingly, Jaffe & Trajtenberg (1998) [82] define  $\alpha$  in a way that can take into account the specific characteristics of citations between countries. So it takes into account knowledge flows between those countries. Thus, this  $\alpha$  specification measures the

proximity between two countries  $l$  and  $L$  and then the importance of knowledge transfers between those two. Popp (2002) [146] uses this  $\alpha$  parameter in order to measure the productivity of a specific patent cohort. A relative important value of this parameter means that this cohort provides high technological opportunities. Pillu & Koleda (2008) [144] mixed these two approaches to evaluate the importance of knowledge transfers between countries in function of selected technologies. The estimated parameter represents the value given by the inventor of country  $L$  in the year  $T$  to the patent cohort published by inventors from country  $l$  in the year  $t$ . Knowing that, this approach is in the center of our research and next chapters are going to detail the different aspects of the citation function (the origin, former developments and future ones).

**The number of citation** Many studies use the simple count of patent citations between two agents to measure knowledge spillovers. The underlying idea is that if an agent  $i$  has more citations toward an agent  $j$  than others, it means this agent benefits more from the results of R&D processes of agent  $j$  than others.

As an example, Lee (2005) [105] uses the number of citations from the country  $i$  to the country  $j$  in order to weight country  $j$  knowledge flows towards country  $i$ .

$$w_{ij} = \frac{C_{ij}}{\sum_j C_{ij}}$$

Where  $C_{ij}$  is the number of citations sent out from country  $i$  to country  $j$  patents. Malerba et al. (2007) [116] and Lee (2006) [106] use this formulation to study inter sectoral and international spillovers.

**Technological flows and patent citations** Verspagen (1997) [185] uses USPTO citation data to build a technological proximity matrix "à la Jaffe". Technological proximity is set off by the number of citations between patents from different technological classes. In order to link technological classes to sectoral classification, Verspagen uses a technological flows matrix. So the author build a matrix where knowledge comes from the sector which filed the citing patent toward the sector of the cited patent. But this approach has not been very used (see also van Pottelsberghe, 1997 [183]).

## 2.5 Geographical distance

### 2.5.1 The different measures used

It is commonly admitted that geographical distance has an influence on the importance of knowledge transfers between agents. More precisely, geographical proximity makes the exchanges of non codified and tacit knowledge possible. The transmission of such knowledge is more about a "face to face" than codified knowledge (such as in a patent). The importance of this transmission then depends on the interaction between agents tied to different organizations (firms, laboratories, universities...). This interaction is easier if agents are geographically close, but also if they belong to the same country or speak the same language... Geographical distances, on a large scale, can be defined as metric distances or institutional ones. Thus we can class them into three groups.

- The first group includes every distance measures realised in a continuous way in metric or in a monetary form. The most commonly used formulation is the exponential one and uses the geographical distance  $d_{ij}$  between two agents :  $w_{ij} = e^{-\beta d_{ij}}$  ( $\beta$  being the localisation parameter. The bigger is  $w_{ij}$ , the more the knowledge are geographically dependent). This formulation allows to calculate the half life of the knowledge, i.e the distance that reduces the spatial interaction of the knowledge of 50%;  $d = \frac{\ln 2}{\beta}$  (Keller 2002 [98], Funke & Neibhur 2005 [50]). We can also consider the distance in terms of economic cost linked to physical capital (UPS distance) and to human capital (Air frame cost) transport costs Conley & Ligon (2002) [33].
- The second group gathers discrete geographical distances. We then consider the overall specified agents into a same geographical area (for instance all the firms within less than 300km, then those lying between 300 and 600km...(Bottazi & Peri 2003 [20], Orlando 2004 [137])) or institutional area (overall firms in the same state, country...).
- The third one gathers qualitative dummies measure: such as sharing the same country, the same language... Scherngell et al. (2007) [159]. This last category is not used in the construction of the weighting parameter  $w$  but allows, for example to test the effects of the belonging to a country or a language group (see next section).

Under the hypothesis that geographical distance conditions the face to face process, one can say that it then conditions the transmission of tacit knowledge. Then, it is theoretically justified to associate knowledge spillovers and geographical distance, see Doring & Schnellbach (2006) [42]. With that,

Channels with a geographical component
Trades
FDI
Patent citations
Channels with no geographical component
intersectoral goods flows
technology flows
Technological proximities

TABLE 2.1 – Channels and geographic dependencies

we can estimate spillovers from tacit knowledge. But if we are interested in codified knowledge, which is by definition less liable to distance, this method is less significant. As an example, a knowledge codified in a patent can be known by an agent, who would not have to meet the inventor before. This is all the more true since the generalisation of ICT and the computerisation of databases. Indeed, one can consult a patent anywhere in the world with a marginal cost close to zero. As we are going to see it in the next section, knowledge spillovers have a geographical component, but it does not explain every knowledge flow patterns. Thus, the weighting of knowledge spillovers only with geographical distances is restrictive, but it is a good method if used as a complement to channels which do not take into account the geographical pattern of the flows; like the ones that are the technological proximity matrices or based on intersectoral input flows (technological flows matrices and I-O). Thus, if we consider these matrices, two firms have the same weight factor, either they are at hundreds of kilometers from each others or in the same agglomeration. Of course, the geographical element is not pertinent in the study of domestic inter sectoral knowledge spillovers where the level of aggregation is the sector. On the contrary, transaction flows or citations already incorporate a geographical element. As an example, import flows are already in function of the geographical distance between two countries (this characteristic clearly reflects itself in the estimation of trading flows through the gravity equations). In the same way, and as we are going to see it in the next section, patent citations are geographically localised. So, considering patent citation as a measure for knowledge spillovers, it is also to take into account the geographical composition of knowledge flows. Table 2.1 presents the different channels in function of their geographical component.

### 2.5.2 What is the geographical component of knowledge spillover?

A major debate on knowledge spillovers is about their geographical components. If they are localised, they will explain most of the agglomeration effects. The underlying question is then to investigate

if knowledge spillovers are localised by nature or because of an *ex-ante* agglomeration effect. As an example, Pr Langer, from MIT, specialist of tissue engineering, receives localised patent citations (from Boston). Is it because living in the area of Pr Langer from Boston increases chances to learn from him? Or is it because Boston is an important area for tissue engineering and that people working on that subject are more likely to live in Boston? Or is it because Boston is one of the top places for tissue engineering because firms settle in the same area in order to benefit from knowledge spillovers? JTH (2005) [75].

JTH (1993) [89] have studied the localisation of spillovers represented by patent citations. They have found out that citations are extremely localised. Patent belonging to the same country (respectively: same state, same statistic zone) is 1.1 (resp 1.7, 3.3) more likely to be cited than any other one (in the control group). Thomson et al. (2005) [176] took up the study again, but they built a control sample with a much more finer technological disaggregation (more than the 3 digit level) and found that the localisation has a much more modest effect. A patent belonging to the same country (state, statistic zone) is 1.2 (1.6, 1.5) more likely to be cited than any other one.

Keller (2002) [98] thinks that R&D expenditures have a decreasing impact on the TFP in function of the distance. Half of the knowledge do not travel more than 1.200 km (it is the “half life”), but this geographical component of spillovers has a decreasing trend in time. In the same way, Bottazzi & Peri (2003)[12] tested the R&D effect on patent registrations at an European regional level. They found out that only R&D expenditures coming from less than 300 km (distance between these two regions frontiers) have a positive impact on the innovative activity of the region (but it is a very light one). Maursth & Verspagen (2002) [121] confirmed that result by pointing out that the distance has a negative impact on patent citations between European regions (EPO data). These ones are more localised between regions of the same country than between regions from different countries, even if sharing the same language. But Mancusi (2008) [117] by extending the Maursth & Verspagen sample (2002) [121] (not only at the European level) found a much less important geographical component. More recently and again using patent citations, Paci & Usai (2007) [138] showed that the impact of spillovers between European regions increased in time (the half life was 549 km in 1990 and became 625 km in 1999).

Peri (2002) [141] tested the effect of geographical variables on knowledge flows with the gravity equation. Knowledge flows between North American regions and European ones are represented by patent citations. The author showed that distance and the fact of belonging to two different countries



have a negative impact on knowledge flows (85% of knowledge remains local). On the opposite, technical specialisation have a positive impact on knowledge flows (knowledge flow between two regions with the same technological specialisation is 90% higher than between two regions with different specialisation). Finally, by comparing his work with classical gravity equation, Peri (2002) [141] shows that knowledge flows are less concentrated than trading flows. It is interesting to remark that the author finds out that knowledge created in the region on the technological frontier spreads less away than the one of average regions. In a same way, Mac Garvie (2005) [114] shows that patent citations are more important between geographically close countries sharing the same language, but that geographical pattern decrease in time.

Thompson (2006) [175] uses the same method than JTH (1993) [89] but studies the impact of the distance on patent citations in function of their origins. He finds that citations made by an inventor are more localised : they are 20% more likely to belong to the same country and 25% to same state in the US. The geographical component has a tendency to disappear with the patent life in the country (USA), but not at an international level. This might be because of the mobility of inventors (who are more mobile into the US than at an international level). Finally the geographical component tends to disappear if we only take into account intra-technological citations. These results are confirmed at an European level by Criscuolo & Verspagen (2005) [38]. The authors use the EPO data which allow to discriminate between citations made by the examiners from those made by the inventors and added during the search. Data show that citations made by inventors are more localised than those made by examiners. EPO examiners have a tendency to cite patents that do not belong to the same technological class. Finally Agrawal et al. (2008) [4] show that spatial and social proximity are “substitutable” in the knowledge accession.

## 2.6 Conclusion

Knowledge spillovers literature has been particularly prolific in the last two decades. Literature shows that those externalities are real and have a positive and significant impact on the economy. This literature highlights the positive effect of external knowledge stocks on the innovative process. This effect comes from characteristics of knowledge as an economic good. Which induces that knowledge cannot be fully controlled by the creator and that it benefits to other agent without pecuniary compensation. The different hypothesis made on the transmission channels lead to different weighted schemes for

knowledge external to the agent. Several schemes are developed in the literature, each of them having a specific implications that do not lead necessary to real knowledge spillovers measure. It is necessary to fully understand both the implications and the underlying hypothesis of the use of these channels to analyze correctly knowledge spillovers. This review tried to shed some light on this concern. But the study of these channels leads to questions on their impact in term of estimation results. For instance, is the use of matrices close to knowledge externalities (as opposed to those close to rent externalities) leads to a stronger measure of the externalities effects? The answer to this question is ambivalent because the value of the estimations depends on several factors inherent to these studies. Moreover, the traditional study of the value of these estimations does not take into account the significativity of the results. The natural perspective of such a review then would be to implement a meta-analysis of the studies on knowledge spillovers, in order to fully understand the “means and ending” of the different hypothesis of the channel in a quantitative way.

## 2.7 Annexes

### 2.7.1 Note on the adoption capacity of technology

The benefits retrieved from foreign knowledge also depends on the capacity of the country to absorb them. It is obvious that the economy must invest into absorption and imitation activities if it wants to benefit from these spillovers. Moreover, the necessity to absorb knowledge also depends on the distance of the economy from the technologically frontier. It is obvious that the leader benefits less from spillovers because of its position. Thus, Cohen & Levinthal (1989) [31] described two aspects of R&D: Innovating and Learning. The underlying idea is that R&D has two roles, varying in function of the distance to the technological frontier. The first role is conventional, i.e the stimulation of the innovative activity. The second one is about absorbing and imitating the foreign inventions. Being committed into a specific technological domain allows to acquire tacit knowledge. Then, this knowledge allows to understand and learn foreign inventions.

This hypothesis has been confirmed by Griffith et al. (2003) [58]. The authors show that R&D is both used for innovation and for adoption of a technology. The more the country is far from the frontier, the more the R&D returns are important. Thus, the more the country is R&D intensive, the more it benefits from knowledge transfers. Therefore, R&D effects can be under-estimated at a global level if only US results are taken into account. The United-States being a technological leader

in almost every sectors, R&D returns only take into account the innovation effects.

The importance of spillovers effects depends on the agent capacity to absorb external knowledge. The foreign knowledge stock elasticity on productivity  $\alpha_j^f$  is then defined by:

$$\alpha_j^f = \alpha_{0,j}^f + \alpha_{1,j}^f * E_j^d$$

Where  $\alpha_{0,j}^f$  is the constant,  $\alpha_{1,j}^f$  represents the capacity of the agent to take benefit from spillovers and  $E_j^d$  represents the cumulative experience of the agent  $j$  in innovative activities. The agent experience can be expressed through different ways. Thus, Guellec & van Pottelsberghe (2004) [69] and Griffith et al. (2004) [59] consider the R&D stock for  $j$ , i.e.  $E_j^d = S_j^d$ , Mancusi (2008) [117] the part of self citations in the total number of citations made by an agent. All these authors find out a positive elasticity.

## 2.7.2 Production Function

The production function is supposed to be a Cobb-Douglas. In addition to traditional inputs, the production function takes into account knowledge stock and knowledge spillovers:

$$Y = C.e^{\lambda t}.L^\alpha.K^\beta.R^\gamma S^\kappa \quad (2.2)$$

$Y$  represents the production,  $C$  is a constant,  $\lambda$  is the exogenous rate of technological change,  $L$  is the labor,  $K$  the capital,  $R$  the knowledge stock and  $S$  the knowledge spillovers.  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\kappa$  are respectively the elasticity of production, labor, capital, knowledge stock and knowledge spillovers. By rewriting equation (2.2) in log, we obtain:

$$\text{Log}Y = \lambda t + \text{Log}C + \alpha \text{Log}.L + \beta \text{Log}.K + \gamma \text{Log}.R + \kappa \text{Log}.S$$

Some studies use directly this formulation in level or in growth rates:

$$\Delta \text{Log}Y = \lambda + \alpha \Delta \text{Log}.L + \beta \Delta \text{Log}.K + \gamma \Delta \text{Log}.R + \kappa \Delta \text{Log}.S \quad (2.3)$$

Where  $\Delta$  represents the difference in log, i.e.  $\text{Log}Y - \text{Log}Y_{-1}$ . No hypothesis is made on return rate of traditional inputs. To take them into account, we rewrite in terms of labor productivity:

$$\text{Log}\frac{Y}{L} = (\alpha + \beta + \gamma - 1)\text{Log}.L + \beta\text{Log}.\frac{K}{L} + \gamma\text{Log}.\frac{R}{L} + \kappa\text{Log}.S \quad (2.4)$$

Where labor elasticity ( $\alpha + \beta + \gamma - 1$ ) in the equation 2.4 now measures the deviation from constant returns. Another approach commonly used is to estimate the effect of knowledge on total factor productivity (TFP):

$$\log TFP = \text{Log}Y - \tilde{\alpha}\text{Log}.L - \tilde{\beta}\text{Log}.K = \text{Log}C + \lambda.t + \gamma\text{Log}.R + \kappa\text{Log}.S$$

Where  $\tilde{\alpha}$  ( $\tilde{\beta}$ ) are calculated according to capital share in total cost.

$$\log TFP = \text{Log}C + \lambda.t + \gamma\text{Log}.R + \kappa\text{Log}.S \quad (2.5)$$

Equation can be thus estimated in growth rate:

$$\Delta\text{Log}TFP = \lambda + \gamma\Delta\text{Log}.R + \kappa\Delta\text{Log}.S \quad (2.6)$$

One alternative specification of equation, suggest by Griliches (1973) [63], is to estimate directly the growth rate instead of elasticity. By definition, elasticity of knowledge stock is  $\gamma = \left(\frac{\partial Y}{\partial R}\right) \cdot \left(\frac{R}{Y}\right)$ , thus 2.6 can be rewritten as follows:

$$\Delta\text{Log}TFP = \lambda + \rho\frac{\dot{R}}{R} + \eta\frac{\dot{S}}{S} \quad (2.7)$$

Where  $\rho(\eta)$  is the return rate of knowledge stock (knowledge externality).



## Chapter 3

# Knowledge Spillovers: A Meta-Analysis

### 3.1 Introduction

Since the seminal work of Solow (1957) [167] on the decomposition of growth, many studies have focused on the role of knowledge as a factor of production. But the special characteristics of knowledge as economic good give it special properties that are different from traditional input. These characteristics involve an incomplete appropriation of its return, consequently knowledge flows across agents and affects positively their productivities, in other words knowledge spill-overs. By knowledge spillovers, we mean that agents can benefit from ideas or knowledge created by innovative activities of other agents, without monetary compensation or at a price below the real value of the quality improvement. Two decades ago, Romer (1990) [150] emphasized the role of spillovers as a source of growth by allowing an escape to the fate of diminishing returns. Moreover, the presence of knowledge externalities generally induces a sub-investment in private R&D because agent doesn't take into account the positive effect of their own R&D investment on the productivity of other agents. The gap between social and private return (spillovers) leaves some places for public interventions and thereby require a good understanding of how spillovers flow.

The empirical literature provides evidence that significant knowledge spillovers do exist but highlights important variability in its empirical estimation. The Australian Industry Commission (1995) [32] points out, for R&D productivity, that "it's virtually impossible to be entirely consistent between studies because of the wide range of factors, including the use of different methodologies, a lack of clarity

in the way finding are presented, major structural differences in the countries covered, difference in the time period covered and whether the return on R&D are estimated from specifications that use R&D stocks or flow expenditure figure". Most of the problems highlighted above are also relevant in the analysis of knowledge spillovers simply because the knowledge spillovers variables are often constructed from R&D expenditures, and because we estimate their effects on productivity just like we do for the estimation of R&D return. Another source of variability in the empirical estimations is related specifically to the study of knowledge spillovers and is linked to the choice of the channel associated to these spillovers. A wide range of possible channels exists in the literature (I-O, import, FDI, technology flows, technology proximity...) and their implication in term of estimates variability is already proved but not measured.

This paper surveys an expansive body of empirical literature (544 estimations from 69 papers) in order to shed some lights and standardize the effects of knowledge spillovers by employing a Meta-Analysis. A Meta-Analysis is a set of quantitative techniques for evaluating and combining empirical results from different studies that distinguish themselves not only by the precise specification they adopt but also by the underlying assumption and the type of information they use (Rose & Standley 2005 [151]). In our case, the MA improves the assessment of spillovers effect through the identification and quantification of different factors that influence its estimates. Results show the presence of a publication effect which positively distorts the average measure of spillovers. Variability is mostly explained by the choice of channel involved and by the level at which the analysis is conducted. The MA regression allows the identification of the genuine effect underling the different studies that is, for instance, 0.235 for between firm spillovers when no weight is used and 0.305 when spillovers are measure through a technological proximity matrix, 0.26 for inter-sectoral spillovers measured with input-output matrix, 0.045 for international spillovers measured with trade flows, 0.116 for intra-sectoral international spillovers measured by trade flows and 0.101 for inter-sectoral international spillovers measured with I-O matrix and Trade flows.

This chapter is organized as follow: section 3.2 outlines the econometric framework which is commonly used in the measurement of spillovers effect; particular attention is made about most used channels. Section 3.3 introduces methodological issues regarding the MA approach; Section 3.4 presents the econometrics results of the Meta Regression Analysis on knowledge spillovers estimates and Section 3.5 concludes.

## 3.2 On the measurement of knowledge spillovers effects

Most econometrics studies on knowledge spillovers attempt to evaluate the impact of an agent activity on another. Most precisely they attempt to evaluate the impact of knowledge, created by an agent, on the innovative activities of a second. Presence and importance of spillovers effect depends on agent's characteristics and the distance between them (the distance between agents can be technologic, economic or geographic). Two agents close to each other are most likely to benefit from their externalities induced by their respective research activities. Studies differ in the definition of the agent (firm, sector, country), of the technology input indicator  $IC$  (R&D, Human capital...), of the technology output indicator  $OC$  (productivity, new product, patents, production factor cost...) and on the weighted scheme used to build the knowledge spillovers indicators  $w$  (I-O matrix, geographic distance, technology flow matrix...). To summarize, studies commonly used the following specification

$$OC = f(Z, w, IC) + u$$

Where  $Z$  is a vector of control variables. Since the seminal work of Griliches (1979) [64], we usually made the distinction between the rent spillovers and the “pure” knowledge spillovers. Rent spillovers arise when the prices of intermediate input purchased from other agent are not fully adjusted for quality improvement resulting from innovative activities. As such, they originate from economic transactions and are the consequences of measurement errors. By contrast, “pure” knowledge spillovers arise because of the imperfect appropriation of ideas. Knowledge spillovers take place when the knowledge generated by an agent contributes to the innovation process of other innovators. They occur when knowledge are “borrowed” by a research team  $i$  from a research team  $j$ . As opposed to rent spillovers, knowledge spillovers are not necessarily synonymous to economic transaction or measurement error. Although the distinction between the two concepts seems clear from a theoretical point of view, their empirical identification is far more problematic.

### 3.2.1 Specifications

The study of spillovers could be done by three main methods: with the production function approach, with the knowledge production function approach, and with the cost function approach. The production function approach estimates the effect of knowledge spillovers on production or productivity. The knowledge production function approach, initiated by Griliches & Pakes (1984) [139], estimates the



effect of research effort and knowledge spillovers directly on knowledge increment, generally approximated by patents (Brandstetter 2001 [21], Peri 2005 [142] for instance). The cost function approach estimates the impact of knowledge spillovers on production factors cost (Bernstein 1989 [15] for instance). Under dual theory, production function and cost function are related together, but the latter is more complex and requires more data that are not always available. For our case, we concentrate on the production function approach and knowledge production function that are the most used.

**The production function approach** Most studies use a simple production function framework. The common production function is assumed to be a Cobb-Douglas. In addition to the traditional input, this function includes knowledge capital at the agent level and external knowledge capital that the agent could benefit from.

$$Y = C.e^{\lambda t}.L^{\alpha}.K^{\beta}.R^{\gamma}S^{\kappa}$$

where  $Y$  is the output,  $C$  a constant,  $\lambda$  is the exogenous rate of technical change,  $L$  is a measure of labor,  $K$  is physical capital,  $R$  is the internal knowledge capital and  $S$  the external one (i.e. knowledge spillovers).  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\kappa$  are respectively the output elasticities with respect to labor, capital, own knowledge and external knowledge.

Taking logs:

$$\text{Log}Y = \text{Log}C + \lambda.t + \alpha\text{Log}.L + \beta\text{Log}.K + \gamma\text{Log}.R + \kappa\text{Log}.S \quad (3.1)$$

Some studies used directly this formulation in level (equation 3.1) or in growth (equation 3.2):

$$\Delta\text{Log}Y = \lambda + \alpha\Delta\text{Log}.L + \beta\Delta\text{Log}.K + \gamma\Delta\text{Log}.R + \kappa\Delta\text{Log}.S \quad (3.2)$$

where  $\Delta$  is the first difference. No assumptions are made on the return to scale with respect to agent input (excluding external knowledge). To account for this we rewrite equation 3.1 in term of labor productivity:

$$\text{Log}\frac{Y}{L} = \text{Log}C + \lambda.t + (\alpha + \beta + \gamma - 1)\text{Log}.L + \beta\text{Log}.\frac{K}{L} + \gamma\text{Log}.\frac{R}{L} + \kappa\text{Log}.S \quad (3.3)$$

Where the coefficient of the logarithm of labor ( $\alpha + \beta + \gamma - 1$ ) in equation 3.3 now measure the

departure from constant return. Another common approach used is to estimate the effect of knowledge on total factor productivity:

$$\text{LogTFP} = \text{Log}Y - \tilde{\alpha}\text{Log}.L - \tilde{\beta}\text{Log}.K = \text{Log}C + \lambda.t + \gamma\text{Log}.R + \kappa\text{Log}.S$$

Where  $\tilde{\alpha}$  ( $\tilde{\beta}$ ) are the share of labor (capital) cost in value added.

$$\text{LogTFP} = \text{Log}C + \lambda.t + \gamma\text{Log}.R + \kappa\text{Log}.S \quad (3.4)$$

Equation 3.4 could also be estimated in growth:

$$\Delta\text{LogTFP} = \lambda + \gamma\Delta\text{Log}.R + \kappa\Delta\text{Log}.S \quad (3.5)$$

An alternative specification of previous equation 3.5, suggested by Griliches (1973) [63], is to estimate directly the “rate of return” instead of elasticities. By definition the elasticities of internal knowledge stock is  $\gamma = \left(\frac{\partial Y}{\partial R}\right) \cdot \left(\frac{R}{Y}\right)$ , thereby equation 3.5 could be written as<sup>1</sup> :

$$\Delta\text{LogTFP} = \lambda + \rho\frac{\dot{R}}{Y} + \eta\frac{\dot{S}}{Y} \quad (3.6)$$

where  $\dot{R} = R.\Delta\text{Log}.R$  and  $\rho$  (respectively  $\eta$ ) is the rate of return with respect to internal (resp. external) knowledge.

Elasticity measures the percentage increase in output or total factor productivity that occurs in response to a one per cent increase in R&D. The rate of return measures the change in total output or TFP that results from a one unit (dollar) increase in R&D (the marginal product of R&D). The choice between the estimation of elasticities and rate of returns involves strong hypothesis. In both cases, underlying hypothesis is the constancy across agent and time of the estimated parameters. Hall (1996) [70] points out that constant return hypothesis across agents “could be a more plausible assumption than the equalization of sales or output elasticities”. To quote Weiser (2005) [189] “...to the extent that the production technology is specific to each firm, firms will use different factor shares, and if inputs are used at their competitive equilibrium levels, firms are unlikely to have the same output elasticities”. Unfortunately most of studies choose to estimate knowledge spillovers elasticities rather

---

<sup>1</sup>We simply use the relation :  $\rho = \frac{\partial Y}{\partial R} = \gamma\frac{Y}{R}$

than their rate of return.

**Knowledge production function approach** An alternative way to measure the impact of knowledge is to use a knowledge production function approach. This approach gives the relationship between newly produced knowledge (often proxied by patents) and research inputs. Thereby, studies using this framework focus directly on how resources devoted to research and development generate new ideas that could have productive economic use (Bottazzi & Peri 2003 [20]). Whereas in the case of production function the aim is to assess the impact of spillovers on productivity, the knowledge production function measures their effect directly on innovation. We can also assume that knowledge production function is a Cobb Douglas:

$$Q = C.R^\varphi.S^\phi$$

where  $C$  is a vector of specific agent characteristics,  $R$  is the internal knowledge capital,  $S$  the external one, (i.e. spillovers) and  $Q$  is some latent measure of new technological output (for our case it will be approximated by patents).  $\varphi$  and  $\phi$  are respectively the output elasticities with respect to own knowledge and external knowledge. Given that, the main difficulties in spillovers measure is the separation between real knowledge spillovers and rent spillovers, the second approach is preferred because it avoids a measurement problem linked to the deflator. The production of innovation does not require intermediate inputs and is not evaluated using prices, but simply require the quantity of innovations, it thus minimizes the role of rent externalities (see Griliches & Lichtenberg 1984 [67] to an extent discussion related to this item).

Finally, it's important to notice some points about knowledge stock. In all previous specifications, either knowledge stock in level or in growth is used. But the construction of knowledge stock is not always an easy way. In the lack of long data set on knowledge flows, studies commonly used the Perpetual Inventory Method namely PIM originally proposed by Griliches (1979) [64]. Suppose that the knowledge stock is built upon R&D expenditures  $R\&D_t$ , it follows that:  $R_t = (1 - \delta)R_{t-1} + R\&D_t$  and  $R_0 = \frac{R\&D_0}{g+\delta}$  where  $\delta$  is the depreciation rate and  $g$  is the average geometric growth rate for the R&D expenditures. This specification involves some assumption about the value of  $g$  and  $\delta$  parameters<sup>2</sup>. Some studies circumvent the problems by ignoring the depreciation of knowledge stock

---

<sup>2</sup>Johnson (2000) investigates the effect on estimated rates of return to the R&D stock when the depreciation rate was varied, and found that the rate of return was remarkably constant across different depreciation rates. Koysk

and  $\dot{R}$  is sometime approximated by R&D expenditure.

Thereafter we will make the distinction between the different measures of the effects of knowledge spillovers. Thereby, we adopt the following formulation:

- the rate of return estimated through a production function  $\eta$  is noted *Return Production Function (RPF)*,
- the elasticity estimated through a production function  $\kappa$  is noted *Elasticity Production Function (EPF)* and
- the elasticity estimated through a knowledge production function  $\phi$  is noted *Elasticity Knowledge Function (EKF)*.

### 3.2.2 Aggregation level

Spillovers estimations could be made at different aggregation levels. We principally distinguish three kinds of spillovers: between firms, between sectors (inter-sectoral spillovers) and between countries (international spillovers). Author could focus on one of the previous kinds of spillovers or choose to do a mix of them, thus it is common to see a study with both intra/inter-sectoral spillovers coming from domestic and foreign sectors. Eventually all theoretical levels of aggregation could be use and different combination could be made. We consider all kinds of aggregation with and from the firm level, this includes inter-firm, inter/intra-industry and inter/intra-national spillovers.

### 3.2.3 Spillovers channels

The main methodology used, in spillovers analysis, is to consider that knowledge diffuses proportionately to the level of potential relationship between agents. This relationship is represented by a parameter  $w_{ij}$  that represents the degree of closeness between two agents namely  $i$  and  $j$ . We will refer to this parameter as the weighted parameter and at the underlying method the weighted method or scheme. The external knowledge  $S_i$  that agent  $i$  could benefit from all other agents is thereby a weighted sum of agents knowledge stocks  $R_j$ :

$$S_i = \sum_{i \neq j} w_{ij} \cdot R_j \quad (3.7)$$

---

transformation allows the estimation of the decay parameter directly from the regression.

The choice of weighting parameters  $w_{ij}$  induces strong assumptions about the channel associated to knowledge diffusion and about the kind of spillovers involved. The kind of spillovers estimated is clearly dependent on the assumption made about the channel associated. Cincera & Pottelsberghe (2001) [28] attempt to point out what spillovers is measured according to the weighted scheme selected<sup>3</sup>. For our purpose we identify six different weighted schemes associated to the measure of spillovers:

1. Based on no weighted scheme. The external knowledge is an unweighted sum of the knowledge stock of other agents (for instance Bernstein & Nadiri 1989 [16] considers the unweighted sum of the R&D spending by other firms in the same sector). Knowledge diffuses with same proportion from all agents:  $w_{ij} = 1$ .
2. Based on econometric transaction. Assumption underlying is that knowledge spillovers is proportional to economic relationship (I-O matrices, import share, capital goods investment...). A common specification, according to Terleckyj (1974) [174], using I-O's table coefficient is given by:  $w_{ij} = \frac{M_{ij}}{\sum_h M_{ih}}$ , with  $M_{ij}$  is the materials flow between sector  $j$  and  $i$ .
3. Based on patent flow data. This approach, based on an idea of Schmookler (1966) [163] and first developed by Scherer (1982) [156], uses patent data to determine the probability that an innovation made in sector  $j$  is used in sector  $i$ . In other words this approach associates to a patent technology class a probability of IOM-SUE (Industry Of Manufacturing - Sector Of Use). Results of these associations is called "technology flow matrix".
4. Based on technology proximity. This approach, first introduced by Jaffe (1986) [80], associates to each agent a vector of characteristics  $F_i$  (this vecto can be defined on patent portfolio distribution, R&D investment across product field...). Distance between agents depends on the euclidean distance between their representative vectors:  $w_{ij} = \frac{F_i \cdot F'_j}{[F_i \cdot F'_i]^{1/2} \cdot [F_j \cdot F'_j]^{1/2}}$
5. Based on patent citation data. Patent documents also include reference to previous patents (citations). According to Jaffe et al. (1993) [89] patent citations could be taken as a paper trial of knowledge flows. External knowledge is thus dependents of the share of citation between agents. Proximity between agent depends on  $w_{ij} = \frac{C_{ij}}{\sum_h C_{ih}}$ ; where  $C_{ij}$  is the number of citations receive by the patents granted by agent  $j$  from the patent of agent  $i$ .

---

<sup>3</sup>For instance they considers that rent spillovers are most likely to be associated with economic transaction scheme whereas "pure" knowledge spillovers are most likely to be associated with agent technology proximity.

6. Based on geographic distance. This scheme made the assumption that spillovers is limited by geographic distance  $d_{ij}$  between agent  $i$  and  $j$ .  $w_{ij} = e^{-d_{ij}}$

Since the choice of channel is expected to have an influence on estimations it is important to distinguish them. Some studies estimates multiple channel in the same time (not a combination of channel, i.e. they include two spillovers variables  $S$  and  $S'$  in a same regression with two different channels) and thus raise the risk of multicollinearity, we will take into consideration this aspect.

### 3.2.4 Others explanatory variables

From a methodological point of view, we distinguish estimates obtained with cross-section data and panel data i.e. when variable entering the production function are taken in level or in time series dimensions (long differences, growth rate...) and when estimates are obtained with Ordinary Least Square estimators or with more sophisticated estimators (Instrumental Variable, General Least Square regressions...). From a temporal point a view, periods analyzed could have some influence on estimation. We distinguish the studies using data explicitly referring to the followings two periods: before and after 1990. Lastly, some authors argue that absorptive capacities or degree of openness could have an implication in the benefit of spillovers. We make a distinction when the studies take into account these aspects.

## 3.3 Meta-Analysis

Meta-analysis serves as a quantitative literature review, it is the analysis of empirical analyzes that attempts to integrate and explain the literature about some important parameters. Rather than using methodological selective bias to discount evidence unfavorable to the reviewer prior theoretical beliefs, as conventional reviews often do, Meta analysis forces the reviewer to include all research and permits it to employ more objective measures of research quality (sample size, methodological aspect, omissions of relevant variables, etc.), see Standley et al. (2008) [171]. In other words, meta analysis is essentially an analysis of analyses, a combination of data from several studies to produce a single estimate and the explanation of variation between studies estimates by moderators variables<sup>4</sup>. MA is specially used in medicine, where any one clinical trial usually has too few observations to achieve statistical validity.

---

<sup>4</sup>A moderator variable is a qualitative (e.g., sex, class...) or quantitative (e.g., level of reward) variable that affects the direction and/or strength of the relation between dependent and independent variables.

MA has recently been growing in popularity in economics (a special issue of the Journal of Economics Surveys was dedicated in 2005) and has been used to explain dozens of area of research (union wage gap, gender wage discrimination, reciprocal trade agreements, R&D productivity...). For our purpose MA will give an idea of the true effect of knowledge spillovers but also of the influence of moderator variables mentioned above. What is the influence of the channel associated with spillovers? Are the effects of spillovers growing in time? What is the difference in the spillovers effect through the different level of aggregation : within and between sectors, countries...?

### 3.3.1 Sample selection

First and most painful step in a Meta analysis is the creation of the meta base<sup>5</sup>. The selected papers are obtained via an extensive search in IDEAS<sup>6</sup>. We have retained papers including at least one of the following keywords in their titles:

“R&D spillovers”, “Research & development spillovers”, “knowledge spillovers”, “technological spillovers”, “innovation & spillovers”, “technology spillovers”, “technology flows” and “knowledge flows”.

The response to our research returns 411 articles. Among these papers, we have retained only those responding to specific criteria that permit their utilization in an MA process. First, paper must be in English and published. Second, it must include an empirical estimation of the effect of an external knowledge stock on a knowledge output indicator (an estimation of  $\eta$ ,  $\kappa$  or  $\phi$ ). Papers which don't make the distinction between own and external knowledge are rejected. Third, in order to avoid heterogeneity in the estimation sample, we eliminate papers that focus only on a particular technology or sector and that focus only on the effect of academic research on private productivity. Fourth, since we want to analysis spillovers at different level of aggregation we drop estimates related to firm which explicitly belong to different sectors or countries. Thereby, between firm spillovers refers to intra sectoral spillovers.

Final sample contains 69 articles providing 544 estimates. The list of papers and their main finding are presented in Appendix (Table 3.11 to 3.14). For a easier lecture of sample characteristics we split it in function of the measure of spillovers effects (*RFP*, *EPF* *EKP*) and in function of the level of analysis. The different level retain are : between firms (intra-sectoral spillovers), between sectors (within a same country, i.e. inter-sectoral spillovers), between countries without sectoral definition (international

---

<sup>5</sup>The search of studies is limited to paper published before January 2009.

<sup>6</sup><http://ideas.repec.org/>; IDEAS uses the RePEc database.

spillovers), between countries within the same sector (intra-inter spillovers) and between countries and between sectors (inter-inter spillovers). Sample characteristics according to previous nomenclature is provided in Table 6.2 (definitions of variables are provided in annex). Table 6.2 highlights the common use in spillovers estimation, for instance technological proximity is the main weighted scheme used in the estimations of firms spillovers and the third for estimations at the intersectorial level (within a same country). At the exception of international level (without sectoral definition) most of study use more than one channel in the same estimation traducing the volition to measure both rent and knowledge spillovers. At the international level (excluding between firms spillovers) most of study use import channel which is considered as a rent spillovers channel. Eventually, most of studies measure elasticity (at the exception of inter-sectoral study) and use a function production approach.

### 3.3.2 Dealing with multiple results per study

It generally appears that an article provides multiple estimates of the effect under consideration. The presence of more than one estimate is problematic because the assumption that multiple observations from the same study are independent draws becomes too strong. On the other hand, counting all estimates equally would tend to overweight studies with many estimates (Standley 2001 [168]). Many solutions have been proposed in the literature. Some authors add a fixed effect when study provides more than one estimates. Other authors choose to represent each study by one estimate. Its selection could be made under different characteristics: the preferred estimates, the medians or randomly selected...(Cipollina & Salvatici 2007 [29]). In this case, important information is lost in the grouping process and it is not clear which estimate one should use (Jeppensen et al. 2002 [90]). Bijmilt & Pieters (2001) [17] show that the procedure using a single value for each study generates misleading results. Furthermore, they demonstrate that the specific procedure selected (using one value per study or using the complete set) has a significant impact on the results and hence on the validity of the generalization of the meta-analysis. They conclude that procedures using complete set outperform procedure using a single value. In our purpose, most of the studies selected contain more than one estimate with a maximum of 30 estimates for a single study.

In order to highlight the implication of the selection process, I provide in Table 3.2 both fixed effect and random effect meta-estimates in function of their nature which are common in the area. The fixed effect is based on the assumption that a single fixed effect underlies every study, so that, in principle, if every study were infinitely large, every study would yield an identical result. This is the same as



### Chapter 3. Knowledge Spillovers: A Meta-Analysis

Firm	RPF	EPF	EKP	Total
<b>Number of obs</b>	4	80	34	118
<b>Number of studies</b>	2	14	5	19
<b>Channel</b>				
Pool	2	28	15	50
I-O	1	0	0	1
Tech flow	0	3	0	3
Tech prox	0	45	19	64
Geo	0	2	0	2
Tech prox+Geo	0	2	0	2
<b>Other</b>				
Nbr channel	0	60	30	100
Growth	1	4	2	7
Complex Reg	1	21	25	47
Cross section	3	53	12	75
Absortion	2	8	6	18

Inter-sectoral	RPF	EPF	EKP	Total
<b>Number of obs</b>	53	30	.	83
<b>Number of studies</b>	8	7	.	15
<b>Channel</b>				
I-O	24	21	.	45
Tech flow	18	6	.	24
Tech prox	8	3	.	11
Cit	3	0	.	3
<b>Other</b>				
Nbr channel	25	27	.	52
Growth	52	9	.	61
Complex Reg	0	12	.	12
Cross section	37	5	.	42

International	RPF	EPF	EKP	Total
<b>Number of obs</b>	9	213	9	231
<b>Number of studies</b>	3	23	3	27
<b>Channel</b>				
Pool	0	17	0	17
Trade	8	150	5	163
Tech prox	0	11	0	11
Geo	0	15	2	17
Cit	0	0	2	2
FDI	1	20	0	21
<b>Other</b>				
Nbr channel	3	74	0	77
Growth	9	84	2	95
Complex Reg	2	42	0	44
Cross section	7	37	0	44
Absortion	0	33	5	38
Open	0	76	9	85

Intra-sectoral International	RPF	EPF	EKP	Total
<b>Number of obs</b>	1	43	9	53
<b>Number of studies</b>	1	9	2	12
<b>Channel</b>				
Pool	0	2	4	6
Trade	1	39	0	40
Geo	0	0	5	5
FDI	0	2	0	2
<b>Other</b>				
Nbr channel	1	30	7	38
Growth	1	2	0	3
Complex Reg	0	21	4	25
Cross section	1	3	5	9
Absortion	0	1	2	3
Open	0	8	0	8

Inter-sectoral International	RPF	EPF	EKP	Total
<b>Number of obs</b>	12	47	.	59
<b>Number of studies</b>	3	8	.	11
<b>Channel</b>				
Pool	0	2	.	2
Trade+I-O	9	25	.	34
Trade+Tech flow	1	3	.	4
Trade+Tech prox	0	3	.	3
Trade+Pool	0	7	.	7
Pool+I-O	2	4	.	6
Pool+Tech flow	0	3	.	3
<b>Other</b>				
Nbr channel	9	46	.	55
Growth	9	12	.	21
Complex Reg	3	22	.	25
Cross section	9	2	.	11
Open	0	6	.	6

Table 3.1: Sample characteristics

assuming there is no heterogeneity across studies. By way of contrast, the random effects estimator assumes that the studies are estimating different treatment effects, drawn from a distribution whose mean is of interest. Methodological issues are detailed in Appendix.

If we look at the Table 3.2, the fixed and random effects based on study's minimum, median and maximum estimates provides very different value between and within them. Manifestly, there is a considerable heterogeneity between estimates, the fixed and random estimators are not similar in magnitude with perhaps the exception of minimum values and median value for rate of return and minimum value for *EKP*. One might be surprised by the fact that the fixed and random variables present in some cases the same three-digit value. When the values are similar, the associated statistics  $H2$  are very low<sup>7</sup>, indicating some form of homogeneity among the studies. Second, these values are often associated with few observations, below 10 in general. In this case, samples are easily influenced by some values (one or two) with a high explanations power (i.e. very low standard error). In all cases, we reject the null hypothesis of homogeneity among estimates (Q-test:  $p=0.00$ ) and the test  $H2$  and  $I2$  confirm the results of Q-test<sup>8</sup>. Concerning the estimates heterogeneity between the preferred values (between min, med and max estimates), the value falls by a minimum factor of 2 between maximum and minimum estimate for random effect and fixed effect for the KPF. However, all estimates are economically substantial and their confidence bounds are positive and reject the null hypothesis of no effect at the exception of minimum value for rate of return which is significant at 10%. Of course, a evident source of this heterogeneity comes from the aggregation of estimates realized at different level of analysis (the sample used aggregates different studies with different levels of analysis because we have mixed studies at micro level with studies at macro-level.). But as highlight the Table 6.2, considering separately the different level of analysis with the principle of one result per study leads to very few observations per case. In addition, most of considered studies use different assumptions that we would like to investigate, thereby by selecting one observation per study we will lose information about the effect of their assumptions on estimates.

---

<sup>7</sup> $H2$  is a measure of the amount of heterogeneity. In absence of heterogeneity the value of  $H2$  is 1 but this value increases with heterogeneity.

<sup>8</sup> $I2$  measures the proportion of inconsistency in individual studies. Values close to 100% represent very high heterogeneity.

<b>Return; production function</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.020	0.010	0.030	0.072	13	0.000	2.61	61.63%
	random-effects	0.020	0.010	0.030	0.073				
Median	fixed-effects	0.043	0.035	0.050	0.000	13	0.000	2.92	65.72%
	random-effects	0.043	0.035	0.050	0.000				
Max	fixed-effects	0.055	0.046	0.065	0.000	13	0.000	7.66	86.94%
	random-effects	0.063	0.052	0.075	0.000				

<b>Elasticity; production function</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.004	0.004	0.005	0.000	50	0.000	69.06	98.55%
	random-effects	0.007	0.006	0.008	0.000				
Median	fixed-effects	0.007	0.006	0.007	0.000	50	0.000	140.24	99.29%
	random-effects	0.041	0.040	0.042	0.000				
Max	fixed-effects	0.018	0.018	0.019	0.000	50	0.000	279.49	99.64%
	random-effects	0.147	0.145	0.150	0.000				

<b>Elasticity; knowledge production function</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.012	0.010	0.014	0.000	10	0.000	12.97	92.29%
	random-effects	0.012	0.010	0.015	0.000				
Median	fixed-effects	0.014	0.012	0.016	0.000	10	0.000	32.33	96.91%
	random-effects	0.015	0.013	0.018	0.000				
Max	fixed-effects	0.024	0.022	0.026	0.000	10	0.000	43.24	97.69%
	random-effects	0.027	0.024	0.030	0.000				

Table 3.2: Sensitivity of meta-estimates to the choice of “preferred” estimates

Given these results, considering that we would lose valuable information from studies estimating the spillovers effect and following the recommendation of Bijmilt & Pieters (2001) [17] we include in the meta-analysis all observations available. Table 3.3 provides combined estimates of MA in function of the nature of the measure and in function of the level of analysis. Difference between fixed and random effect confirms the important heterogeneity among estimates also confirmed by Q-test, H2 and I2 (at the exception of case considering few observations). The hypothesis of no effect is easily rejected for all models. By the way, it seems that the effect of knowledge spillovers estimated with a knowledge production is higher than those measured with a production function. Finally EPF seems to have the more heterogeneous values perhaps due to the high number of estimates include in the sample.

### 3.3.3 Publication Bias

Publication bias refers to the greater likelihood of research with statistically significant results to be reported in comparison to those with null or no significant results. Publication bias is a socio-economic phenomenon and has long been a major concern for Meta analysts. Publication bias could occur as the results of author’s empirical finding comportment. Authors may find results that are opposite to their expectations and decide to file these results away, leaving unpublished and unknown to other

### Chapter 3. Knowledge Spillovers: A Meta-Analysis

<b>Inter-Firm</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
RPF	fixed-effects	0.183	0.145	0.221	0.017	4	0.038	2.81	64.46%
	random-effects	0.183	0.145	0.222	0.017				
EPF	fixed-effects	0.016	0.016	0.017	0.000	80	0.000	334.00	99.70%
	random-effects	0.111	0.109	0.112	0.000				
KPF	fixed-effects	0.153	0.144	0.162	0.000	34	0.000	12.38	91.92%
	random-effects	0.191	0.179	0.203	0.000				

<b>Inter-sectoral</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
RPF	fixed-effects	0.057	0.043	0.071	0.000	53	0.000	3.93	74.58%
	random-effects	0.151	0.129	0.173	0.000				
EPF	fixed-effects	0.111	0.107	0.115	0.000	30	0.000	17.64	94.33%
	random-effects	0.117	0.112	0.122	0.000				

<b>International</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
RPF	fixed-effects	0.037	0.029	0.045	0.002	9	0.000	4.52	77.90%
	random-effects	0.037	0.029	0.046	0.002				
EPF	fixed-effects	0.005	0.005	0.005	0.000	213	0.000	24.20	95.87%
	random-effects	0.007	0.006	0.007	0.000				
KPF	fixed-effects	0.016	0.015	0.018	0.000	9	0.000	36.52	97.26%
	random-effects	0.016	0.015	0.018	0.000				

<b>Intra-sectoral International</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
EPF	fixed-effects	0.015	0.013	0.016	0.000	53	0.000	52.02	98.08%
	random-effects	0.032	0.030	0.034	0.000				
KPF	fixed-effects	0.118	0.104	0.132	0.000	9	0.000	4.57	78.12%
	random-effects	0.118	0.104	0.133	0.000				

<b>Inter-sectoral International</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
RPF	fixed-effects	0.115	0.083	0.146	0.004	12	0.003	2.54	60.62%
	random-effects	0.113	0.080	0.146	0.006				
EPF	fixed-effects	0.068	0.066	0.070	0.000	47	0.000	38.80	97.42%
	random-effects	0.087	0.084	0.089	0.000				

Table 3.3: Meta-analysis of spillovers effects on productivity output according to their nature

researchers. Furthermore, authors, reviewer and editors treat statistical results more favourably; hence, they are more likely to be published. The problem for summary is that publication bias makes empirical effect seem larger than they are. For example, publication selection distorts the average measure of the price elasticities of water demand by a factor of three fold or four fold, see Stanley (2005) [169]. Some meta-regression and graphical methods can differentiate genuine empirical effect from publication bias. In the presence of publication effect, the reported effect  $b_{js}$  will be proportional to its standard error. A simple test for publication is to use a Meta Regression Analysis:

$$b_i = \beta_0 + \beta_1 S e_i + e_i \quad (3.8)$$

In the absence of publication bias the reported effect will vary randomly around  $\beta_0$  the true effect

and independently of its standard error  $Se$ . Due to obvious heteroscedasticity, a common principle in the literature is to use a weighted least square (WLS) version which divides equation 3.8 by  $Se$ .

$$t_i = \frac{b_i}{Se_i} = \beta_0 \left( \frac{1}{Se_i} \right) + \beta_1 + v_i \quad (3.9)$$

Where  $t_i$  is the conventional t-value for  $b_i$ , the intercept and slope of coefficient are reversed from previous equation. Equation 3.9 could be estimated by OLS. Egger et al.(1997) [45] argue that the conventional t-test of the intercept of equation is a test for publication selection, named the funnel asymmetry test or FAT. Similarly testing  $\beta_0$  can be considered as a test of authentic effect, corrected for publication selection, named the precision test or PET (Stanley 2005 [169]).

### 3.3.4 Meta Regression Analysis

The standard meta regression analysis usually entails a conventional regression model:

$$b_{js} = \beta_0 + \sum \alpha_k X_{k,js} + w_{js} \quad (j = 1, \dots, J, s = 1, 2, \dots, S) \quad (3.10)$$

Where  $b_{js}$  is the  $j$  th reported estimates, from a total of  $J$ , of the knowledge spillovers impact (in our case  $\eta$ ,  $\kappa$  and  $\phi$ ) in study  $s$  from a total of  $S$  studies,  $\beta_0$  is the “true” value of the parameter of interest,  $X_{k,js}$  is the meta-independent variable which measures relevant characteristics of an empirical study and explains its systematic variation from other results in the literature,  $\alpha_k$ 's are the meta-regression coefficients, which reflect the biasing effect of particular study characteristics, and  $w_{js}$  is the meta-regression disturbance term (Stanley & Jarrell, 1989, p 165 [170]). In order to control to the “publication effect”, we add the standard error of the related studies estimate  $Se_{js}$ .

$$b_{js} = \beta_0 + \beta_1 Se_{js} + \sum \alpha_k X_{k,js} + w_{js} \quad (j = 1, \dots, J, s = 1, 2, \dots, S) \quad (3.11)$$

Due to obvious heteroscedasticity, we use a weighted least square (WLS) version of equation 3.11.

$$t_{js} = \frac{b_{js}}{Se_{js}} = \beta_0 \left( \frac{1}{Se_{js}} \right) + \beta_1 + \sum \frac{\alpha_k X_{k,js}}{Se_{js}} + w'_{js} \quad (j = 1, \dots, J, s = 1, 2, \dots, S) \quad (3.12)$$

## 3.4 Econometric results

### 3.4.1 FTA and PET tests

Figure 3.4.1 presents the kernel density estimates<sup>9</sup> of the effect sizes. The mean value of parameters is 0.26 and the median is 0.13. Parameter value range from -2.41 to 3.43, though the majority of coefficients are clustered between 0 and 0.2. These simple statistics do not make use of any information on the precision of each estimate.

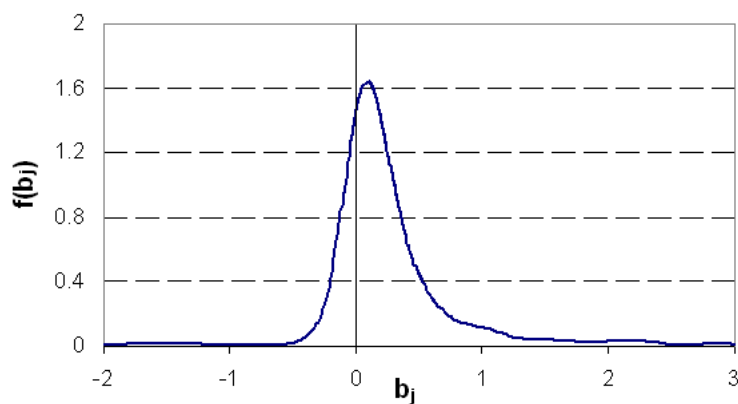


Figure 3.1: Distribution of the knowledge spillovers effects ( $\gamma$  and  $\rho$ )

Figure 3.4.12 graphs the knowledge spillovers parameter estimates with the precision of their estimations (i.e.  $1/Se_j$ ). This graph is named a funnel graph because in the absence of publication selection it should look an inverted funnel (Sutton 2000 [173]). The symmetry is a necessary condition to exclude publication bias. Even though this figure resembles to a funnel, it does not present a clear symmetry. With the exception of few negatives estimates, it appears that the left side of the graph has been cut. But the visual interpretation of the funnel graph may be ambitious, for this reason FAT is generally associated with the MRA test of equation 3.9.

<sup>9</sup>Kernel density estimation is a non-parametric way of estimating the probability density function of a random variable.

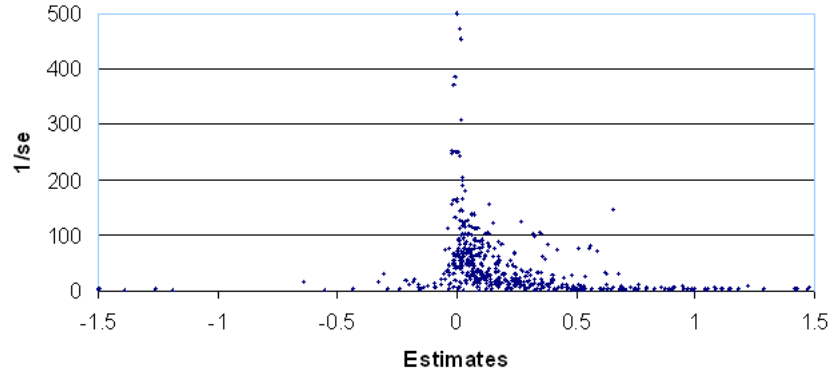


Figure 3.2: Funnel graph

Table 3.4 reports the result of the FTA and PET tests. Regression results of equation 3.9 present evidence of publication bias for all spillovers measure, in the light of test we reject the absence of publication effect ( $\beta_1 \neq 0$ , p-value=0.00). The slope of these MRA also serves as a test for genuine empirical effect corrected of publication bias. These PET do not find a significant evidence of a positive effect of knowledge spillovers for two of the three samples considered, only effect of knowledge spillovers estimated through a knowledge production function seem to be significant.

Sample	Return; production function		Elasticity; prodution function		Elasticity; knowledge production function	
Intercept	1.580	0.227(***)	4.883	0.822(***)	3.184	0.740(***)
1/Se	0.011	0.012	0.001	0.002	0.009	0.003(***)
Obs	79		413		52	
Adj R-squared	0.004		-0.002		0.036	
S.E. of regression	1.322		8.473		3.319	
Fisher F-stat[p-value]	1.331	[0.252]	0.164	[0.685]	2.890	[0.095]

S.E reported in parenthesis are heteroscedasticity-consistent  
 \*\*\*,\*\*,\* denote statistical signifiance at the 1%, 5%, 10% level respectively

Table 3.4: FTA and PET test for the complete sample

This result is surprising in regard of the whole literature but before concluding to a not significant genuine effect we might look further in the sample characteristics. We should think that estimates are too heterogeneous to be aggregated into a single significant value. Note that for the two samples which present no significant genuine effect the F-test associated is strongly not significant meaning that we must look for other variables to explain  $t_{j_s}$  and it is that we will do in the next section.

### 3.4.2 MRA econometrics results

We now turn to attempt an explanation of the variation in the estimates across sample ( $\eta$ ,  $\kappa$  and  $\phi$ ). The specification of the MRA equations is not a easy way. Since the common characteristics of spillovers estimations are function of the level of analysis (between firm spillovers, between sectors...) we will implement the MRA for each level in order to catch up the effect of the moderator variables, thus we will implement 5 MRA as specified in equation 3.12. Concerning the true effect, we will estimates one  $\beta_0$  for each kind of measure (*RPF*, *EPF*, *KPF*) in all MRA estimations when it is possible.

#### 3.4.2.1 Definition of the moderators variables

Concerning the channels associated to spillovers, we add in the MRA, a dummy *Pool* equal to 1 when study doesn't use any weighted scheme for the construction the spillovers variable (when  $w_{ij} = 0$  in equation 3.7) and 0 otherwise. In the same way, we add a dummy *I-O* when the study uses a matrix based on economic transaction, *Trade* for the use of international transaction, *Tech flow* for technology flow matrix, *Tech prox* for technology proximity matrix, *Geo* for geographic distance, *Cit* for patent citation and *FDI* for foreign direct investment flows. For firms and inter-sectoral international spillovers the combination of channel follows the same methodology, for instance *Tech prox + Geo* is equal to one when the study combines both geographic distance and technology proximity channel.

Because multiple dummies are added to account for differences induced by the channel used, we define one of them as the reference or as the base method that we won't include into the MRA. We choose this channel in a way which it would be present among all the different kind of measure (*RPF*, *EPF*, *KPF*). For instance, for between firm spillovers we will set the pool scheme as the reference method whereas for inter sectoral spillovers we will set the *I-O* scheme as the reference method. Negative MRA estimates associated with a specific channel means that the use of this channel in the estimation of the effects of spillovers leads to a value lower than that obtained with referenced method.

Concerning the other variables, the dummy *Nbr channel* is equal to 1 when the study include more than one spillovers variable into a same regression; the dummy *Growth* is equal to 1 when the study estimations are made in growth rate or lag difference; the dummy *Com\_reg* is equal to 1 when the estimates are obtained through advanced econometric methods; the dummy *Cross section* is equal to 1 when the study uses cross-sectional data; the dummy *Absorption* and *Open* is equal to 1 when the study uses respectively a measure of absorptive capacities and openness in the construction of its spillovers variable *S*. The summary of all the moderators variables used in the MRA are provided in



Table 3.15.

#### 3.4.2.2 MRA of Spillovers at a firm level within a same sector

Table 3.5 reports the meta-regression results of studies reporting spillovers between firms. Results show an authentic spillovers effect with a positive and significant impact<sup>10</sup> on innovative activities about 0.207 for knowledge production elasticities, 0.202 for production function and 0.108 for rate of return. Note that this value hold for the assumption of knowledge spillovers effects are estimated with the reference weighted method that is the pool scheme (no weights). These values are higher if spillovers variables were measure with different weighted method because of the positive sign of the moderator variables associated with the others channels (at the exception of I-O method). For instance *technology flow* estimate presents a positive value meaning the use of technology flow matrix in the construction of spillovers variable leads to higher estimate than if we don't use weight. Concerning the other moderator variables, only one of them is significant. The use of multiple channels in the same estimation has an expected negative size effect.

#### 3.4.2.3 MRA of intersectoral spillovers within a same country

Table 3.6 reports MRA results for inter-sectoral spillovers within a same country. Results also find a genuine spillovers effect that is about 0.322 for EPF and 0.241 for RPF. MRA estimation doesn't find significant difference in effect size according to the use of different channels with the exception of technological proximity which presents a negative size effect. The use of cross section data has a unexpected negative size effect. The use of multiple channels as well as the use of data referring exclusively to period before 1990 has a negative size effect. Finally, the use of advanced methods of regression leads to lower estimates.

#### 3.4.2.4 MRA of Spillovers between countries

We now turn to international spillovers without sectoral definition, results are providing in table 3.7. The authentic effect is positive and significant for the three measures of spillovers. It is about 0.04 that is significantly lower than for national spillovers. Results suggest that all channels have a significant impact on estimates at the exception of *FDI*. Patent citation and no weight scheme lead to

---

<sup>10</sup>at the exception of KPF, but this elasticity become significant when it is associated with technology proximity weighted scheme

Variable	Coefficient	S.E.	t-statistic	P-value
Std. error of regression = 9.87766      F (zero slopes) = 9.35322 [.000]				
Adjusted R-squared = 0.499883      Log likelihood = -429.666				
C	2.13788	1.05195	2.03231	[.045]
EPF	0.207462	0.111846	1.85489	[.066]
KPF	0.202236	0.134066	1.50848	[.134]
RPF	0.107736	0.063701	1.69129	[.094]
CHANNELS				
Pool	base method			
I-O	-0.102893	0.089981	-1.1435	[.255]
Tech Flow	0.320739	0.099701	3.217	[.002]
Tech Prox	0.068499	0.060172	1.13838	[.258]
Geo	0.078426	0.113333	0.691998	[.490]
Tech Prox + Geo	0.094609	0.100941	0.937273	[.351]
OTHER VARIABLES				
Nbr Channel	-0.230978	0.106692	-2.16491	[.033]
Growth	0.092901	0.102426	0.907007	[.367]
Complex Reg	0.028083	0.030219	0.929317	[.355]
Cross section	0.090477	0.097643	0.926612	[.356]
Absorption	-0.071291	0.088956	-0.801418	[.425]
1989	-0.116075	0.082059	-1.41452	[.160]

Standard Errors are heteroskedastic-consistent

Table 3.5: MRA of between firms spillovers

higher estimates comparing to the use of trade channel. At the opposite, technological proximity and geographical distance schemes reduce the value of estimates. The variable *Open* which traduces the presence of a measure of openness in the construction of spillovers variable seems to have a comparable and negative effect than the Absorptive variable at the exception that it is significant. Once again, multiple channels lead to lower estimates. The variable growth is negative meaning that the use of variable in time difference has a negative size effect.

### 3.4.2.5 MRA of Spillovers within sectors and between countries

The results of intra sectoral spillovers between countries are provided in Table 3.8. Here we have drop the only estimates for the rate of return. Results present also a positive and significant authentic effect that is about 0.319 for EPF which is higher than for simple international spillovers. Elasticity of KPF is not significantly different from zero but this elasticity concern only 4 estimates which could be not representative. The moderators variables associated with the channels are significant. Whereas *Trade*

Variable	Coefficient	S.E.	t-statistic	P-value
Std. error of regression = 1.93522      F (zero slopes) = 15.2602 [.000]				
Adjusted R-squared = 0.634909      Log likelihood = -166.670				
C	1.48655	0.252951	5.87683	[.000]
EPF	0.322666	0.106822	3.02059	[.003]
RPF	0.241609	0.24717	0.977502	[.332]
CHANNELS				
I-O	base method			
Tech Flow	-0.059519	0.051257	-1.16119	[.249]
Tech Prox	-0.091095	0.051889	-1.75555	[.083]
Cit	9.10E-03	0.107278	0.084844	[.933]
OTHER VARIABLES				
Nbr Channel	-0.15108	0.092872	-1.62676	[.108]
Growth	0.214301	0.222219	0.964371	[.338]
Complex Reg	-0.058212	5.90E-03	-9.86113	[.000]
Cross section	-0.078929	3.78E-03	-20.8767	[.000]
1989	-0.399785	0.09524	-4.19765	[.000]

Standard Errors are heteroskedastic-consistent

Table 3.6: MRA of inter-sectoral spillovers

is negative the geographic distance and the *FDI* are positives. Other significant results are as expected at the exception of *Open* which presents similar results than previous regression.

### 3.4.2.6 MRA of Spillovers between sectors and countries

Regarding inter sectoral and international spillovers results in Table 3.9, we find a genuine positive and significant effect about 0.18 both for EPF and RPF. Authentic effect are higher than simple inter country spillovers and lower than intra sectoral-international spillovers but this remarks has to be taken with conscious since the reference channel is not the same. Concerning the channels, only the pool scheme is not significantly different from the *I-O & Trade* weighting scheme. Values associated with *trade & pool* channel present lower estimates although for it is the opposite for the others. Most important, channel size effect are *technological flow* and *technological proximity* matrices when they are associated to *trade*. Concerning others variable unexpected results is for *Nbr of channel* and *C-S* that present opposite value than expected.

---

Variable	Coefficient	S.E.	t-statistic	P-val
Std. error of regression = 3.08737		F (zero slopes) = 5.75088 [.000]		
Adjusted R-squared = 0.236548		Log likelihood = -579.895		
C	2.28352	0.471098	4.84722	[.000]
EPF	0.04753	0.012958	3.66809	[.000]
KPF	0.038603	0.013033	2.96191	[.003]
RPF	0.045006	0.015142	2.97231	[.003]
CHANNELS				
Trade	Base method			
Pool	0.082354	0.014279	5.76741	[.000]
Tech Prox	-0.062865	0.0298	-2.10956	[.036]
Geo	-0.028678	0.012536	-2.28767	[.023]
Cit	0.346511	0.017933	19.3225	[.000]
FDI	2.08E-03	8.14E-03	0.255992	[.798]
OTHER VARIABLES				
Nbr Channel	-5.68E-03	2.86E-03	-1.98404	[.049]
Growth	-0.028094	0.012378	-2.26976	[.024]
Complex Reg	-2.83E-03	2.46E-03	-1.15383	[.250]
Cross section	0.012062	0.02225	0.542093	[.588]
Absorption	-0.015156	0.010958	-1.38303	[.168]
Open	-0.022409	0.01153	-1.94351	[.053]
1989	0.058577	0.064191	0.912545	[.363]

---

Standard Errors are heteroskedastic-consistent

Table 3.7: MRA of international spillovers

### 3.4.2.7 General remarks

Let's turn to general conclusion about common variable through the different MRA. First when studies refer explicitly to data covering the 1990 and /or period afterward, they seem to get higher estimates (the moderator variable *1980* is negative on four of the five MRA). This result confirms recent studies (Wieser 2005 [189] for instance) that highlight the reinforcement of spillovers effect over time notably with the development of ICT.

Concerning variables introduced to explain the diversity of estimates from a specification point of view, the addition of absorptive parameters in the construction of the spillovers variable doesn't seem to have a significant impact on estimates through the different studies. The use of absorptive capacities variable is generally included in order to take into account the second characteristics of R&D (the two face of R&D of Cohen & Levinthal 1989 [31]). This hypothesis cannot be confirmed here since absorptive moderator variable, *Absorption*, is never significant. At the opposite the use of

Variable	Coefficient	S.E.	t-statistic	P-val
Std. error of regression = 3.22017		F (zero slopes) = 9.66305 [.000]		
Adjusted R-squared =0.670876		Log likelihood = -127.116		
C	4.73947	0.79727	5.94462	[.000]
EPF	0.328126	9.57E-03	34.2968	[.000]
KPF	3.94E-03	0.061733	0.063827	[.949]
CHANNELS				
Pool	base method			
Trade	-0.178196	0.067409	-2.64351	[.012]
Geo	0.728833	0.099136	7.35182	[.000]
FDI	0.226667	0.067533	3.35636	[.002]
OTHER VARIABLES				
Nbr Channel	-0.146903	0.063782	-2.3032	[.027]
Growth	-0.024426	0.01082	-2.25743	[.030]
Complex Reg	-0.033781	0.011258	-3.00055	[.005]
Cross section	-0.61396	0.10614	-5.78445	[.000]
Absorption	-8.24E-03	0.014493	-0.568832	[.573]
Open	-0.130819	0.066661	-1.96245	[.057]
1989	-0.22189	0.072634	-3.0549	[.004]

Standard Errors are heteroskedastic-consistent

Table 3.8: MRA of international - intra-sectoral spillovers

an *Open* variable seems to lead to lower estimates for studies which use it. This result is not expected since theory suggests that more a country trade which another more it's able to benefit from external knowledge.

As expected, estimations using more than one estimates per regression tend to provide lower estimates (it is true for all the MRA at the exception of inter-inter spillovers, see Table 3.9, for which this parameter is positive but presents very low value). This result confirms the presence of strong collinearity arising when authors attempts to integrate multiple proxy for spillovers effects in the same regression generally to take into account both rent and "pure" knowledge spillovers.

Concerning a methodological point of view, all MRA with significant values for the *Cross section* or *Growth* variable present negative values meaning that estimations in cross section and/or in growth rate provide lower value for estimates. These results are expected for the *Growth* moderator variable but not for the *Cross section* one. At the opposite, estimations made with simple estimation techniques, *Complex reg*, provide lower results than those using simplest estimators.

Concerning the channels associated to spillovers, generalization appears to be more difficult. We

Chapter 3. Knowledge Spillovers: A Meta-Analysis

Variable	Coefficient	S.E.	t-statistic	P-value
Std. error of regression = 2.1569      F (zero slopes) = 23.7039 [.000]				
Adjusted R-squared = 0.845685      Log likelihood = -120.415				
C	0.686743	0.570606	1.20353	[.235]
EPF	0.178368	8.05E-03	22.1697	[.000]
RPF	0.181269	0.027778	6.52562	[.000]
CHANNELS				
Trade + I-O	base method			
Pool	-0.048092	0.031463	-1.52851	[.134]
Trade + Tech Flow	0.47895	0.016146	29.6635	[.000]
Trade + Tech Prox	0.469904	0.01977	23.768	[.000]
Trade + Pool	-0.083072	0.036918	-2.25016	[.029]
Pool + I-O	0.167804	4.68E-03	35.8911	[.000]
Pool + Tech Flow	0.167125	1.42E-03	117.299	[.000]
OTHER VARIABLES				
Nbr Channel	4.43E-03	1.85E-03	2.40025	[.021]
Growth	-0.106477	0.031507	-3.37952	[.002]
Complex Reg	-0.077356	0.032857	-2.35435	[.023]
Cross section	-0.12013	0.017186	-6.98991	[.000]
Open	-0.530166	3.84E-03	-138.03	[.000]
1989	-0.324345	0.021851	-14.8434	[.000]

Standard Errors are heteroskedastic-consistent

Table 3.9: MRA of international - inter-sectoral spillovers

have seen that the true effect of the estimates is greatly dependent of the channel used. This difficulty is reinforced by the lack of common channel through the different studies.

In order to summarize results about the true effect, we will re-estimate all MRA equations with all different channels as the reference weight method. For instance, MRA equation related to inter firm spillovers will be estimated five more times, one for each channel. Results of the exercise are provided in Table 3.10. This table shows that most of the channel and specification (kind of measure) used lead to a positive and significant effect. If we retain the authentic effect that is related to the most numbers of estimates reported we can attempt to make some generalization. By this way the authentic effect is 0.276 for firms, 0.323 for sectors, 0.048 for countries, 0.116 for intra sectoral international and 0.178 for inter sectoral international. Not surprising intrasectoral spillovers is higher than intersectoral spillovers both at national and international level. It seems that spillovers associated with *technology flow* are always higher than those associated with *technological proximity*. *Citation* is always the channel associated with the higher value. The use of a matrix without geographic dimension, likes

*technological proximity*, is not relevant for the study of international spillovers. The simple use of geographic distance is ambiguous because this channel presents high value for inter firms spillovers and intra sectoral international but not for simple international spillovers. Concerning the level of analysis of study, simple international spillovers present the lower authentic effect but when a sectoral definition is added the value could be 4 times higher. Concerning national spillovers values are relatively close between intra and inter sectoral with some higher values for intra sectoral (inter firms spillovers).

Level	Kind	Channel	Estimates	S.E.		Nbr Obs	
Firm	RPF	pool	0.108	0.064	*	3	
		I-O	0.005	0.132		1	
	EPF	pool	0.207	0.112	**	28	
		tech flow	0.528	0.167	***	3	
		tech prox	0.276	0.108	**	45	
	KPF	pool	geo	0.286	0.035	***	2
			mix	0.302	0.032	***	2
		tech prox	pool	0.202	0.134		15
			tech prox	0.271	0.123	**	19
Sector	RPF	I-O	0.242	0.247		24	
		tech flow	0.182	0.235		18	
		tech prox	0.151	0.234		8	
		cit	0.192	0.296		3	
	EPF	I-O	0.323	0.107	***	21	
		tech flow	0.263	0.093	***	6	
		tech prox	0.232	0.093	**	3	
Country	RPF	trade	0.045	0.015	***	8	
		fdi	0.047	0.015	***	1	
	EPF	pool	0.130	0.014	***	17	
		trade	0.048	0.013	***	150	
		tech prox	-0.015	0.028		11	
	KPF	geo	fdi	0.019	0.013		15
			fdi	0.050	0.014	***	20
		trade	geo	0.039	0.013	***	5
			geo	0.010	0.001	**	2
			cit	0.385	0.021	***	2
Intra-inter	EPF	pool	0.294	0.008	***	2	
		trade	0.116	0.063	*	39	
	KPF	fdi	0.521	0.066	***	2	
		pool	-0.030	0.053		4	
Inter-inter	RPF	geo	0.699	0.058	***	5	
		trade I-O	trade	0.181	0.028	***	9
			tech flow	0.660	0.033	***	1
	EPF	pool I-O	0.349	0.031	***	2	
		pool	0.130	0.026	***	2	
		trade I-O	0.178	0.008	***	25	
		trade tech flow	0.657	0.011	***	3	
		trade tech prox	0.648	0.014	***	3	
		trade pool	0.095	0.031	***	7	
		pool I-O	0.346	0.012	***	4	
		pool tech flow	0.345	0.009	***	3	

S.E reported in parenthesis are heteroscedasticity-consistent  
 \*\*\*,\*\*, \* denote statistical significance at the 1%, 5%, 10% level respectively

Table 3.10: Authentic spillovers effects

### 3.5 Conclusion

In this paper, we have analysed the contribution of knowledge spillovers to the productivity and to the innovative activity of agents at different level of aggregation (inter firm spillovers/intra sectoral spillovers, inter sectoral spillovers, international spillovers, intra sectoral-international spillovers and inter sectoral-international spillovers) by reviewing the main empirical findings in the literature. We have used a meta regression analysis to assess the authentic effect of knowledge spillovers on economic performance. The use of MRA allows, in addition of the identification of genuine or authentic effect, to highlight the presence of a strong publication effect that tends to affect positively the value of reported estimates through the literature. MRA also allows the identification of different factors affecting the reported estimates. The mains empirical findings can be summarized as follow: First, knowledge spillovers do exist and are strongly significant at all the levels considered and for all kinds of measures with one exception. When we correct the rates of return obtained by the intersectoral studies by the publication effects, the authentic effect becomes insignificant. The value of the effect presents considerable heterogeneity that is mainly due to the channel involved. If we focus on the most used channel the elasticity of knowledge spillovers on productivity lies between 0.207 to 0.276 for inter firm spillovers, 0.263 to 0.323 for inter sectoral spillovers, 0.048 to 0.050 for international spillovers, 0.116 for intra sectoral-international spillovers and 0.178 for inter sectoral-international spillovers. Second, we find a strong and significant publication effect for four of five MRA (the presence of publication effect is significant for all levels of analysis at the exception of intersectoral-international spillovers). For instance, the estimated value of simple international spillovers is two times more important when we don't take into account the presence of publication effect. Third, we find a significant increase in spillovers effect estimates over time, studies referring explicitly to data covering period after 1990 report systematically lower estimates.

### 3.6 Appendix

#### 3.6.1 Methodological issues

The easiest way to combine estimates in a meta analysis is to compute a fixed effect model. Fixed effect model assumes that differences across studies are only due to within-variation. Once fixed effect model is computed, we can implement common MA homogeneity test. A field of the literature show that high



heterogeneity can not be summarized by fixed effect, it should be preferable to use a random effect model. The random effect model considers both between and within study variability, and assumes that the studies are random samples from the universe of all possible studies (Sutton et al. 2000). Fixed effect models assume that differences across studies are only due to within-variation  $\epsilon_i$ . More specifically it assumes that all the studies are measuring the same thing, thereby a “true” effect  $\theta_F$  underlines every study:

$$\theta_i = \theta_F + \epsilon_i$$

Where  $\theta_i$  is the individual estimate of the spillovers effect and  $\epsilon_i$  is typically assumed normal with a mean 0 and a variance  $Se(\hat{\theta}_i)^2$ . Following Higgins and Thompson 2002, a optimal estimator of  $\theta_F$  is calculated as a weighted average of the study estimates:

$$\hat{\theta}_F = \frac{\sum \hat{\theta}_i w_i}{\sum w_i}$$

Where weights  $w_i$  are inversely proportional to the square of the standard errors:

$$w_i = \frac{1}{Se(\hat{\theta}_i)^2}$$

Alternatively, the meta analyst might conclude that the studies are heterogeneous implying that differences in the  $\theta_i$  come both from between and within study variability. Unlike the fixed effect model, the individual studies are not assumed to be estimating a true single effect size, rather the true effects in each study are assumed to have been sampled from a distribution of effects. More specifically the random effect model considers both between  $\varsigma_i$  and within  $\epsilon_i$  study variability, and assumes that the studies are random samples from the universe of all possible studies (Sutton et al. 2000). The random effects models assume that:

$$\theta_i = \theta_R + \epsilon_i$$

$$\theta_R = \theta_F + \varsigma_i$$

Where  $\varsigma_i$  is assumed to be normal with mean 0 and variance  $\tau^2$  and  $\varsigma_i$  &  $\epsilon_i$  are considered independent. An optimal estimator of  $\theta_R$  is calculated as a weighted average of the study estimates:

$$\hat{\theta}_R = \frac{\sum \hat{\theta}_i w_i^*}{\sum w_i^*}$$

Where the weights are equal to  $w_i^* = (Se(\hat{\theta}_i)^2 + \hat{\tau}^2)^{-1}$ .  $\hat{\tau}^2$  may be obtained by:

$$\hat{\tau}^2 = \frac{Q - n + 1}{\sum w_i - (\sum w_i^2 / \sum w_i)}$$

Allowing for the extra between-study variation has the effect of reducing the relative weighting given to the more precise value. Hence the random effect models produces a more conservative confidence interval for the pooled effect estimate. A test of homogeneity of the  $\theta_i$  is provided by referring the statistic:

$$Q = \sum w_i (\hat{\theta}_F - \hat{\theta}_i)^2$$

$Q$  follows a  $\chi^2$  distribution with  $n - 1$  degree of freedom. If  $Q$  exceeds the upper-tail value, the observed variance in estimated effect sizes is greater than that we would expect by chance if all studies shared the same true parameters. If there is no heterogeneity, then  $Q$  should be approximately equal to  $n - 1$ . The statistic  $H^2$  provide a possible measure of the amount of heterogeneity:

$$H^2 = \frac{Q}{n - 1}$$

In the absence of heterogeneity  $E[Q] = 1$  so that  $H^2 = 1$  indicates homogeneity in effect size. Another alternative is  $I^2$  (Higgins 2003), a statistic that measures the proportion of inconsistency in individual studies that cannot be explained by chance.

$$I^2 = \frac{H^2 - 1}{H^2} \times 100$$

$I^2$  is bounded above by 100% and values close to 100% represent very high degrees of heterogeneity. This measure is preferred to Cochran's  $Q$ . The problem with Cochran's  $Q$ , the authors claim, is that it tends to have too little power with a collection of studies with small sample sizes and too much power with a collection of studies with large sample sizes. In the following we assign adjectives of low, moderate, and high heterogeneity to values of  $I^2$  equal to 25%, 50%, and 75% respectively.

3.6.2 Tables

Authors	Database	Weights	Struct.	Model	Parameters		
Aiello & Cardamone 2005	Italy, 1017 firms	Pool I-O	panel	Y/L, R&DS/L	0.007 0.002	to	0.0167 0.0169
Aiello & Cardamone 2007	Italy, 557 firms	Pool Prox tech (firm charac) Distance	panel	Y, R&DS	0.3235 0.32	to	0.65
Aiello & Cardamone 2008	Italy, 1203 firms	Prox tech (firm charac) Distance	panel	Y, R&DS	0.1364 0.3177	to	0.3476
Anon Higon 2007	UK, 8 indus.	O-O Import + O-O	panel	$\Delta Y/K$ , $\Delta R&DS$	0.421 -0.162	to	2.553 0.009
Antonelli 1994	Italy, 92 firms	Pool	CS	$\Delta Y$ , R&D/Y	0.176	to	0.353
Barcenilla et al. 2008	6 countries, 10 indus	Import intra	panel	TFP, R&DS	0.327	to	1.216
Barrio-Castro et al. 2002	21 countries	Import	panel	TFP, R&DS	0.092	to	0.141
Bitzer & Geishecker 2006	17 countries, 7 indus 9 countries, 7 indus	Import Import intra Import Import intra	panel	Y, R&DS	0.0329 -0.0134 0.013 -0.018	to	0.0211
Bloom et al. 2005	US, 736 firms	Prox tech (pat) Prox tech (sales) Prox tech (pat) Prox tech (sales)	panel	Y, R&DS Pat, R&DS	-0.038 -0.008 0.192 0.024	to	0.111 0.009 0.403 0.049
Bonte 2004	West Germany, 25 indus	Input	panel	TFP, R&DS	0.217		
Bottazi & Peri 2003	Europe, 86 regions	Prox tech (pat) Distance	CS A	Pat, R&D	0.015 0.012	to	0.03 0.072
Braconier et al. 2001	Sweden, 84 firms	Pool intra	4 sub CS	Y/L, R&DS	-0.79	to	0.093
Branstetter 2000	Japan, 226 firms	Prox tech (pat)	panel CS	Pat, R&D $\Delta Y$ , $\Delta R&DS$	0.859 0.489		
Branstetter 2001	Japan, 205 firms & US, 209 firms	Prox tech (pat)	panel CS	Pat, R&D $\Delta Y$ , $\Delta R&DS$	0.797 0.503	to	1.108 0.829
Chinkov 2006	10 countries	Import	panel	TFP, PatS	0.655	to	1.921

Table 3.11: Papers included in the database

### Chapter 3. Knowledge Spillovers: A Meta-Analysis

Authors	Database	Weights	Struct.	Model	Parameters		
Cincera 2005	Us, 485 firms	Prox tech (R&D)	panel	Sales, R&DS	0.51		
Crespo et al. 2004	28 countries	Import	panel	$\Delta$ TFP, R&DS	-0.004	to	0.0035
Crespo et al. 2008	21 countries	Import	5 sub CS	$\Delta$ Y, $\Delta$ R&DS	0.014		
Edmond 2001	21 countries	Import Pool	panel	TFP, R&DS	-0.2357	to	0.4459
					0.149	to	0.4476
Engelbrecht 1997	21 countries	Import	panel	TFP, R&DS	0.198		
Engelbrecht 2002	61 countries	Import	4 sub CS	$\Delta$ Y, $\Delta$ R&DS	-1.829	to	0.49
Falvey et al. 2004	21 countries	Import Export	panel	$\Delta$ Y/L, $\Delta$ R&DS	0.031	to	0.076
					0.02	to	0.056
Fischer & Varga 2003	Austria, 72 regions, 6 indus	Distance intra	CS	Pat, R&D	0.1	to	0.402
Frantzen 1998	21 countries	Import	panel	TFP, R&DS	0.17	to	0.253
Frantzen 2000	21 countries	Import	CS	$\Delta$ TFP, R&D/Y	2.079	to	2.184
Frantzen 2002	14 countries, 22 indus	I-O Import intra Import+ I-O	panel	Y/L, R&DS	0.163	to	0.243
					0.045	to	0.053
					0.176	to	0.181
Fritsch & Franke 2004	Germany, 349 firms, 7 indus	Pool intra	panel	Patent, R&D	0.08	to	0.09
Goto & Suzuki 1989	Japan, 50 indus	I-O	CS	$\Delta$ TFP, R&D/Y	0.8		
Greunz 2001	153 European regions	Distance	panel	Pat/L, R&D/L	0.02		
Griliches & Lichtenberg 1984	Us, 193 sectors	Tech flow	2 sub CS	$\Delta$ TFP, R&D/Y	0.289	to	0.904
Hanel 2000	Canada, 22 sectors	Tech flows Import FDI Import +Tech flow	3 sub CS	$\Delta$ TFP, R&D/Y	1.537	to	1.904
					0.126	to	
					0.167	to	
					0.176	to	
Hanel & St Pierre 2002	Canada, 278 firms	Prox tech (pat) intra Tech flow	CS	Profit/Y, R&DS/Y	-0.015	to	-0.009
					0.0009	to	0.046

Table 3.12: Papers included in the database (continued)

### Chapter 3. Knowledge Spillovers: A Meta-Analysis

Authors	Database	Weights	Struct.	Model	Parameters		
Harhoff 2000	Germany, 439 firms	Prox tech (R&D)	2 sub CS	Y, R&DS	-0.327	to	0.037
Hejazi & Safarian 1999	20 countries	Import FDI	panel	TFP, R&DS	0.003 0.080	to	0.106 0.136
Jacob & Mesiter 2005	Indonesia, 19 sector	Export + I-O	panel	Y/L, R&DS/Y	-0.160	to	0.253
Jaffe 1986	US, 432 firms	Prox tech (pat)	CS	Pat, R&D	0.509	to	0.628
Jaffe 1989	US, 500 firms	Prox tech (pat)	CS	Pat, R&D Y, R&DS	0.104 0.038		
Kao et al. 1999	22 countries	Import	panel	TFP, R&DS	0.289	to	0.060
Keller 1997	8 countries, 13 indus	Foreign pool Foreign I-O Foreign tech flow Foreign random	panel	TFP, R&DS	0.385 0.373 0.359 0.191		
Keller 1998	22 countries	Import	panel	TFP, R&DS	0.060	to	0.329
Kuo & Yang 2008	China, 31 regions	Distance	panel	Y, R&DS	0.000	to	0.152
Kwark & Shyn 2006	103 countries	Export	5 sub CS	$\Delta$ TFP, $\Delta$ R&DS	0.007	to	0.197
Lai et al. 2006	China, 30 regions	FDI Import	panel	$\Delta$ Y, R&D	0.107 0.012	to	0.241 0.072
Lejour & Nahuis 2005	4 world regions, 6 indus	I-O Import + I-O	panel	TFP, R&DS	1.675 0.126	to	3.430 0.617
Lichtenberg & Van Pottelsberghe 1998	22 countries	Import	panel	TFP, R&DS	0.004	to	0.310
Link 1983	US, 302 firms	Input	CS	$\Delta$ TFP, R&D/Y	0.247		
Liu & Buck 2007	China, 21 indus	Foreign pool intra	panel	NP/Y, R&D/Y	0.054	to	0.086
Lopez-Pueyo et al. 2008	6 countries, 10 indus	I-O Import intra Import + I-O	panel	TFP, R&DS	0.014 -0.046 -0.001	to	0.068 0.119 0.272
Los & Verspagen 2000	US, 485 firms	Pool Tech flow	CS	Y/L, R&DS	0.392 0.005	to	0.623 0.483

Table 3.13: Papers included in the database (continued)

### Chapter 3. Knowledge Spillovers: A Meta-Analysis

Authors	Database	Weights	Struct.	Model	Parameters		
Lumenga-Nesoa et al. 2005	22 countries	Import Pool	panel	TFP, R&DS	0.055 0.053	to	0.080 0.054
Madsen 2007	13 countries	Import	panel	TFP, PatS	0.137		
Muller & Nettekoven 1999	22 countries	Import	panel	TFP, R&DS	-0.199	to	0.233
Park 2004 a	21 countries	Import Pool	panel	TFP, R&DS	0.006 0.129	to	0.180 0.175
Park 2004 b	17 countries, 2 indus	Import intra	panel	$\Delta$ TFP $\Delta$ R&DS	-0.002	to	-0.096
Park 1995	10 countries 3 countries	Prox tech (R&D)	panel	$\Delta$ Y/L, $\Delta$ R&DS/L	0.170 0.095	to	0.180 0.180
Peri 2005	113 regions	Pat cit	panel	Pat, R&DS	0.400	to	0.540
Scherer 1982a	Us, 87, 51 & 37 indus	Tech flow product	CS	$\Delta$ Y/L, $\Delta$ R&DS/Y	0.874	to	1.472
Scherer 1982b	Us, 87 indus	Tech flow product	CS	$\Delta$ Y/L, R&D	0.096	to	1.039
Scherer 2002	Us, 49 indus	Tech flow product	CS	$\Delta$ Y/L, R&D/Y	0.698	to	0.742
Todo 2006	Japan, 21404 firms	Pool intra dom	CS	TFP, R&DS	0.247	to	0.343
Van Pottelsberghe 1997	7 countries, 22 indus	I-O Tech flow	panel	$\Delta$ TFP, $\Delta$ R&DS/Y	0.204 2.123	to	1.039 2.804
Verspagen 1997	14 countries, 22 indus	Tech flow	CS	Y/L, R&DS	0.056		
Vuori 1997	Finnish, 20 indus	Prox Tech (R&D)	3 sub CS	$\Delta$ TFP, R&D/Y	0.240	to	0.530
Wang 2007	25 countries, 16 indus	Import	panel	TFP, R&DS	0.297	to	0.357
Wei & Liu 2006	China, 7697 firms	Pool regional	panel	VA, R&DS	0.038	to	0.998
Wolf & Nadiri 1993	US, 19 indus	Input/Invest	5 sub CS	$\Delta$ TFP, R&D/Y	0.076	to	0.143
Xu & Wang 1999	21 countries	Import	panel	TFP, R&DS	-0.004	to	0.247
Yao 2006	US, 1365 firms	Pool	panel	Pat, R&D Pat, Pat	-1.806 0.760	to	-0.090 2.119
Zhu & Jeon 2007	22 countries	Pool Import FDI	panel	TFP, R&DS	0.132 0.048 0.026	to	0.175 0.101 0.110

Table 3.14: Papers included in the database (continued)

### Chapter 3. Knowledge Spillovers: A Meta-Analysis

Moderator variables	
Channel	
<b>Pool</b>	Take the value 1 if the spillovers variable is a no weight sum of external knowledge
<b>I-O</b>	Take the value 1 if the channel associated with spillovers is based on national economic transactions
<b>Tech flow</b>	Take the value 1 if the channel associated with spillovers is based on technology flows
<b>Tech prox</b>	Take the value 1 if the channel associated with spillovers is based on technology proximity
<b>Cit</b>	Take the value 1 if the channel associated with spillovers is based on patent citation
<b>FDI</b>	Take the value 1 if the channel associated with spillovers is based on foreign direct investment flows
<b>Geo</b>	Take the value 1 if the channel associated with spillovers is based on geographic distance
<b>Trade</b>	Take the value 1 if the channel associated with spillovers is based on international economic transaction
<b>Tech prox+Geo</b>	Take the value 1 if the channel associated with spillovers is based both on technologic proximity and geographic distance
<b>Pool+I-O</b>	Take the value 1 if the channel associated with spillovers is based on pool scheme and economic transaction for the international and national dimension resp.
<b>Pool+Tech flow</b>	Take the value 1 if the channel associated with spillovers is based on pool scheme and technology flow for the international and national dimension resp.
<b>Trade+I-O</b>	Take the value 1 if the channel associated with spillovers is based on economic transaction for both the international and national dimension
<b>Trade+Pool</b>	Take the value 1 if the channel associated with spillovers is based on international transaction and pool scheme for the international and national dimension resp.
<b>Trade+Tech flow</b>	Take the value 1 if the channel associated with spillovers is based on international transaction and technology flows for the international and national dimension resp.
<b>Trade+Tech prox</b>	Take the value 1 if the channel associated with spillovers is based on international transaction and technologic proximity for the international and national dimension resp.
Others	
<b>Nbr channel</b>	Take the value 1 if the estimates are estimated with a specification including more than one spillovers variable into a same regression
<b>Cross section</b>	Take the value 1 if the estimation are made with cross sectional data
<b>Growth</b>	Take the value 1 if the estimation are made in growth rate or lag difference
<b>Complex Reg</b>	Take the value 1 if the estimation use complex estimators
<b>Open</b>	Take the value 1 if the spillovers variable includes some measure of the agent openness
<b>Absorption</b>	Take the value 1 if the spillovers variable includes some measure of the absorption capacity of the agent
<b>1989</b>	Take the value 1 if the estimate refers explicitly to data before 1990

Table 3.15: Definition of the moderators used in the MRA

Rate of return

<b>Inter-Firm</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.132	0.087	0.177	0.208	2	0.093	2.81	64.45%
	random-effects	0.132	0.087	0.177	0.208				
Median	fixed-effects	0.229	0.158	0.300	0.191	2	0.663	0.19	-427.86%
	random-effects	0.229	0.158	0.300	0.191				
Max	fixed-effects	0.300	0.242	0.358	0.121	2	0.359	0.84	-18.77%
	random-effects	0.300	0.242	0.358	0.121				

<b>Inter-sectoral</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	-0.005	-0.020	0.011	0.778	8	0.065	1.90	47.38%
	random-effects	-0.005	-0.020	0.011	0.780				
Median	fixed-effects	0.263	0.217	0.310	0.001	8	0.000	3.73	73.21%
	random-effects	0.292	0.239	0.345	0.001				
Max	fixed-effects	0.307	0.224	0.390	0.008	8	0.000	4.63	78.39%
	random-effects	0.629	0.478	0.781	0.004				

<b>International</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.035	0.021	0.049	0.135	3	0.036	3.34	70.04%
	random-effects	0.035	0.021	0.049	0.135				
Median	fixed-effects	0.039	0.032	0.047	0.014	3	0.104	2.05	51.24%
	random-effects	0.039	0.032	0.047	0.014				
Max	fixed-effects	0.039	0.029	0.048	0.058	3	0.016	4.17	75.99%
	random-effects	0.039	0.029	0.049	0.058				

<b>Inter-sectoral International</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.003	-0.060	0.066	0.969	3	0.029	3.54	71.77%
	random-effects	0.003	-0.062	0.068	0.967				
Median	fixed-effects	0.077	0.029	0.124	0.354	3	0.199	1.65	39.44%
	random-effects	0.077	0.029	0.124	0.354				
Max	fixed-effects	0.239	0.194	0.284	0.034	3	0.029	3.53	71.71%
	random-effects	0.238	0.193	0.284	0.035				

Table 3.16: Sensitivity of meta-estimates to the choice of “preferred” estimates



## Elasticity estimated with a production function

<b>Inter-Firm</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.009	0.009	0.010	0.000	14	0.000	248.83	99.60%
	random-effects	0.015	0.014	0.016	0.000				
Median	fixed-effects	0.106	0.103	0.109	0.000	14	0.000	298.14	99.67%
	random-effects	0.184	0.173	0.196	0.000				
Max	fixed-effects	0.212	0.209	0.215	0.000	14	0.000	585.91	99.83%
	random-effects	0.285	0.260	0.309	0.000				

<b>Inter-sectoral</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.045	0.034	0.056	0.007	7	0.000	4.17	76.02%
	random-effects	0.045	0.034	0.056	0.007				
Median	fixed-effects	0.128	0.120	0.136	0.000	7	0.000	9.34	89.30%
	random-effects	0.127	0.119	0.136	0.000				
Max	fixed-effects	0.156	0.148	0.165	0.000	7	0.000	14.82	93.25%
	random-effects	0.156	0.147	0.165	0.000				

<b>International</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.001	0.000	0.002	0.305	23	0.000	12.73	92.15%
	random-effects	0.001	0.000	0.003	0.268				
Median	fixed-effects	0.003	0.003	0.004	0.000	23	0.000	30.05	96.67%
	random-effects	0.003	0.003	0.004	0.000				
Max	fixed-effects	0.007	0.006	0.008	0.000	23	0.000	53.33	98.13%
	random-effects	0.009	0.009	0.010	0.000				

<b>Intra-sectoral International</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.003	0.001	0.005	0.213	9	0.000	123.05	99.19%
	random-effects	0.015	0.012	0.018	0.001				
Median	fixed-effects	0.021	0.018	0.023	0.000	9	0.000	122.03	99.18%
	random-effects	0.059	0.055	0.063	0.000				
Max	fixed-effects	0.012	0.010	0.015	0.001	9	0.000	168.64	99.41%
	random-effects	0.119	0.112	0.126	0.000				

<b>Inter-sectoral International</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.051	0.045	0.056	0.000	8	0.000	24.09	95.85%
	random-effects	0.053	0.046	0.059	0.000				
Median	fixed-effects	0.016	0.013	0.019	0.000	8	0.000	26.39	96.21%
	random-effects	0.018	0.015	0.021	0.000				
Max	fixed-effects	0.076	0.072	0.081	0.000	8	0.000	27.28	96.33%
	random-effects	0.087	0.082	0.092	0.000				

Table 3.17: Sensitivity of meta-estimates to the choice of “preferred” estimates (continued)

Elasticity estimated with a knowledge production function

<b>Inter-Firm</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.056	0.043	0.069	0.009	5	0.000	8.90	88.76%
	random-effects	0.058	0.044	0.072	0.009				
Median	fixed-effects	0.210	0.190	0.229	0.000	5	0.000	6.52	84.66%
	random-effects	0.210	0.190	0.230	0.000				
Max	fixed-effects	0.331	0.303	0.358	0.000	5	0.000	10.51	90.48%
	random-effects	0.332	0.299	0.365	0.001				

<b>International</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.011	0.009	0.013	0.037	3	0.000	33.64	97.03%
	random-effects	0.011	0.009	0.013	0.037				
Median	fixed-effects	0.022	0.019	0.024	0.010	3	0.000	89.88	98.89%
	random-effects	0.022	0.020	0.024	0.011				
Max	fixed-effects	0.022	0.019	0.024	0.010	3	0.000	86.41	98.84%
	random-effects	0.022	0.019	0.024	0.011				

<b>Intra-sectoral International</b>		Pooled estimates	95% confidence interval		Ho: no effect	Obs	Q-Test (p-value)	H2	I2
Min	fixed-effects	0.079	0.052	0.106	0.211	2	0.399	0.71	-40.31%
	random-effects	0.079	0.052	0.106	0.211				
Median	fixed-effects	0.090	0.068	0.112	0.155	2	0.175	1.84	45.76%
	random-effects	0.090	0.068	0.112	0.155				
Max	fixed-effects	0.300	0.255	0.344	0.094	2	0.001	11.10	90.99%
	random-effects	0.295	0.248	0.342	0.100				

Table 3.18: Sensitivity of meta-estimates to the choice of “preferred” estimates (continued)



## Chapter 4

# Patent Citation Data and Knowledge Flows

### 4.1 Introduction

Knowledge is presented as a key factor of economic growth, development and the well-being of nations. Our understanding of what is happening in a knowledge-based economy is constrained by the scope and the quality of the available knowledge-related indicators. Accordingly, great interest has been placed in measuring the amount of knowledge. Traditional measures of knowledge were based on R&D indicators which are excellent indicators of resource invested in innovative activities but say nothing about the output and the diffusion process of innovation. In recent years, following the guideline of the OECD (1996) [131], several new indicators were built (intangible investment, patents, bibliometrics<sup>1</sup>, measures of innovation though direct survey...). One of them, patents, was used extensively in many works due to recent improvements in the quality of patent data (Kleinknecht et al. 2002 [100]). Since the seminal work of Griliches (1990) [65], focus was placed on the adequacy of patent data in providing output-based measures of the knowledge formation. But few studies have considered possibilities offered by patents in the creation of an input-based indicator. Such indicator would be very useful to compensate lack of R&D data (R&D data are often not available for specific technologies and/or for long periods of time) because the great advantage in using a patent database lies in the large volume of information available, related to inventors, technological fields, citations, etc, covering a

---

<sup>1</sup>Bibliometrics is a set of methods used to study or measure texts and information. Historically bibliometric methods have been used to trace relationships amongst academic journal citations. Citation analysis, which involves examining an item's referring documents, is used in searching for materials and analyzing their merit.

long period of time, on an international level. Thus, patents are one of the richest sources of data on inventions and certainly the one with the widest coverage providing an unique insight into the processes and outcomes of inventive activities. Specifically, patent data allows the analysis of the knowledge diffusion process. It is well known that knowledge has the characteristics of a public good, flowing across agents affecting positively their productivities, in other words knowledge spill-over. From a theoretical point of view, Romer in his seminal 1990 [150] article has highlighted the role of spillovers as a source of growth by allowing an escape to the fate of diminishing returns. From an empirical point of view, literature provides evidence that significant spillovers exist and take place (see chapter 2 and 3). Understanding knowledge diffusion is therefore fundamental in the study of the knowledge-based economies, unfortunately “knowledge flows [...] are invisible; they leave no paper trail by which they may be measured and tracked” Krugman 1992 [101]. But thanks to patent data, knowledge flows do sometimes leave a paper trail, in the form of patent citations, Jaffe et al. (1993) [89]. Nevertheless, the extent to which citations are a good indicator of knowledge flows is still controversial especially about citations set up by examiners.

The first part of this chapter is dedicated to the definition and specification of patents and how we can use them for different purposes into the measurement of economic knowledge based values. An extensive literature has been made in this direction, but it focused almost exclusively on US patent data. Nowadays, with the introduction of new database with largest offices coverage, new light needs to be shed especially on patent and patent citation data that do not arise from the USPTO. By a detailed review of the technical issues related to patents and patent procedures for the six major patent offices in the world, we will discuss the opportunities offered by patent data as a measure of the creation of new ideas and knowledge flows.

This review allows us to lay the groundwork for the development of a new input based knowledge indicator that takes into account international knowledge spillovers and to focus on the legitimacy of using patent citations as a measure of technology flows. By analyzing patent citations and their characteristics through the “patent citation function”<sup>2</sup> we will be able to identify the productivity of knowledge embodied in specific patent cohorts<sup>3</sup> and the usefulness of this knowledge for subsequent inventors. This point is linked on the accumulative property of knowledge: “In the process of innovation,

---

<sup>2</sup>the patent citation function is a structural model developed for the analysis of patent citations’ patterns, see Caballero and Jaffe (1993) [25] and Jaffe and Trajtenberg (1996, 1998) [84, 85]

<sup>3</sup>i.e all patents applied (or granted, depending on what we want to measure) during the related year with specific characteristics like geographic origin, technological field...

inventors rely and build on the insights embodied in previous ideas represented by the public stock of knowledge that accumulates from the spillovers of previous inventions.” Schmookler (1966) [163]. We follow Popp (2002) [146] in using the productivity parameters estimated by the citation function to build an input-based indicator or in more common words, a stock of knowledge. The knowledge stock thus developed is an international stock of patents weighted by their subjective productivities (depending also on the inventor’s characteristics who could take advantage of this knowledge). In this way, we expand the previous methodology on an international level taking into account international knowledge flows. Another advantage, using the citation function lies in the direct estimation of the diffusion lag and the obsolescence rate of knowledge that is one of the gaps highlighted by the literature:

“As regards the formalization of R&D spillovers, several issues should be addressed in order to improve the measure of these effects. For instance, the assumed rate of obsolescence for computing the spillover stock should actually be much less than the corresponding one used for constructing the firm’s own R&D capital to the extent that the public R&D stock, given its public nature, can be expected to depreciate less rapidly.” Cincera & Pottelsberghe (2001) [28].

We will begin this chapter by defining and detailing what is a “patent” (section 4.2.1) and how it operates (section 4.2.2). Emphasis is placed on the differences between the six major patent offices in the world. We review the patent procedures from the DPMA, EPO, INPI, IPO, JPO, and USPTO (respectively the German, the European, the French, the Great Britain, the Japan and the US patent office). Secondly, we investigate how patent data can be used as an innovative based indicator (section 4.2.3) and how patent citations can be a good measure of knowledge flows (section 4.2.4). The third point will be devoted to the presentation of the patent citation function used in the analysis of patent citations. This presentation comes back on the development of this function (section 4.3) and presents its technical properties (section 4.4). Finally, the last point will detail the construction of our indicator, i.e. a stock of knowledge based on international patent and patent citation data that takes into account international knowledge spillovers (section 6.2).

## 4.2 Patent and Patent Citation as indicator of knowledge and knowledge flows

A patent for an invention is the grant of a property right to the inventor excluding other from making or using this invention. This exclusive right is given by a state or regional office in a limited period of

time in exchange for a public disclosure of the invention. Information embodied in patent is very useful for economic analysis of technological change in that it constitutes the richest database on innovative activity. The first part of this section reviews and describes the different aspects of patent in term of the legal framework (4.2.1) and administrative process for obtaining it in major offices (4.2.2). The second part investigates the property of patent in term of economic indicator of innovative activity (4.2.3). Finally, the last part of this section will conclude by introducing evidences that patent citation is an accurate indicator of knowledge flows (4.2.4).

### 4.2.1 What is a patent ?

A patent is a legal title protecting an invention (Article 28 of the Trade-Related Intellectual Property Rights (TRIPS) Agreement<sup>4</sup>):

1. “A patent shall confer on its owner the following exclusive rights:
  - (a) where the subject matter of a patent is a product, to prevent third parties not having the owner’s consent from the acts of: making, using, offering for sale, selling, or importing for these purposes that product;
  - (b) where the subject matter of a patent is a process, to prevent third parties not having the owner’s consent from the act of using the process, and from the acts of: using, offering for sale, selling, or importing for these purposes at least the product obtained directly by that process.”
2. “Patent owners shall also have the right to assign, or transfer by succession, the patent and to conclude licensing contracts”.

An invention can belong to any field of technology, with few specified exception. Furthermore, are considered as patentable:

- discoveries, scientific theories and mathematical methods
- aesthetic creations
- schemes, rules and methods for performing mental acts, playing games or doing business

---

<sup>4</sup>National patent laws have to comply with international standards, now laid down in the TRIPS (Trade-related Aspects of Intellectual Property Rights), an international treaty which is part of the WTO (World Trade Organization) package signed in 1994.

- programs for computers,
- presentations of information

Are not considered to be inventions if the patent application only relates to such subject-matter or activities as such. In addition to this, inventions falling into any of the following categories are excluded from patentability: inventions whose commercial exploitation would be contrary to "ordre public" or morality; plants and animal varieties or essentially biological processes for producing plants or animals (Microbiological processes and products thereof are not excluded); methods for treatment of the human or animal body by surgery or therapy and diagnostic methods practiced on the human or animal body. (Products, substances and compositions for use in such methods, e.g. medicament's or surgical instruments, are not excluded).

Thereby, a patent is the legal right of an inventor to exclude others from making or using a particular invention. This right is customary limited in time, to 20 years from the date of the application in most countries. The principle behind the modern patent is that an inventor is allowed a limited amount of time to exclude others from supplying or using an invention in order to encourage inventive activity by preventing immediate imitation. In return, the inventor is required to make the description and implementation of the invention public rather than keeping it secret, allowing others to create more easily on the knowledge contained in his invention. Although the process for granting a patent varies slightly according to the jurisdiction for which the protection is desired it is approximately the same everywhere in the world. To be eligible for a patent, an invention (device, process, etc.) needs to meet certain standards. Invention has

- to be new,
- to mark an inventive step,
- to be commercially applicable.

Patents are obtained after following specific administrative procedures. In order to obtain a patent, the applicant (a person, either legal or natural, who has the "ownership" or control of the invention) has to file an application at a patent office with jurisdiction to grant a patent in the geographic area over which coverage is required. The office checks whether the invention fulfils the relevant legal criteria, and grants or rejects it accordingly. There are different alternative "routes" for protection available to



applicant, this will often be the national route but could be also an international route<sup>5</sup> or a regional route<sup>6</sup>. The three main patent offices are the European Patent Office (EPO), the Japan Patent Office (JPO) and the United States Patent and Trademark Office (USPTO). The largest national patent offices at the European level are the National Industrial Property Institute (INPI-France), the German Patent and Trade Mark Office (DPMA-Germany) and the Intellectual Property Office (IPO-United Kingdom).

Table 4.1 shows frequencies of the routes selected by applicants. The first line represents the percentage of inventors who first file their inventions in their national offices<sup>7</sup>. For instance, 94% of French inventors first file their inventions at the French patent office. All rates are very high, especially for German, French and US inventors whose inventions are almost always filed in the national office. This is also true for the UK but at a lower level. Concerning the JPO, we have to take care about these results because of the low quality of JPO data concerning the inventor address. Only 1.5 % of JPO publications present information about the country code of the inventor. Since it is not the case for the other selected offices, the ratio of inventors who first filed at their national office is under estimated for Japanese inventors. The second line presents the percentage of first patent applications which are made at the national office by domestic inventors. For instance, 98% of first patent application carried out at the JPO was made by Japan inventors. Once again rates are very high. For the national European countries, ratios are very closed the previous ones. In contrast, the ratio of JP and US are slightly different, higher for Japan and lower for the US. For US inventors 98 % first file to their national offices, but 12 % of first patent applications made at the USPTO were carried out by foreign inventors.

---

<sup>5</sup>Any resident or national of a contracting state of the PCT may file an international application under the PCT that specifies the office which should conduct the search. The PCT application serves as an application filed in each designated contracting state. However, in order to obtain patent protection in a particular state, a patent needs to be granted by that state to the claimed invention contained in the international application. The advantage of a PCT application is that fewer searches need to be conducted and the process is therefore less expensive.

<sup>6</sup>A regional patent application is one which may have effect in a range of countries. The European Patent Office (EPO) is an example of a Regional patent office. The EPO grants patents which can take effect in some or all countries contracting to the European Patent Convention (EPC), following a single application process. Filing and prosecuting an application at a regional granting office is advantageous as it allows patents in a number of countries to be obtained without having to prosecute applications in all of those countries. The cost and complexity of obtaining protection is therefore reduced. A similar situation exists with respect to the African Regional Intellectual Property Organization (ARIPO).

<sup>7</sup>We use the EPO Worldwide Patent Statistical Database (called Patstat). PATSTAT was developed by patent information experts at the EPO's Vienna sub-office, and includes patent data from 73 offices world-wide and post-grant data from about 40 offices. It was developed specifically with the needs of policymakers, academics, analysts and IP institutions in mind. Researchers working in this field have previously had to assemble data sets from various and disparate sources and were obliged to perform extensive "cleaning" of the data at considerable cost and time. The PATSTAT dataset addresses these issues, efficiently harmonizing data, resolving issues over family members and addressing such problems as applications from one applicant appearing under several different names. The database also contains related information on citations, procedural information and legal status, which are all of interest to statisticians. We actually use the September 2008 version that included information on about 75 millions of entry.

Office	DPMA	INPI	IPO	JPO	USPTO
Ratio of the inventor who first filed at their national offices	95%	95%	82%	45%	98%
Ratio of the first filing made by domestic inventor	94%	92%	79%	98%	88%

---

Source: Own calculation from national database (Patstat)

Table 4.1: First filing in major patent offices (1978-2006)

Examples of patent application are given in Appendix (first page of the patents and the search report for the first two patents). Patent P1 (figure 4.7) is a first application made by French inventor to the French patent office, second application of P1 was done at the EPO and is referenced to P21 (figure 4.9). Actually patents P1 and P21 cover the same invention but concern different geographic coverage, although the INPI patent concerns exclusively France the EPO patent extends the applicant right to almost all EPO members. Patents P2 and P22 are others examples of first and second patent application (resp figure 4.11 and 4.12) at the difference that first application was made at the JPO by Japan inventors.

Figure 4.1 & 4.2 present trends of first patent filing at selected patent offices. The most notable increase concerns the USPTO, the JPO and the EPO where first applications have more than doubled during the period. Germany also presents an increase trend but only since the late 1980s. France and the United Kingdom present a steady pattern over the period, at the exception of a shift upward for the UK in the early 1980s.

### 4.2.2 Patent Procedures

In order to obtain a patent right, applicant must apply to a patent office and go through the examinations process to determine whether the application fulfils all the necessary requirements. After a first application (a prior application), applicant has a priority period of 12 months to file a subsequent application in another country for the same invention. Once the patent specification complies with the laws of the office concerned, a patent may be granted for the invention described and claimed by the specification. The first step in the granting procedure is the examination on filing. This involves checking whether all the necessary information and documentation has been provided, therefore a filing date can be accorded to the application (this is also the application date). If this application is the first application filed worldwide (in any patent office) it is known as the priority application, to which is

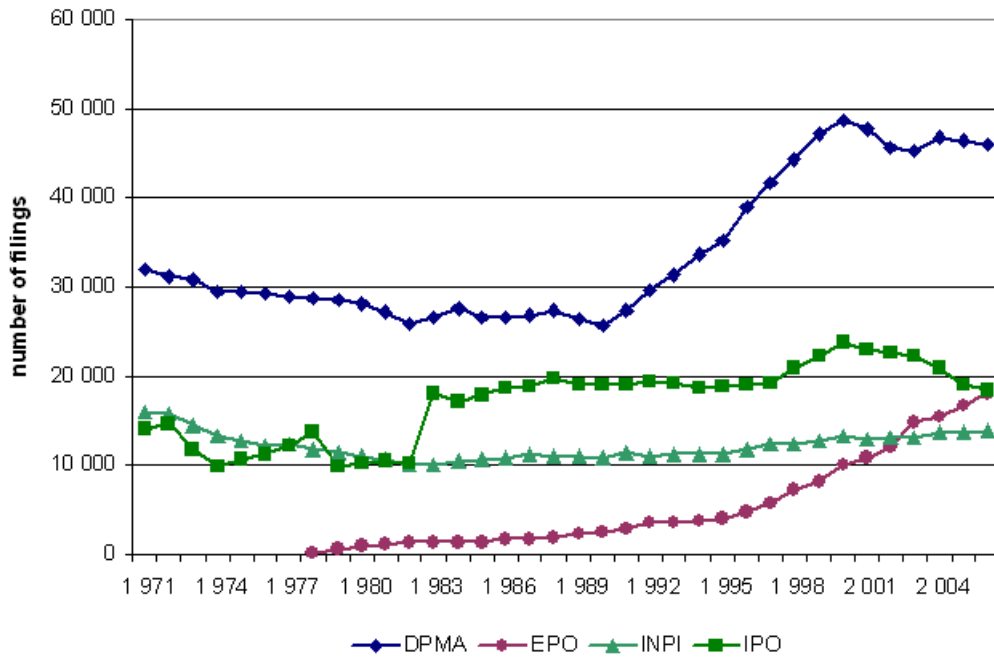


Figure 4.1: Number of first patent filing at selected patent offices, 1971-2006

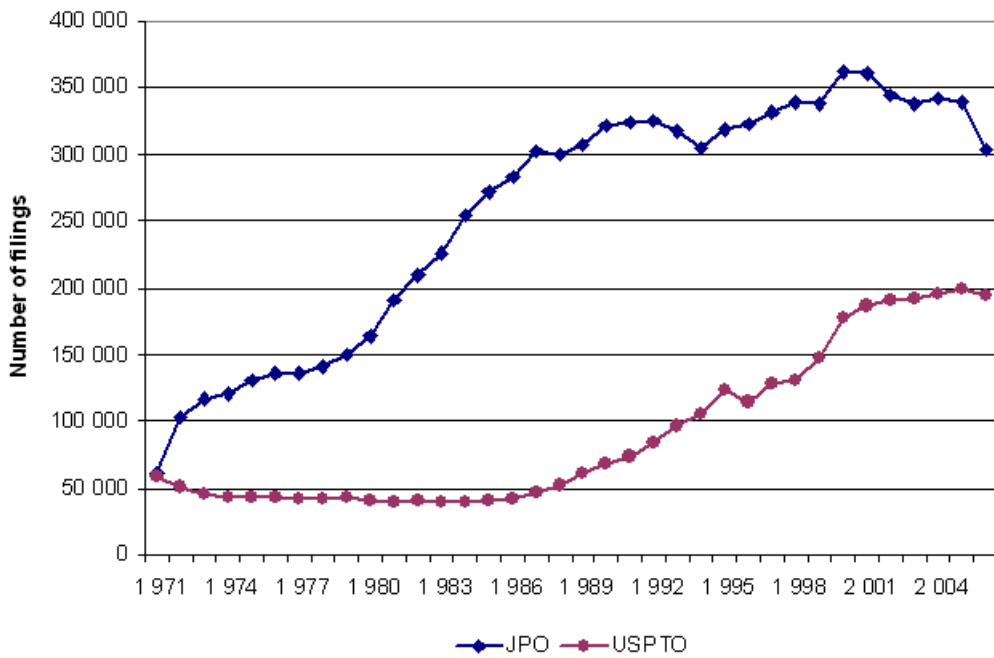


Figure 4.2: Number of first patent filing at selected patent offices, 1971-2006

associated a priority date. The patent office then drawn up the search report, listing all the documents available to the Office that may be relevant to assess novelty and inventive step. The search report is based on the patent claims but also takes into account the description and any drawings.

The application is generally published 18 months after it is filed (publication date). The lag between filing and grant or refusal is not fixed; it ranges from two to eight years, with significant differences across patent offices. Delay at which the authorized offices confer the patent rights to the applicant (the grant date) may occur on average after three years at the USPTO, five years at the EPO, but can take up to ten years in some cases. The EPO, the JPO and the USPTO differ slightly in the patent procedures; European national offices differ also but are comparable to one of the previous three offices. We will descriptive differences and practices into the three major offices afterwards we compare their procedures with the three major European national patent offices (France, Germany and United Kingdom). Summarize of procedures for the EPO, JPO and USPTO are provided in figure 4.3 and 4.4.

### 4.2.2.1 EPO.

A European patent application can arise from: direct filing to the EPO without a priority claim (first filing), extension of an earlier national patent application (within 12 months of first filing), or from an international application filed using the PCT procedure.

1. First step is common in all offices, it's the formality examination. It is just an administrative examination for compliance with formal requirements so that the applicant can be accorded a filing date.
2. In a second step a search is done in order to establish the state of the art with respect to the invention. The applicant receives the search report with an initial opinion on patentability.
3. The application is published with the search report 18 months after the date of filing.
4. Once the search report has been published, the applicant has six months to request a substantive examination. Once the request is filed, the EPO examines whether the invention is patentable (inventive steps and industrial applicability). Any application for which a request for substantive examination has not been filed is deemed to be withdrawn.
5. After the "Announcement to grant" the patent, the patent specifications are published. If the application cannot be granted, the intention to reject the application is communicated in the

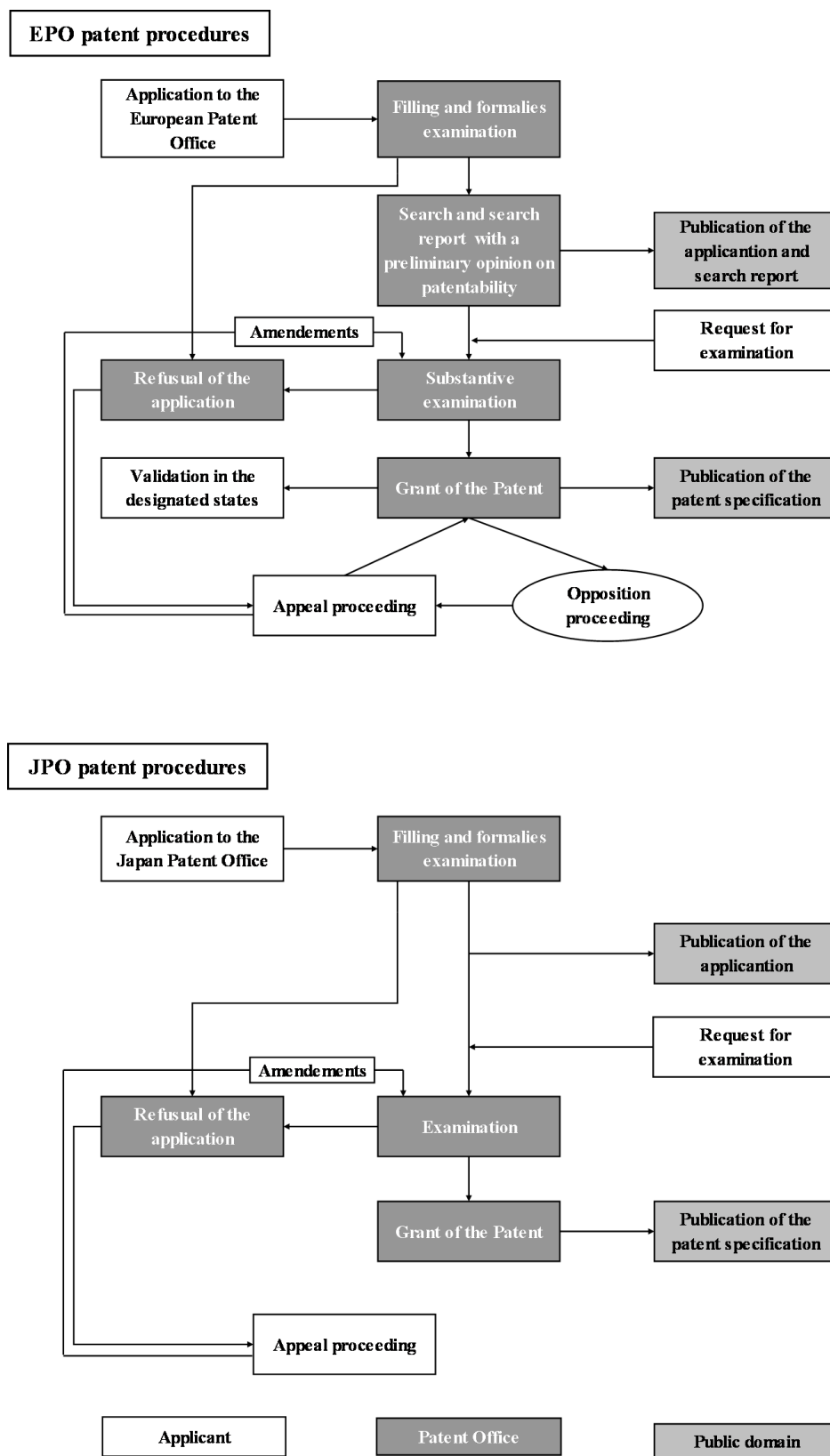


Figure 4.3: EPO and JPO patent procedures

“examination report”. The applicant may then make amendments to the application, generally in the claims, after which examination is resumed. This procedural step is iterated as long as the applicant continues to make appropriate amendments. Then, either the patent is granted or the application is finally rejected or withdrawn by the applicant.

6. The granted European patent is a “bundle” of individual national patent. In most contracting states, for the patent to retain its protective effect and be enforceable against infringement, it must be validated. The patent owner has to file a translation of the specification into an official language of that state with the national patent office.

Decisions of the EPO concerning issues such as the refusal of an application are open to appeal. An opposition to patent granted by the EPO can be filed by a third part within a period of nine months. The patent holder and the opponent can then debate. Finally, the EPO will make a decision based on facts and arguments presented by both sides. Such procedure does not exist in the two other patent offices (i.e. JPO and USPTO). According to Reitzig and Harhoff (2004) [74], the rate of opposition at the EPO is about 6.5%, one third of the oppositions are regarded as unfounded by the EPO, one third leads to the removal of the patent and one third to amendment of the patent.

### 4.2.2.2 JPO

1. First step is similar to the EPO and concerns the formality examination allowing applicant to get a filing date.
2. At the opposite of the EPO, the search and the substantive examination are undertaken in one phase.
3. The JPO will publish the content of an “unexamined application” after 18 months have elapsed from the date of filing.
4. As for the EPO, the “formality examination” is done only on the requirement of the applicant. At the JPO, the applicant has three years to request for examination (this delay was reduced from seven to three years in 2001). It is during this examination that the examiner discloses the prior art with respect to the invention.
5. After the “Decision to Grant” the patent, the patent specifications are published. If the application cannot be granted, the intention to reject the application is communicated in the “Noti-

fication of reason for refusal”. Like to the EPO applicant can made amendments and begin the same iteration process.

Decisions of the JPO concerning issues such as the refusal of an application are open to appeal.

### 4.2.2.3 USPTO

1. First step is similar to the EPO and concerns the formality examination allowing applicant to get a filing date.
2. At the opposite of the EPO, the search and the substantive examination are undertaken in one phase.
3. Until the passage of the American Inventors Protection Act of 1999 (P.L. 106-113), the United States was unusual in allowing patent application to remain secret until it was granted. The act includes a requirement that U.S. patent application is published 18 months after the USPTO first filing, unless the inventor signs an agreement to forgo patent protection in other countries.
4. Unlike to the EPO and the JPO, filing an application to the USPTO is taken to imply a request for examination, so that applicant doesn't have to request an examination since its application is filing.
5. After the “Notice of allowance” (decision to grant the patent), the patent specifications are published. If the application cannot be granted the intention to reject the application is communicated in the “Office action of rejection”. Like to the EPO, applicant can make amendments and begin the same iteration process.

Decisions of the USPTO concerning issues such as the refusal of an application are open to appeal. The USPTO has a re-examination procedure, which is slightly different to the EPO opposition procedure. Anyone can present reasons and evidences to the USPTO to challenge the validity of a granted patent. However, only the patent holder can be engaged in discussion with the USPTO examiner to establish the validity of the reasons. The challenger is not a part of these proceedings. Lanjouw & Schankerman (2001) [104] estimate the incidence of validity challenges to be at 1% of all patent grants.

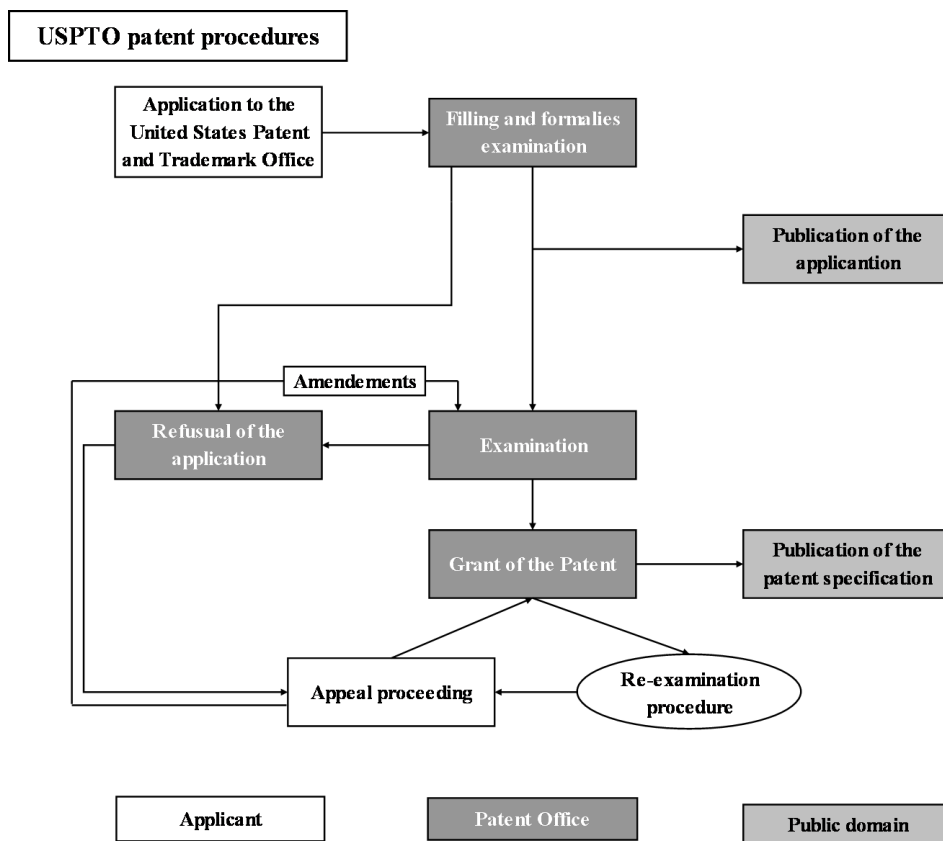


Figure 4.4: USPTO patent procedure



### 4.2.2.4 National European patent offices

The German patent office is the Deutsches Patent- und Markenamt (DPMA). Its procedures are similar to the JPO. No search reports are provided before the request for an examination. Applicant has seven years to request such an examination. Patent application is published 18 months after the filing date. DPMA allows, like the EPO an opposition procedure within three months after the publication of the patent specifications.

The French patent office is the National Industrial Property Institute (INPI). Concerning the INPI, the patent procedure is a hybridization form of the EPO and USPTO procedure. Once the application is filed, applicant does not have to request for examination. But this examination is done into two steps like at the EPO. In a first time the INPI provides a search report with an opinion on patentability. Applicant has to make observations on the relevance of the prior art cited and / or make changes in his claims within a delay of three months. Afterward, patent application is published (18 months after the filing date). Then, INPI proceeds to a substantial examination according to the modification provided by the applicant.

The United Kingdom patent office is the Intellectual Property Office (IPO). The IPO is closer to the EPO. Like the EPO, two distinct examinations are conducted (a search report and a substantive examination). But unlike the EPO, applicant has to request an examination within 12 months after the filing date (to draw up the search report). Finally, application is also published after 18 months and applicant has to request for a substantive examination within 6 months after the publication.

### 4.2.2.5 Cost for obtaining a patent

Cost for acquire a patent is not negligible. Whereas official administrative fees (filing fees, search, examination, country designation, grant/publication fees and validation fees) to obtain a patent granted and validated it are EUR 3 470 at the EPO, EUR 2 050 at the USPTO and EUR 1 570 at the JPO, a survey conducted in 2004 (EPO/Roland Berger, 2005 [9]) evaluates that a company from a European country (EPO member state) will pay on average in 2003 EUR 24 100 to have a Euro-direct patent granted and validated (costs for covering 6 countries); a US company will pay EUR 10 250 to receive a USPTO grant and a Japanese company will pay EUR 5 460 to acquire a JPO grant. These differences, between offices, are mainly explained by translation cost. Translating the very specific documents linked to a patent, especially the claims, involves very high translation costs. The renewal fees (the maintenance fees require to keep patent under protection) for procedure are also higher at the EPO

compared to the two other offices. Note that the renewable fees increase over the life of the patent (for a European patent applications it is about EUR 400 for the first year and EUR 1 350 for the 10th and each subsequent year).

In these sections we have reviewed the patent procedures and the technical characteristics of patents. This review allows us to turn to the debate concerning the utilisation of patent as an economic indicator of knowledge life.

### 4.2.3 Patent as an indicator of innovative activities

The idea of using patent data on a large scale for economic researches was first introduced by Schmookler (1966) [163], followed by Scherer (1982) [156] and Griliches & Lichtenberg (1984) [67]. As was noted long ago by Schmookler in his 1966 book, *Invention and Economic Growth*, patent data can be very helpful as a measure of innovative activities because it conveys information on the output and processes of inventive activities. Since the availability of information from the USPTO in machine readable form in late 1970s, patent was used to measure the innovative activities. As highlighted by Griliches in 1990 [65] “After all, a patent does represent a minimal quantum of invention that has passed both the scrutiny of the patent office as to its novelty and the test of the investment of effort and resources by the inventor and his organization into the development of this product or idea, indicating thereby the presence of a non-negligible expectation as to its ultimate utility and marketability”. The assumption that patent data reflect innovative activity has been validated in a number of studies. Main results of these studies are summarized as follows:

- patents are strongly related to R&D across firms,
- there are some difficulties to determine lag relationship between them because R&D is very smooth over time within firm,
- in the presence of R&D, patents added little explanatory power for sales, profits, and market value maybe because of the skewness distribution of patent value or importance.

These results point out that in absence of detailed R&D data, patents are the most interesting and useful indicator of innovative activities that economists could use. It is a second best: “The practical implication of these findings is that in the absence of detailed R&D data, the much more plentiful patent data can be used instead as an indicator of both, inventive input and output [...] patents statistics remain a unique resource for the analysis of the process of technical change. Nothing else even comes

close in the quantity of available data, accessibility, and the potential industrial, organizational, and technological detail.” Griliches (1990) [65].

To go further on the possibilities offer by patent into the measurement of innovative activities, the next two sections present the advantages and the disadvantages of patents into reflecting inventive activities.

### 4.2.3.1 Advantages

Advantages using patent as an indicator of innovative processes is threefold:

- First, patents have a close (if not perfect) link to inventions. Patent is by definition an output of the innovative activities.
- Second, patents cover a broad range of technologies, years and countries (patent data provide information on technology on which there are sometimes few other sources of data and so for nearly all countries in the world, back to the 19th century in most OECD countries). Each patent document contains detailed information on the inventive process. Details of information available in patent are provided in Table 4.2.
- The third advantage is that patents provide information on links between different quanta of knowledge via the citations to other patents and non-patent documents that they contain. With the possible exception of data on scientific paper publication, no other data source comes even close to provide this level and quantity of information about the creation and dissemination of new knowledge. We will go back to this point in the section 4.2.4.

### 4.2.3.2 Drawbacks

Even if it is useful, patents are however only an imperfect indicator of inventive activities, there are certain drawbacks in using patents as a proxy for innovation, namely :

“not all inventions are patentable, not all inventions are patented and the inventions that are patented differ greatly in “quality”, in the magnitude of inventive output associated with them.” (Griliches, 1990, p.1669) [65]

According once again to the OECD (2008, 2009) [132, 133], drawbacks can be summarized as following:

<b>Technical description of the invention</b>
Title and abstract (describes the invention).
The list of claims. This describes the innovative content of the claimed field of exclusivity. The claims define the scope of protection of the patent rights (legal boundaries). It can be more or less broad or narrow, depending on the content and number of claims.
The technical classes to which the invention pertains (based on patent classification). These are fixed by patent examiners. The most commonly used classification is the International Patent Classification (IPC) system.
Prior art. Each patent lists prior art relevant to the invention. The prior art determines the boundaries of what is in the public domain and what the applicant is entitled to in relation to the claims. The cited (patent and non-patent) references help to define the patent claims and its specific uses and applications. Citations may be made by the examiner or the applicant.
<b>Development and ownership of the invention</b>
The list of inventors and their respective addresses. The address given in the patent document is usually the professional address of the inventor (laboratory etc.).
The list of applicants (assignees in the United States) and their respective addresses. Applicants designates the ownership or control of the invention
<b>History of the application</b>
Publication number, application number, patent (grant) number. A patent number is a unique identifier of a patent. Patent numbers are assigned to each patent document by the patent-issuing authority. The first two letters designate the issuing patent office i.e. EP for EPO patents and US for USPTO patents.
Priority number. This is the application or publication number of the priority application, if applicable. It makes it possible to identify the priority country, reconstruct patent families, etc.
Priority date. First date of filing of a patent application, anywhere in the world, to protect an invention. It is the earliest and therefore closest to the invention date.
Application date: date when a patent is filed at a specific patent office. There is usually a 12-month lag between residents and foreigners. The lag extends up to 30 months for PCT procedures.
Publication date. Patents are normally published (i.e. the information is available to the public) 18 months after the priority date. Prior to the publication of a patent document, the content of the document remains secret. List of designation. For patent applications filed using the European Patent Convention or Patent Cooperation Treaty procedures, applicants are required to designate the member countries in which protection is being sought.
Refusal or withdrawal date. This indicates that the invention did not fulfil the statutory criteria (novelty, non-obviousness or industrial applicability) for patentability, or that the applicant decided to suspend the patent application during the examination process.
Grant date. date when the patent rights are conferred to the applicant by the authorised body.

Source OECD (2008, 2009)

Table 4.2: Information available in patent

1. The range of patentable innovations constitutes a sub-set of all research outcomes. Purely scientific advances which are not industrially applicable and small technological improvements are excluded from an analysis which uses patent data as proxy for innovation.
2. Patenting is a strategic decision and not all the patentable innovations are actually patented. Limitation in patenting could be due to a strategic decision (for instance, innovators could decide to protect their innovation by secrecy instead) or could results from regulatory framework. Most representative example comes from the US where until 1980 and the Bayh-Dole Act, US universities could not collect royalties for the use of patents derived from federally funded research. This situation was clearly a negative incentive for university to patent.
3. The propensity to file patent applications differs significantly across technical fields. For instance, in the electronics industry (e.g. semiconductors) a patented invention can be surrounded by patent applications on incremental variations of the invention, with a view to deterring the entry of new competitors and to negotiate advantageous cross-licensing deals with competitors. This situation leads to an over representation of this technical field.
4. Not all patents are equal in economic value or in technological opportunity embodied. Studies show that value distribution of patented innovation is highly skewed, with a few patents worth a lot and a lot patents worth nothing (e.g. Pakes & Schankerman, 1986 [155]; Harhoff et al., 1999 [73]).
5. Differences in patent law, practices and economic costs around the world limit the comparability of patent statistics across countries. Major differences between EPO, JPO and the USPTO are described above.
6. Changes in patent laws over years call for caution when analyzing trends over time. For instance see the the Bayh-Dole Act above.

Patent literature has validated the use of patent as economic indicator of innovative activities. It provides helpful information when R&D data are missing. But this indicator presents several limitations that we have to keep in mind when we attempt to measure technological change. One of the particular characteristic of patent is the inclusion of references toward previous patents and scientific literature which can be used as a way for tracking knowledge flows, this will be the subject of the next section.

### 4.2.4 Citations and knowledge flows

Patent data include references or citations to previous patents and to the scientific literature. Citations perform a legal function identifying previous patents that delimits the property right covered by the patent. These citations offer the possibility to trace multiple linkages between inventions, inventors, scientists, firms, locations... In particular, patent citations allow us to study the cumulative property of knowledge, and thus to measure the “importance” or technological impact of individual patent.

#### 4.2.4.1 Concepts

When a patent is published, the patent document identifies a list of references or citations, which are previous patents or non patent materials such as scientific articles upon which the current patent is built. There are basically two kinds of citations. Patent references are citations to previous patents filed anywhere in the world, at any time, in any language. References categorized as non-patent literature (NPL) are scientific publications, conference proceeding, books, database guides, technical manuals, standard descriptions, etc. Patent and non-patent citations are the references provided in the search report which are used to assess an invention’s patentability and help to define the legitimacy of the claims of the new patent application. As they refer to the prior art, they indicate knowledge that precede the invention and may also be cited to show the lack of novelty of the citing invention. However, citations also indicate the legal boundaries on the claims of the patent application in question. Therefore, they serve an important legal function, since they delimit the scope of the property rights awarded by the patent.

The granting of the patent is a legal statement that the idea embodied in the patent represents a novel and useful contribution over and above the previous state of knowledge, as represented by the citations. Thus, in principle, a citation of patent *A* by patent *B* means that *A* represents a piece of previously existing knowledge upon which *B* is built, and over which *B* cannot have a claim. Actually a patent is similar to a bibliographic citation. Inventor must disclose all prior art related to its invention. But at the difference of bibliographic citation, patent citations are controlled by an examiner who is supposed to be expert in the area and which could removes some biases. Because of the role of the examiner and the legal significance of patent citations, there is reason to believe that patent citations are less likely to be contaminated by extraneous motives in the decision of what to cite than other bibliographic data. Figures 4.8 & 4.10, in Appendix, present citations made by patents P1 and P12, as we have seen above, citations are included and published in the search report.

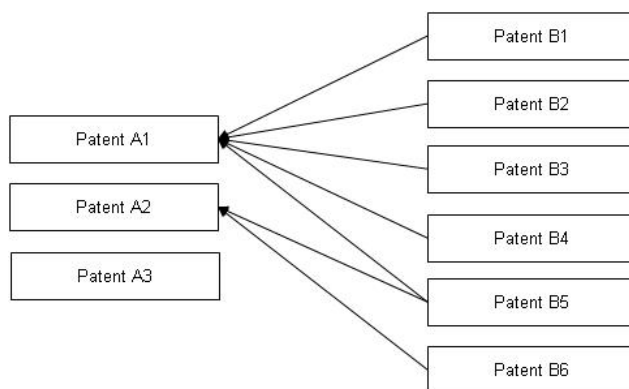


Figure 4.5: Patent Citations

Figure 4.5 illustrates the patent citation mechanisms. For instance, Patent *A1* is cited by Patents *B1-5*, meaning, under the assumption that patent citation is a proxy for knowledge flows, that Patent *A1* embodies knowledge or prior art upon which Patents *B1-5* are built. Patent *B5* cites both Patent *A1* and *A2* indicating that these two Patents were significant in the creation of invention embodied in Patent *B5*. On the contrary, Patent *A3* is not cited by any subsequent patents meaning that the knowledge embodied in this patent was not useful for the development of any invention embodied in Patents *Bx*.

#### 4.2.4.2 Practical differences between offices

Citation process differs slightly across the different patent offices. Although European patent offices and the Japan office are quite close each other, USPTO differs in citation procedures:

- USPTO: For an application to the USPTO, inventor or his lawyer must provide a list of references describing the state of the art on which the invention is based (known as the “duty of candour” rule). Per consequence, they tend to include all the cited references as possible, even if some are not directly related to what is patented. As the examiners do not seem to be bothered by the limitations described in inventor’s list, they generally leave the list unchanged. There is a strong incentive to provide too much citation. For the U.S. patents about 60% of all citations are added by inventors (Alacer & Gittelman, 2004 [8]).
- EPO: In contrast, no such provision exists for the EPO. The applicant may cite prior art within the application document but this is optional (only 3.9 % of all EPO citations are added by inventors). Nevertheless the list of references is in part based on the applicant disclosure of

prior art submitted with the patent application (at the EPO; this is done in the information disclosure statements). According to EPO philosophy, a good search report contains all relevant information within a minimum number of citations, i.e. the minimum number needed to covert prior art. (Michel & Bettels 2001 [122]). Finally, examiners have the obligation to favour recent documents over older ones.

- The European National patent procedures are in general similar between all European countries (Felix 2006 [49]). A patent examiner determines the state of the art relevant for the invention and examine whether a patent can be granted against this background. Applicants do not have to refer the prior art relating to the invention.
- At the JPO, patent examiners conduct the search of the prior art. In this sense JPO citation data is similar in nature to the citation information derived from search reports produced by examiners at the European Patent Office (Michel & Bettels 2001 [122], Goto & Motohashi 2007 [55]). On the other hand, since 2002, Japan adopted the prior art reference disclosure practice. According to this, applicants must file a detailed list of prior art references known at the time of the patent application. However, there are no associated penalties for failing to do so. It seems that most applicants nevertheless list references extensively.

### 4.2.4.3 Patent citations and knowledge flows

The extent to which patent citations represent spillovers is controversial. Despite patent citations were used since 1993 to proxy knowledge spillovers, the empirical validation of such assumption was done nearly a decade latter. First attempt to measure this extent was made by Jaffe, Trajtenberg and Fogarty (2000) [88]. Authors surveyed US 166 responses of inventors about citing USPTO patents. Through the responses surveyed they conclude that patent citations can be interpreted as “providing a (noisy) signal of spillovers”. About 38% of the answers indicate that they had learned about the research or work underlying the cited patent before and during the development of their invention, 30% after essentially completing their invention and the remaining (less than one third) indicate that they never know before the cited patent. Jaffe et al. (2000) stress that this result is not surprising because inventors of the citing patent are not always aware of patents cited in the search report, as



citations are frequently given by patent office examiners or by patent attorneys<sup>8</sup>. They conclude:

“the likelihood of knowledge spillovers(...) is significantly greater (...)than the likelihood without a citation (...but...) a large fraction of citations, perhaps something like half, do not correspond to any apparent spillovers (...) citations are a noisy signal of the presence of spillovers”.

The main criticism raised against patent citations, as a proxy for knowledge flows, comes from the fact that examiners are ultimately responsible for citations and that a large fraction is added by them. These critics are more problematic when we turn to the EPO citations because only a small fraction is added by inventors, about 3.9 %. All knowledge flows are not necessary captured by inventor’s citations. As highlighted in the study (Jaffe et al. 2000), inventors may cite prior art by conducting a search after completing the invention (40% of all citations were learnt during the patent application process) or inventor identified citations can result from an attorney addition, thereby these citations do not reflect a knowledge flow. Moreover, Jaffe et al. (2000) interpret unawareness of cited patent as an absence of spillovers.

For Breschi & Lissoni (2004) [23] this reduction is questionable. Even if inventor’s citations are more likely to represent a true knowledge flow, examiner’s citations are more likely to reflect ignorance on the part of the inventor. In this sense, inventors may fail to cite prior art that they knew about, and these are eventually added by the examiner. In other hand, it is reasonable to assume that as patents are public knowledge, professional R&D laboratories would have a reasonable knowledge about existing patent in their field, thereby examiner citations can be interpreted as an useful indicator of technological relevance and of potential knowledge flows (Crisculuo et al. 2005 [38]).

Breschi & Lissoni (2004) investigate this question and the extent by which examiner’s citations and thereof EPO citations could be used to measure knowledge flows. While Jaffe et al. (2000) associate the presence of knowledge spillovers with the awareness of cited patent before completing the invention, Breschi and Lissoni prefer the concept of cumulative effort, traducing a cumulative link between the technological content of the two patents and not between the two patents (the two patents are : the cited and the citing patent). Thus, the technical knowledge embodied in the cited patent may reach the citing inventor through a social chain, while they did not know the existence of the related patent. For the authors it is clear that knowledge flows may be captured by patent citations even if inventors

---

<sup>8</sup>Concerning the mode of spillovers, 18% indicated that they had had either exposed to some kind of presentation or demonstration of the cited invention or had a direct communication with the inventor. 18% indicated that they learn through “word of mouth” or had read the patent document. 40% indicate that its due to the process of their own application and subsequent search.

are unaware of those citations:

“There is no reason to exclude that examiner’s citation (i.e. “unaware citations”) may indeed hide a knowledge flow. At most, we can presume that the cited and the citing inventors do not know each other, otherwise the former would have passed both the information on the patent’s contents and the information on the existence of the patent itself. But it may well be the case that the two are linked by one common acquaintance, or a social chain of acquaintances, who are responsible for passing on the information on the patent contents, as well as the necessary “tacit bits” to build upon them: the longer the chain, the more likely the case that other bits of information, specifically those on the existence of the cited patent, have gone lost during the diffusion process.” (p623) Breschi & Lissoni (2004).

To confront their assumptions, authors investigate the relation between social chains and EPO patent citations. They construct an index of geodesic distance<sup>9</sup> of Italian inventors and test how this variable explains patent citations. They find that inventors linked by not too long social chain (about four edges) are more likely to be linked by patent citations than unconnected ones. This result shows that EPO citations (made by examiners) well reflect the scientific social chain, which is an obvious channel for knowledge spillovers. After this preliminary work, the use of EPO citation data in the study of spillovers was empirically confirmed principally by two studies. Duguet & Mac Garvie (2005) [44] assess the relevance of EPO patent citations as proxies of knowledge flows by exploring the correlation between citation patterns and the answers to the Community Innovation Survey (CIS) questionnaire using a sample of French firms. They find that patent citations are more likely to occur between firms involved in partnership that may induce knowledge transfers (these partnership activities can be R&D cooperation, right to use the inventions of a third party, business acquisition, joint venture ...). The correlation between citations and activities varies significantly according to the source and destination of the technology transfer. Nevertheless, the existence of citations in the presence of activities that are sources of knowledge transfer tends to reinforce the idea that they are well correlated with knowledge spillovers. MacGarvie (2006) [115] uses EPO citations to measure the impact of foreign knowledge on French firms and control the validity of EPO citations by using the same sample with USPTO citations. The findings are quite similar to those obtained with EPO citation data.

This section has extensively presented patent through its different facets. In one hand, we have defined what is a patent, what are the difference practices over the different offices and in the other

---

<sup>9</sup>The geodesic distance is defined as the minimum number of steps (or, more formally, “edges”) that separate two distinct inventors in the network.

hand, how it could well proxy the innovative activities. One of its aspects, the inclusion in the search report of the prior art over which the new invention is built, is particular relevant in our attempt to measure technology flows. Patent citations could be seen as a window to the cumulative characteristics of knowledge and thus offers a measure of knowledge spillovers. The next section will be based on this property and will present the way by which patent citations could be used in order to build knowledge stocks while taking into account international spillovers.

### 4.3 The citation function

We follow the Jaffe & Trajtenberg (2002) [87] conception about the use of patent citations. We have shown that patent citations are a predicate on a cumulative view of the process of technological development, by which each inventor benefits from previous inventors' work, and in turn contributes to the base of knowledge upon which future inventions will be built. Thereby, thanks to patent citation data, it is possible to identify, analyze and estimate the attributes of knowledge which leads to the technological development of the economy. The first empirical analysis of patent citations was done by Caballero & Jaffe in 1993 [25]. Authors developed a structural model to estimate knowledge flows' determinants according to patent citation data. This model is nowadays commonly called the "patent citation function". This function was developed by Jaffe and Trajtenberg during the nineties through different applications but away from the first aim which was the construction of knowledge stocks. Here, we will present the origin of this model and its major developments. Secondly, we will turn to the technical characteristics of this function before presenting the way by which the model could be extended in order to take into account international knowledge spillovers in the construction of knowledge stocks.

#### 4.3.1 The origin of the patent citation function

The citation function was first introduced by Caballero and Jaffe in 1993. Authors attempt to provide a framework for incorporating the microeconomics of creative destruction and knowledge spillovers into a model of growth following Grossman & Helpman (1991) [68] and Aghion & Howitt (1992) [3]. They developed a quality ladder model where innovation is function of resource devoted to research and its productivity:

$$\dot{S}_T = \theta_T L_T^r$$

Where  $S_T$  is a quality index of new produced good,  $L_T^r$  is the quantity of labor devoted to research activity and  $\theta_T$  is the average productivity of research that depends of past knowledge accumulation. The productivity of research is function of the public stocks of past knowledge and the probability to use it:

$$\theta_T = \int_{-\infty}^T p^*(T, t) \dot{S}_t dt \quad (4.1)$$

Productivity or stock of knowledge available is assumed to be identifying as the sum of past ideas potentially usable represented by  $\dot{S}_t$  and by the probability,  $p^*(T, t)$ , that these ideas developed in year  $t$  are used in subsequent year  $T$ .  $p^*(T, t)$  could be interpreted as a knowledge diffusion function that explains the transmission mechanism of ideas. Authors assume that this probability is dependent on three factors. First factor is linked to the concept of endogenous obsolescence. Very old ideas are unlikely to contain useful information for generating new ideas. Endogenous concept means that it is not the passage of time that makes old ideas less useful, it is the accumulation of new ideas. Second factor is underlying to the time needed by the idea to be diffused. Inventors take time in seeing others inventions which suggests that there are diffusion lags. Unlike obsolescence, diffusion depends on time. Moreover, the intensity of using previous ideas may vary independently of the effect of obsolescence and diffusion. Many factors could influence the mechanism of transmission of ideas, such as the geographical or institutional origins of knowledge (recall the large literature on the extent by which public knowledge create externalities with respect to private knowledge). Thus, the third factor is not well defined. Authors assume that it could vary in function of specific attributes of the different idea cohorts. By grouping these three factors into a function measuring the transmission mechanism of knowledge authors created that we called the ‘‘citation function’’:

$$p^*(T, t) = \delta e^{-\beta_1(S_T - S_t)} (1 - e^{-\beta_2(T-t)}) \quad T \geq t \quad (4.2)$$

with  $\beta_2 \geq 0$ ,  $\beta_1 \geq 0$  and  $0 \leq \delta \leq 1$ .

The citation function  $p^*(T, t)$  is the product of the usefulness of old ideas  $\delta e^{-\beta_1(S_T - S_t)}$  and the probability of having seen them  $(1 - e^{-\beta_2(T-t)})$ . The probability of seeing or knowing an idea created in year  $t$  at the date  $T$ , thus to see a  $(T - t)$  year old idea, is  $(1 - e^{-\beta_2(T-t)})$ . The usefulness of the ideas

is a decreasing function of the technology advancement since the creation of considered idea ( $S_T - S_t$ ) and it is equal to  $e^{-\beta_1(S_T - S_t)}$ . The parameter  $\delta$  could capture two distinct effects. It could represent the potential of spillovers emanating from each cohort of ideas or an absorption capacity, measuring the intensity of use of old ideas by new ideas.

This equation captures the situation where inventors can build over the existing stock of knowledge, but there are limits on its usefulness in generating new ideas. Equation (4.1) captures the more general case in which the research productivity parameter  $\theta_T$  depends on the stock of existing ideas weighted by their probability to generate new knowledge at time  $T$ . This probability is, in turn, dependent on the likelihood that the previous idea is known to a current inventor, and the likelihood that it is useful.

Caballero and Jaffe assume that  $\dot{S}_T$  is proportional to the rate of patenting at time  $T$ . They use patent citations between patent cohort granted in year  $T$  to patent cohort granted at year  $t$  as an indicator of the use of knowledge of vintage  $t$  in the production of new ideas at time  $T$ . Thus the probability at year  $T$  of using an idea developed in year  $t$  is linked to patent citations :

$$p(T, t) = \frac{C_{T,t}}{N_T N_t} \quad (4.3)$$

where  $p(T, t)$  is an estimates of the probability of citation,  $C_{T,t}$  is the observed citations from patents granted in year  $T$  to patents granted in year  $t$ ,  $N_T$  is the number of patents granted in year  $T$ , and  $N_t$  is the number of patents granted in year  $t$ . They assume that citations are proportional to “used ideas” with a proportionality factor  $\phi_T$ , representing the citing behavior of agent over time because practices of citing could affect the number of citations.  $\beta_1 S$  is proportional to patent with a proportionality factor,  $\psi_T$  (they assume that parameter  $\beta_1$  is not identified by the patent equation but by a subsequent growth equation). Patents can encompass a number of distinct ideas,  $\psi/\beta_1$  controls the average size of a patent (the inverse of the “propensity to patent” of Griliches 1990). Since not all ideas are patented,  $\psi/\beta_1$  can also be the reciprocal probability to any given idea to be patented. Thus,  $C_{T,t}/\phi_T$  represents “ideas used” and  $(\psi/\beta_1)S$  represents new ideas  $\dot{S}$ . The probability that new ideas  $\dot{S}_T$  use previous ideas  $\dot{S}_t$  is

$$p^*(T, t) = \frac{(C_{T,t}/\phi_T)}{\dot{S}_T \dot{S}_t} = \frac{(C_{T,t}/\phi_T)}{(\psi_T/\beta_1)N_T(\psi_t/\beta_1)N_t}$$

Identifying the probability of citation as defined in equation (4.3):

$$p^*(T, t) = \frac{C_{T,t}}{N_T N_t} \frac{1}{(\phi_T / \beta_1^2) \psi_T \psi_t} = p(T, t) (\beta_1^2 / \phi_T) \psi_T^{-1} \psi_t^{-1}$$

Isolating it:

$$p(T, t) = (\phi_T / \beta_1^2) \psi_T \psi_t p^*(T, t)$$

Using equation (4.2):

$$p(T, t) = (\phi_T / \beta_1^2) \psi_T \psi_t \delta_t e^{-\beta_1(S_T - S_t)} (1 - e^{-\beta_2(T-t)}) \quad (4.4)$$

$N_T - N_t$  is the number of patent granted between the years  $t$  and  $T$ , if we weight it by  $(\psi / \beta_1)$ , we can substitute  $S_T - S_t$  by  $\sum_{x=t}^T (\psi_x / \beta_1) N_x$  in equation (4.4) and finally obtain:

$$p(T, t) = (\phi_t / \beta_1^2) \psi_T \psi_t \delta_t e^{-\beta_1 \sum_{x=t}^T (\psi_x / \beta_1) N_x} (1 - e^{-\beta_2(T-t)}) \quad (4.5)$$

To summarize, variables entering into the citation function are:

- $\psi$  is the number of ideas per patent over time, i.e. the “propensity to patent”.
- $\phi$  is the propensity factor between “ideas used” and citations (it represents change in the number of citation across patents over time).
- $\delta$  is the potency of the spillovers emanating from each cohort of ideas (its variation represents changes in the potency of knowledge spillovers).
- $\beta_2$  is the rate at which new ideas diffuse.
- $\beta_1$  is the rate at which new ideas outdated old ones.

This specification raises several identification problems. First,  $\beta_1$  can not be directly estimated from the equation (4.5) (it is possible to estimate  $\beta_1 \dot{S}$  but not  $\beta_1$ ). Secondly, other parameters can be identified up to a normalization that sets one to  $\delta_t$  over time because equation (4.5) encompass both  $\delta_t$  and  $\psi_t$  that are indexed by the same year index  $t$ . Nevertheless, this specification was very useful in allowing for the first time an empirical estimation of the diffusion rate of ideas and in providing a way for tracking the influence of past inventions across time. This model paves the way for the measurement of spillovers capacities arising from a specific cohort of ideas with specific characteristics.

As ideas are identified with patent data and that patent data are available at a very detailed level, the definition of these cohorts can take several forms allowing the investigation of multiple knowledge diffusion patterns.

### 4.3.2 Rationalization of the patent citation function.

#### 4.3.2.1 Jaffe and Trajtenberg

The second step in the development of the citation function was realized by Jaffe & Trajtenberg in two stages: 1996 [84] and 1998 [85]. In a first time, they used the citation function to analyze knowledge flows arisen from University and Federal Laboratories. Secondly, they used the function to analyze international knowledge flows. The two developments were relatively similar and have the merit of simplifying considerably the original model. Authors adapted the formulation of Caballero & Jaffe 1993 and since they did not attempt to build any stock of knowledge, they focused directly on patents dropping the formulation linked to ideas. Moreover, Jaffe & Trajtenberg did not consider any endogenous growth model in which the citation function could be encompassed. They directly assumed that the likelihood that some particular patents  $K$  granted in year  $T$  cite some particular patents  $k$  granted in year  $t$  is:

$$p(k, K) = \alpha(k, K) \cdot e^{-\beta_1(k, K)(T-t)}(1 - e^{-\beta_2(T-t)}) \quad (4.6)$$

The closest Jaffe and Caballero equation to this formulation is the equation (4.4) but unlike this one, the rate of obsolescence  $\beta_1$  is function of time and not of the number of past ideas ( $S_T - S_t$ ) and differs according to characteristics of cited  $k$  and citing  $K$  patent cohorts. In addition of the combination of the double exponential process, citations are not only function of patent characteristics related to time ( $T, t$ ) but can also be function of other attributes of the cited  $k$  or/and citing patent  $K$ . Influences of patent attributes on citations are taking into account by the shift parameter  $\alpha(k, K)$ . Such characteristics or attributes can be the kind of organization where the inventions are assigned (university, firm...), the inventor residence country (US, Japan ....), the technologic field, etc...

In their 1996 paper [84], they considered that the citation process is, moreover than the double exponential process, function of the attributes of the cited and citing patent. For the potentially cited patent  $k$  the considered attributes are:

- $t$ , the grant year

- $g$  the institutional nature of the assignees (corporate, university, or government)
- $i$  the technological field

For the potentially citing patent  $K$  the considered attributes are:

- $T$ , the grant year
- $L$  the localization of the citing inventor

Thus, equation (4.6) is rewritten in order to be estimate as:

$$p_{itgTL} = \alpha_i \alpha_{tp} \alpha_g \alpha_T \alpha_L \exp [ -(\beta_1) \beta_{1,i} \beta_{1,g} \beta_{1,L} (T - L) ] \times [ 1 - \exp(-\beta_2(T - t)) ] + \epsilon_{igtTL} \quad (4.7)$$

Where the shift parameter,  $\alpha(k, K)$ , is divided into five components,  $\alpha_i, \alpha_{tp}, \alpha_g, \alpha_T, \alpha_L$ , each of them corresponding to an attribute defined above. To deal with identification issues, they define for each component/attribute a base group and they constraint the shift parameter corresponding to this base group to one. A higher value for another group belonging to the same attribute is interpreted as a higher probability of citation. As specified in equation (4.7), the obsolescence parameter,  $\beta_1$ , is also a function of the patent cohorts attributes, namely: the institutional nature and the technological field of the cited patent and the location of the citing patent, allowing different diffusion patterns among attributes.

In their 1998 paper [85], the citation function is slightly different:

$$p(k, K) = [1 + \mu D(k, K)] \alpha(k, K) \cdot e^{-\beta_1(k, K)(T-t)} (1 - e^{-\beta_2(T-t)}) \quad (4.8)$$

$D(k, K)$  is a dummy variable, set equal to unity if the patent  $k$  is in the same patent technological class as the patent  $K$ , and zero otherwise. Thus the parameter  $\mu$  represents the overall increase in citation frequency associated with the two patents matching in the same class. For the potentially cited patent  $k$  the considered attributes are:

- $t$ , the grant year
- the geographical location of the cited inventor  $l$
- the technological field  $i$



For the potentially citing patent  $K$  the considered attributes are:

- $T$ , the grant year
- the geographical location of the citing inventor  $L$

The overall citation intensity measured by variation in  $\alpha$ , according to cited patent attributes, is composed in two parts. (i) Citation intensity that is the product of the fertility, or importance of the underlying ideas in spawning future technological developments, and (ii) the average size of a patent, i.e. how much of the unobservable advance of knowledge is packaged in a typical patent. But as we will see below, the two studies of Jaffe & Trajtenberg (1996, 1998) do not fully focus on that topic. Due to some convergence issues in the estimations process, they can not allow  $\alpha$  to vary accordingly to each cited  $t$  and citing year  $T$ . They favored the last one, by letting it to vary over each citing year, while the cited years are grouped into a five year interval. Per consequence, changes in the importance of ideas could only be identifying between grouped cohorts, limiting the dynamic analysis of knowledge entering into the cumulative process of technological change.

The reason why the obsolescence parameter, and not the diffusion one, is function of the patent attributes is also a question of identification. The maximum value of the citation frequency is approximately defined by  $\frac{\beta_2}{\beta_1}$ . An increase in  $\beta_2$ , holding  $\beta_1$  constant, rises the overall citation intensity and is roughly equivalent to an increase in the citation frequency at every value of  $(T - t)$ . As a consequence, a variation in  $\beta_2$ , with  $\beta_1$  unchanged, is not separately identified from variations in the multiplicative term  $\alpha$ . Thus, since the model is somewhat easier to estimate and to interpret with variations in  $\alpha$ , they did not allow variations in  $\beta_2$ .

First limitation of such approach lies in considering the grant year for the citing patent. At the USPTO, the year of granting is in average 3 years after the application date and is thereby not the date closest to the invention date. OECD (2008) [132] recommends the use of the priority date (the date of the first application) because it is the closest date to the invention. Second limitation that holds for Jaffe & Trajtenberg (1998) [85] is the investigation of an international dimension with a national database. Domestic applicants, proportionate to their innovative activities, tend to file more patents with their home country patent office than foreign applicants do. Consequently, foreign applicants are under represented. This effect is known as the “home advantage effect” and considerable international issue with the USPTO data induces a positive bias towards US applicants. We will recur on this point at the end of this chapter (in section 4.5.3).

### 4.3.2.2 Popp and the return on the original idea

Second set of development was made by Popp (2002) [146]. The main purpose of the author was to develop an input based knowledge indicator of energy efficiency technology in order to implement an induce innovation test. Since data on R&D are not available at this level (on technological level for energy technologies), the author uses a formulation of knowledge stocks close to that of Caballero & Jaffe (1993) while incorporating the improvements made by Jaffe and Trajtenberg (1996, 1998). His stock of knowledge is based on patent flows weighted by their technological opportunities which are estimated with the citations function. Unlike previous developments, Popp estimates the productivity of patent cohorts for each cited year  $t$ . His work is based on the assumption, extensively discussed on previous sections, that citation to an earlier patent suggest that this patent provides knowledge upon which the current inventor develop his invention. Thereby, frequent citations to a patent provide evidence that the knowledge embodied in that patent have been particularly useful to other inventors and that it encompass high technological opportunities. In the example shown in figure 4.5, it becomes clear, all other things being equal, that Patent  $A1$  provides high technological opportunities comparing to Patent  $A2$  and even more to patent  $A3$  which seems not to provide any. These technological opportunities are taking into account in the shift parameter of the citation function that is allowed to vary over the cited patent year index  $t$ . For the potentially cited patent  $k$  the considered attributes are:

- $i, t$ , the grant year  $t$  of a specific technological filed  $i$ ,
- $i$  the technological field.

For the potentially citing patent  $K$  the considered attributes are:

- $T$ , the grant year

Recall that in previous developments, the year indexes of cited patent were grouped into five years interval. The productivity component of the shift parameter, noted  $\alpha_{i,t}$ , is interpreted as the likelihood that patents of the year  $t$  will be more or less cited by subsequent patents than those belonging to the base group. Higher values for  $\alpha_{i,t}$  means that the knowledge embodied in these patents are particularly useful.

His estimating equation is:

$$p(k, K) = \alpha_{i,t} \alpha_i \alpha_T \cdot e^{-\beta_1(T-t)} (1 - e^{-\beta_2(T-t)}) + \varepsilon \quad (4.9)$$

Firstly, the depreciation parameter unlike previous developments does not depend on the attributes of the cited or citing patent. Secondly, author refers to the application date and not to the grant date for the citing patent  $T$ . Thanks to the parameters estimated with equation (4.9), Popp build a stock of knowledge,  $K(i, t)$  which is based on patent flows  $N_{i,t}$  weighted by their technological opportunities  $\alpha(i, t)$  through the following specification:

$$K(i, t) = \sum_{s=0}^t \alpha(i, t) \cdot N_{i,t} \cdot e^{-\beta_1(t-s)} (1 - e^{-\beta_2(t-s)}) \quad (4.10)$$

Where  $s$  is the base year,  $N_{i,t}$  is the number of patent granted in year  $t$  belonging to the technological class  $i$ . In this way, his formulation is similar to the one developed by Caballero & Jaffe (1993) at the difference that the productivity parameters (the potency of the spillovers for Caballero & Jaffe) can vary depending on all the cited years  $t$ . Note that Popp does not matter about the number of ideas per patent and builds directly his stock with patent as the unit of measurement. In such specification, it is the cited cohort that matter. Moreover, considering productivity parameters allows a kind of control over the number of ideas per patent. If previous patents embody more ideas, all other things being equal, they are more likely to be cited and the productivity parameters increase. As highlighted in sections 4.2.3, not all patents present the same usefulness and using patent citations is one way to deal with the highly skewed distribution of patented innovation. Popp (2002) constructed two different stocks of knowledge, one weighted by the technological opportunity as presented in equation (4.10) and one without weights. The author found that only the first stock is statistically significant in explaining the development of new technologies, meaning that when we want to consider patent as an input indicator, it is necessary to control it by its technological value.

## 4.4 Underlying assumptions and properties of the citation function

Before turning to our own developments of the citation function and the way by which we can use it as a measurement of knowledge stocks and international knowledge spillovers, we will describe some important properties of this function.

### 4.4.1 Underlying assumptions

The citation function identifies separately the contribution of citation lag distribution (obsolescence and diffusion of knowledge), technological opportunity (fertility) and the propensity to cite. But the use of this function induces some assumptions:

- Proportionality: the shape of the lag distribution over time is independent of the total number of citations received, and hence more highly cited patents are more highly cited at all lags.
- Stationary: the lag distribution does not change over time, i.e. it does not depend on the cited patent cohort.

According to Hall et al. (2001) [71], there are some evidences supporting the proportionality assumption, there is almost zero correlation between the average forward citation lag per patent and the number of citation received. The second assumption is difficult to test because the observed citation-lag distribution could have different reasons to shift over time.

### 4.4.2 Properties of the function

Now and thereafter, we consider the following citation function :

$$p(k, K) = \alpha \cdot e^{-\beta_1(T-t)}(1 - e^{-\beta_2(T-t)}) \quad (4.11)$$

#### 4.4.2.1 The modal lag

The modal or most frequent lag in citation is a robust measure of diffusion's speed. The modal lag is the lag at which the citation probability peaks. It is given by the first order condition of the citation function with respect to the lag  $L = T - t$ , i.e.  $\frac{\partial p}{\partial L} = 0$ . After some calculations detailed in Appendix we get:

$$L_{Mod} \simeq \frac{1}{\beta_1}$$

#### 4.4.2.2 The cumulative citation probability

The total number of citations, that a patent could receive, is in theory the cumulative citation probability corresponding to  $L \rightarrow \infty$ . This is given by :

$$C(\infty) = \int_0^{\infty} \alpha \cdot e^{-\beta_1 L} (1 - e^{-\beta_2 L}) dL$$

which after some calculations becomes:

$$C(\infty) = \frac{\alpha \beta_2}{\beta_1 (\beta_1 + \beta_2)} \quad (4.12)$$

#### 4.4.2.3 Median lag

To compute the average lag in citation we just have to multiply the probability of citation by the lag  $L$  and to integrate it to infinity,  $L \rightarrow \infty$ :

$$M(\infty) = \int_0^{\infty} L \times \alpha e^{-\beta_1 L} (1 - e^{-\beta_2 L}) dL \quad (4.13)$$

$$M(\infty) = \alpha \frac{(\beta_1 + \beta_2)^2 - \beta_1^2}{\beta_1^2 (\beta_1 + \beta_2)^2} \quad (4.14)$$

However, this average is based on a cumulative “probability” that does not sum to unity, as equation (4.12) shows. To obtain the mean we divide equation (4.14) by (4.12) so that the probability mass is normalized to 1 and leads to :

$$\frac{M(\infty)}{C(\infty)} \simeq \frac{2}{\beta_1}$$

$$L_{mean} = 2L_{Mod}$$

## 4.5 The input based patent citation indicator

Popp (2002) considers citations only as a temporal spillovers measure, i.e. the extent by which previous knowledge contributes to the discovery of new knowledge. This specification can be extended to other kind of spillovers such as geographic spillovers. To fill this gap, we will improve previous specifications in order to take into account not only inter temporal knowledge spillovers, but also international knowledge spillovers. We will start by defining the input based knowledge indicator and discuss the meaning of the productivity parameters in an international context. Since our developments require

multiple database we will end this section by explaining the implications raised by this coverage.

### 4.5.1 Citations and technological opportunities embodied in cited patents at an international level

Following previous interpretations on patent citations as a measure of knowledge flows, we assume that patents receiving many citations from subsequent patents must have provided greater technological opportunity. Nevertheless, we differ from previous model by considering that the importance of technological opportunity does not depend exclusively on the cited patent but also depend on the characteristics of the citing patents.

Let's assume three patents cohorts belonging to the same technological field but located in different regions ( $a$  is the region of cited patents,  $b$  and  $b'$  of the citing patents with  $b$  different from  $b'$ ). If cohort  $a$  is cited by cohort  $b$  and cohort  $b'$ , it does not mean that the technological opportunity created by cohort  $a$  is the same for cohort  $b$  and  $b'$ . The reason of that is underlying on the characteristics of region  $b$  and  $b'$ . In other words, the same knowledge arising from a specific region can have different productivity impacts in function of the citing innovator's location. For instance, assume that  $a$  is a country on the technological frontier and presents high productivity values for  $b$  and  $b'$  patents (respectively  $\alpha(b, a)$  and  $\alpha(b', a)$  with  $\alpha(b, a)$  the productivity of cohort  $a$  for cohort  $b$ ). We could expect that country  $a$  is out-sourcing technology to countries  $b$  and  $b'$ . But the importance of spillovers created depends on the underlying country pair characteristics ( $a, b$  and  $a, b'$ ) and  $b$  and  $b'$  absorption capacities. If  $b$  has a higher absorption capacity than  $b'$ , all other things being equal,  $b$  is expected to benefit more from  $a$  knowledge and to present a higher value for its patent productivity parameters than do  $b'$  ( $\alpha(b, a) > \alpha(b', a)$ ).

For instance in figure 4.6, Patent  $A1$  (belonging to cohort  $a$ ) seems to provide equal technological opportunities for cohort  $b$  and  $b'$ , but Patent  $A2$  seems to provide technological opportunities only for region  $b$  patents. Since the importance of technological opportunities embodied in cohort  $a$  depends on the frequency of citation received, region  $a$  provides greater technological opportunities for region  $b$  than for region  $b'$ . This is principally due to the number of citations received by Patent  $A2$ . This case can be explained by some minimal technological requirements needed to benefit from knowledge embodied in Patent  $A2$ .

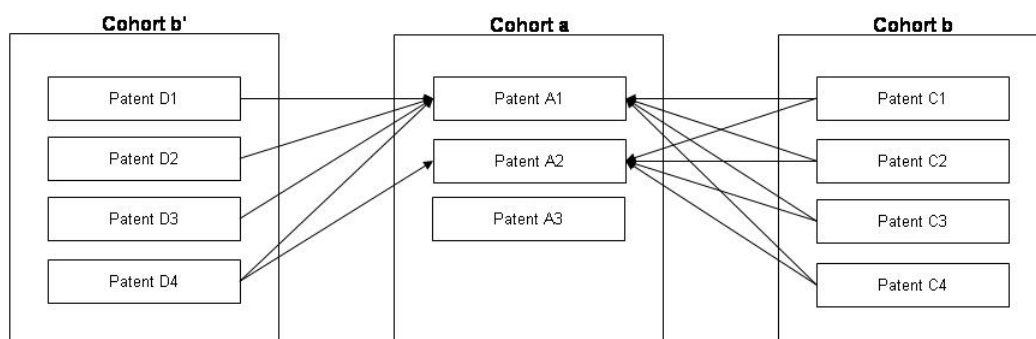


Figure 4.6: Patent Citation between Regions

### 4.5.2 Construction of our input-based patent citation indicator

We define two stocks of knowledge: a domestic stock and a foreign stock of knowledge. The domestic stock of knowledge,  $Knowd$  follows the formulation of Popp (2002), as in equation (4.10). We assume that the domestic knowledge stock available for an inventor located in country  $L$  at date  $T$  is built on a flow of domestic patents publication  $N_{L,t}$  weighted by their productivities  $\alpha(L, L, t)$  :

$$Knowd_{L,T} = \sum_{t=0}^{T-1} \alpha_{L,L,t} \cdot N_{L,t} \cdot e^{-\beta_1 \cdot L \cdot (T-t)} \left( 1 - e^{-\beta_2 (T-t)} \right)$$

Unlike Popp, we consider that knowledge can come from abroad, thus we consider a foreign stock of knowledge  $Knowf$  which follows the same logic. The foreign knowledge stock available for an inventor located in country  $L$  at date  $T$  is built on a flow of foreign patents publication  $N_{l,t}$  weighted by their productivities  $\alpha(L, l, t)$  :

$$Knowf_{L,T} = \sum_{l \neq L} \sum_{t=0}^{T-1} \alpha_{L,l,t} \cdot N_{l,t} \cdot e^{-\beta_1 \cdot l \cdot (T-t)} \left( 1 - e^{-\beta_2 (T-t)} \right)$$

This stock is the sum of foreign knowledge stocks indexed by  $l$ . Here we consider that patent productivities are not only intrinsic to the characteristics of the related knowledge (indexes by  $l$  and  $t$ ). We assume that they can also depend on the characteristics of the inventors that could benefit from this knowledge (indexes by  $L$ ).  $N$  represents the creation of new knowledge. In this specification, productivity parameters embody not only the characteristics of inter temporal spillovers of knowledge, it also takes into account international knowledge spillovers.

We can interpret  $\alpha$  as an indicator of inter temporal spillovers and as an indicator of international spillovers. Imagine in a first time, that  $\alpha$  is constant over time. In that case,  $\alpha$  can be interpreted as

the weighted scheme of external knowledge (it correspond to the weighted scheme  $w_{ij}$  of the chapter 2). Lower values relatively to the based group (i.e.  $\alpha_{L,L} = 1$ ) indicate that the country is most likely to benefit for its own knowledge than from knowledge developed abroad.

Now, if we let this parameter varying over time, the temporal dimension of knowledge spillovers is added. Changes of  $\alpha$  over time may then come from two reasons. First, we could imagine that for some independent reasons, relationship between two countries improves (student exchanges, R&D cooperation projects...) inducing a positive impact on  $\alpha$ . Secondly, variation in the productivity parameter may be due to changes in the technological productivities of knowledge embodied in the relating cohorts. If a particular cohort encompass more useful knowledge it will be more likely to be used for future developments and its  $\alpha$  will present a higher value.

In addition to the productivity parameter of knowledge, the extent by which an inventor could benefit from this knowledge also depends of its obsolescence and diffusion pattern. Thus,  $\beta_1$  is function of  $l$  because we expect that knowledge diffuse faster from domestic location than from foreign one.

Finally, domestic and foreign knowledge stocks are calculated according to the Perpetual Inventory Method (PIM). Unlike conventional methods, the depreciation rates are estimated previously with the citation function. To illustrate this point, we consider a specific knowledge stock defined with  $l$  and  $L$  country pair characteristics:

$$Know_{L,l,T} = \sum_{t=0}^T \alpha_{L,l,t} \cdot N_{l,t} \cdot e^{-\beta_1(T-t)} \left(1 - e^{-\beta_2(T-t)}\right)$$

We define  $b$  as a base year:

$$Know_{L,l,T} = \sum_{t=b+1}^T \alpha_{L,l,t} \cdot N_{l,t} \cdot e^{-\beta_1(T-t)} \left(1 - e^{-\beta_2(T-t)}\right) + \sum_{t=0}^b \alpha_{L,l,t} \cdot N_{l,t} \cdot e^{-\beta_1(T-t)} \left(1 - e^{-\beta_2(T-t)}\right)$$

To simplify we note thereafter :  $\alpha_{L,l,t} \cdot N_{l,t} = K_{L,l,t}$

$$Know_{L,l,T} = \sum_{t=b+1}^T K_{L,l,t} \cdot e^{-\beta_1(T-t)} \left(1 - e^{-\beta_2(T-t)}\right) + Know_{L,l,T(b)}$$

$Know_{L,l,T(b)}$  is built using the perpetual inventory method (PIM). If we suppose that the growth rate of knowledge stocks  $g$ , is constant over time; we can write  $K_{L,l,b} = (1+g)^b K_{L,l,0}$ . The knowledge stock of the base year  $b$  available at year  $T$  is:



$$Know_{L,l,T(b)} = e^{-\beta_1(T-b)} \left[ \left( \frac{1+g}{(1+g) - e^{-\beta_1}} \right) - \left( \frac{1+g}{(1+g) - e^{-(\beta_1+\beta_2)}} \right) e^{-\beta_2} \right] K_{L,l,b}$$

Finally, the existing knowledge stock available for an inventor in a given year  $T$  with an initial stock for base year  $b$  is given by the following equation:

$$Know_{L,l,T} = \sum_{t=b+1}^T \alpha_{L,l,t} \cdot N_{l,t} \cdot e^{-\beta_1(T-t)} \left( 1 - e^{-\beta_2(T-t)} \right) + e^{-\beta_1(T-b)} \left[ \left( \frac{1+g}{(1+g) - e^{-\beta_1}} \right) - \left( \frac{1+g}{(1+g) - e^{-(\beta_1+\beta_2)}} \right) e^{-\beta_2} \right] K_{L,l,b} \quad (4.15)$$

### 4.5.3 Dealing with different database

The use of patent citation function is not really widespread because of the difficulty raised by dealing with huge patent citation database. To our knowledge, in addition to previous cited and published studies, Bacchiocchi & Montobbio (2009) [13], Hu & Jaffe (2003) [77], Johnson & Popp (2003) [91] have used the patent citation function<sup>10</sup>. On these studies only one have used another database than the USPTO<sup>11</sup>. This study is that realized by Bacchiocchi & Montobbio (2009) whose used the EPO database. We have already emphasis the risk of studying international spillovers with a national database because it induces a positive bias in favor of the national country. In contrast, patent statistics based on EPO data should not be affected by the “home advantage effect” since in almost all cases, EPO patent was first filed to a national office and belong to a patent family with at least two applications. The most widespread database used for studying European citations is the EPO Patent Database ‘Building Blocks’ [188] which concerns only citations made by EPO patents. The problem underlying this database is that only EPO patents are documented. The database provides information on the citing patent but when the patent cited has no EPO equivalence, no information are available (especially no information relating to inventor country are available). Generally authors using this database consider only intra EPO citations and eclipse an important part of cited patents. Furthermore, Criscuolo (2006) [38] has shown that contrary to previous belief, EPO data are also affected by the home advantage affect. Considering these points, and in order to take into account all information available, we choose to combine different patent office data (these data are provided in the

---

<sup>10</sup>There exists another papers using the patent citation function but these papers have not yet been published

<sup>11</sup>USPTO data are provided by the NBER Patent Database thanks to Browyn Hall and her associates, [71]

PATSTAT database which will be presented in the next chapter). Our indicator will be built using patent data taken from both the DE, EPO, FR, UK, JP, US and the WIPO offices, that allow us, contrary to most studies focusing on a single patent office, to conduct a global analysis of innovative activities without the presence of the home advantage affect.

### 4.5.3.1 Advantages of such approach are multiple

First, by considering every national office we avoid national bias. Second, we are able to identify and to refer explicitly to the first application both for the cited and citing patent. For instance, if we consider the use of the EPO database on our patent sample provided in Appendix, we will find that the Patent P12 cites the Patent P22, although the real process of citation in term of knowledge flows is the citation of the Patent P2 by the Patent P1. Using only the EPO database, this flow cannot be investigated. Third, by using national office information, we expand considerably our sample, leading to more accurate results.

### 4.5.3.2 Some difficulties have to be challenged

The first objection to the use such mix-based patent data may result from a problem of consistency between patents taken from different patent offices. The difference in the number of publication in the JPO from the USPTO and the EPO is not necessarily due to differences in innovative activities but it is usually the result of different standards and procedures. In Japan, the unity of an invention is tackled differently than in Europe. While the EPO grants only one patent for any given inventive system, the same invention in Japan could constitute up to 10 different patents. One solution to deal with this issue is implemented by Dechezleprêtre et al. (2008) [39]. The authors examined several international patent families in order to establish the number of patent a particular invention require to be protected in different geographic areas<sup>12</sup>. For example, they find that on average, seven patents protect an invention at the JPO while there are only five at the EPO for protecting the same invention. This means that one EPO patent is equivalent to 1.4 Japanese patents. The country weight coefficients calculating following this methodology are provided in Table 4.3. We will use these parameters for future aggregation of patent arising from different offices.

---

<sup>12</sup>A patent family is a set of inter-related patent applications filed in one or more countries to protect the same invention.

---

Offices	DE	EP	FR	UK	JP	US
Breadth	1.12	1.00	0.99	0.93	0.71	0.96

---

Source: Dechezlepretre et al. (2008)

Table 4.3: Patent office breadth coefficients

Secondly, we have established that USPTO and EPO patent citations well proxy knowledge flows, but what about DPMA, INPI, IPO and JPO patent citations? It seems that national European patent citations are close to the EPO ones. Since most of the EPO applications are second application we just need to look how citations embodied in first application are similar to the EPO citations. Patent P1 and P21 suggest that citations taken from the first national application and from the subsequent EPO application are close each others. Concerning JPO patents, things are more complicated because a large part of patent that not include backward citations (Goto & Motohashi 2007 [55]). Moreover, none studies have investigated the extent to which JPO citations could represent a part of knowledge flows. We will attempt to deal with these issues in the next chapters. In chapter 5, we will investigate the similitude between EPO citations and FR, GE, UK and JPO citations. In chapter 6, we will estimate the explanatory power of the stock of knowledge defined above, i.e. our PCI, on the productivity growth.

## 4.6 Conclusion

In recent years, patent citations have been increasingly used to measure knowledge flows. Specially, since the development of the patent citation function by Jaffe and Caballero in 1993, literature has opened ways to analyze the determinants of patent citations. Recently, Popp (2002) has used patent citations as a proxy of the cumulative property of knowledge into the construction of an input based knowledge indicator. His indicator is based on the assumption that citations to previous patents means that these patents represent a piece of knowledge upon which the new invention is built. By considering only US inventors and USPTO patents Popp only takes into account inter temporal knowledge spillovers. In this chapter, we have extended previous specification in order to take into account inter-national knowledge spillovers. Our specification retained is based on data taken from different national databases in order to avoid the home advantage effect highlighted by Crusciolo (2005) [38].

This present chapter provides a method to construct a knowledge stock upon inter temporal and inter national knowledge flows. Knowledge flows are proxied by patent citations from one inventor

to another inventor and that between patents that could belong to different offices. The resulting indicator measures knowledge available to inventor according to its geographical origin. To justify our choice, in the first part of this chapter, we have extensively discussed the different facets of patents and how they can be used as an indicator of technological change. Literature has validated the use of patent as economic indicator of innovative activities especially when R&D data are missing. But the use of patent citations is still controversial, although literature has recognized the utility of EPO and USPTO patent citation data into the measure of knowledge flows, it has not yet explored the case of national European and Japanese offices. Second part of this paper was dedicated to the theoretical construction of our indicator. This indicator presents several advantages. First, it is based on patent data, available at a very detailed technological classification. Indicator could respond to the need of data at a more disaggregated level. Second, it takes into international knowledge spillovers. Third, indicator's key factor is built on patent productivity that represents technological opportunities embodied, which could respond to the problem of patent heterogeneity.

The next step will be the empirical implementation of this indicator. Before doing this, some upstream works are required. In one hand, there is a huge load of work concerning the construction of the relevant database from the raw PATSTAT database and in a second hand the pattern of European national and Japanese citations need to be detailed. In the next chapter, we will empirically explore the pattern of these citations and we will test their similitude with EPO citations. In the chapter 6, the investigation of patent citations will be endogenously linked to the construction of our indicator. By comparing our indicator with the usual R&D indicators and then by analyzing the relationships between the indicators and the productivity growth empirically, we could assess the validity of our indicator and the utilization of these patent citations into the analysis of knowledge spillovers.

## 4.7 Appendix

### 4.7.1 The modal lab calculi

The modal or most frequent lag in citation is a robust measure of speed to diffusion. The modal lag is the lag at which the citation probability peaks. It is given by the first order condition of the citation function with respect to the lag. Letting  $L=T-t$  and dropping indexes  $K$  and  $k$ , we rewrite equation (4.11) in the following:

$$p = \alpha e^{-\beta_1 L} (1 - e^{-\beta_2 L}) = \alpha \left[ e^{-\beta_1 L} - e^{-(\beta_1 + \beta_2)L} \right]$$

The First Order Condition is such as:

$$\frac{\partial p}{\partial L} = \alpha \left[ (-\beta_1) e^{-\beta_1 L} + (\beta_1 + \beta_2) e^{-(\beta_1 + \beta_2)L} \right] = 0$$

we get

$$\beta_1 e^{-\beta_1 L} = (\beta_1 + \beta_2) e^{-(\beta_1 + \beta_2)L}$$

$$\frac{\beta_1}{\beta_1 + \beta_2} = e^{-\beta_2 L}$$

and thus

$$\log \left[ \frac{\beta_1}{\beta_1 + \beta_2} \right] = -\beta_2 L$$

$$L = -\frac{1}{\beta_2} \log \left[ \frac{\beta_1}{\beta_1 + \beta_2} \right] = -\frac{1}{\beta_2} [\log(\beta_1) - \log(\beta_1 + \beta_2)] = \frac{1}{\beta_2} [\log(\beta_1 + \beta_2) - \log(\beta_1)]$$

$$L = \frac{1}{\beta_2} \log \left[ \frac{\beta_1 + \beta_2}{\beta_1} \right] = \frac{1}{\beta_2} \log \left[ 1 + \frac{\beta_2}{\beta_1} \right] = \frac{1}{\beta_2} \left[ \frac{\beta_2}{\beta_1} - \left( \frac{\beta_2}{\beta_1} \right)^2 + \left( \frac{\beta_2}{\beta_1} \right)^3 - \dots \right]$$

If  $\beta_2$  is very small and  $\frac{\beta_2}{\beta_1} \simeq 0$

$$L_{Mod} \simeq \frac{1}{\beta_2} \frac{\beta_2}{\beta_1} = \frac{1}{\beta_1}$$

#### 4.7.2 The cumulative citation probability calculi

The total number of citations that a patent could received is in theory the cumulative citation probability corresponding to  $L \rightarrow \infty$ . This is given by :

$$C(\infty) = \int_0^{\infty} \alpha e^{-\beta_1 L} (1 - e^{-\beta_2 L}) dL$$

Using linear property of integration:

$$C(\infty) = \alpha \int_0^{\infty} e^{-\beta_1 L} dL - \alpha \int_0^{\infty} e^{-(\beta_1 + \beta_2)L} dL$$

$$C(\infty) = \alpha \left[ -\frac{1}{\beta_1} e^{-\beta_1 L} \right]_0^{\infty} - \alpha \left[ -\frac{1}{\beta_1 + \beta_2} e^{-(\beta_1 + \beta_2)L} \right]_0^{\infty}$$

we get

$$C(\infty) = -\frac{\alpha}{\beta_1} [0 - e^0]_0^{\infty} + \frac{\alpha}{\beta_1 + \beta_2} [0 - e^0]_0^{\infty} = \frac{\alpha}{\beta_1} - \frac{\alpha}{\beta_1 + \beta_2}$$

and thus

$$C(\infty) = \frac{\alpha \beta_2}{\beta_1(\beta_1 + \beta_2)}$$

### 4.7.3 Median lag calculi

To compute the average lag in citation, just multiply the probability of citation by the lag  $L$  and integrate it with respect to a lag of infinity,  $L \rightarrow \infty$ :

$$M(\infty) = \int_0^{\infty} L \times \alpha e^{-\beta_1 L} (1 - e^{-\beta_2 L}) dL$$

Using integration by part with  $v = L$ ,  $v' = 1$ ,  $u' = e^{-\beta_1 L} - e^{-(\beta_1 + \beta_2)L}$  and  $u = -\frac{1}{\beta_1} e^{-\beta_1 L} + \frac{1}{\beta_1 + \beta_2} e^{-(\beta_1 + \beta_2)L}$ , we can rewrite equation (4.13):

$$M(\infty) = \alpha \int_0^{\infty} u' \times v dL = \alpha \left( [uv]_0^{\infty} - \int_0^{\infty} u \times v' dL \right)$$

In one hand

$$[uv]_0^{\infty} = \left[ L \times e^{-\beta_1 L} - e^{-(\beta_1 + \beta_2)L} \right]_0^{\infty}$$

Using the property that  $\lim_{L \rightarrow \infty} \frac{e^L}{L} = \infty$  and its reciprocal  $\lim_{L \rightarrow \infty} \frac{L}{e^L} = 0$

$$[uv]_0^{\infty} = 0 - 0 = 0$$

In the other hand

$$\int_0^{\infty} u \times v' dL = \int_0^{\infty} -\frac{1}{\beta_1} e^{-\beta_1 L} + \frac{1}{\beta_1 + \beta_2} e^{-(\beta_1 + \beta_2)L} dL$$

$$\int_0^{\infty} u \times v' dL = -\frac{1}{\beta_1} \int_0^{\infty} e^{-\beta_1 L} dL + \frac{1}{\beta_1 + \beta_2} \int_0^{\infty} e^{-(\beta_1 + \beta_2)L} dL$$

$$\int_0^{\infty} u \times v' dL = -\frac{1}{\beta_1^2} (0 - 1) + \frac{1}{(\beta_1 + \beta_2)^2} (0 - 1) = \frac{\beta_1^2 - (\beta_1 + \beta_2)^2}{\beta_1^2 (\beta_1 + \beta_2)^2}$$

Finally

$$M(\infty) = \alpha \frac{(\beta_1 + \beta_2)^2 - \beta_1^2}{\beta_1^2 (\beta_1 + \beta_2)^2}$$

However, this average is based on a cumulative “probability” that does not sum to unity, as 4.12 shows. To obtain the mean divide equation (4.14) by 4.12 so that the probability mass is normalized to 1 :

$$L_{mean} = \frac{M(\infty)}{C(\infty)} = \frac{\frac{(\beta_1 + \beta_2)^2 - \beta_1^2}{\beta_1^2 (\beta_1 + \beta_2)^2}}{\frac{\beta_2}{\beta_1 (\beta_1 + \beta_2)}} = \frac{\beta_1^2 + \beta_2^2 + 2\beta_1\beta_2 - \beta_1^2}{\beta_2\beta_1(\beta_1 + \beta_2)}$$

$$\frac{M(\infty)}{C(\infty)} = \frac{\beta_2^2 + 2\beta_1\beta_2}{\beta_2\beta_1(\beta_1 + \beta_2)} = \frac{\beta_2 + 2\beta_1}{\beta_1(\beta_1 + \beta_2)}$$

$$\frac{M(\infty)}{C(\infty)} = \frac{1}{\beta_1} + \frac{1}{(\beta_1 + \beta_2)} \simeq \frac{2}{\beta_1}$$

$$L_{mean} = 2L_{Mod}$$

#### 4.7.4 PIM calculi

The knowledge stock in base year  $b$  can be expressed as the sum of the knowledge input from the base year  $b$ :

$$Know_{T,b} = K_b \cdot e^{-\beta_1(T-b)} \left(1 - e^{-\beta_2(T-b)}\right) + (1+g)^{-1} K_b \cdot e^{-\beta_1(T-b+1)} \left(1 - e^{-\beta_2(T-b+1)}\right) + \dots +$$

$$(1+g)^{-(b-1)} K_b \cdot e^{-\beta_1(T-1)} \left(1 - e^{-\beta_2(T-1)}\right) + (1+g)^{-b} K_b \cdot e^{-T\beta_1} \left(1 - e^{-T\beta_2}\right)$$

Divide by the common input  $K_b$ :

$$\frac{Know_{T,b}}{K_b} = e^{-\beta_1(T-b)} \left(1 - e^{-\beta_2(T-b)}\right) + \frac{e^{-\beta_1(T-b+1)} \left(1 - e^{-\beta_2(T-b+1)}\right)}{(1+g)} + \dots +$$

$$\frac{e^{-\beta_1(T-1)} \left(1 - e^{-\beta_2(T-1)}\right)}{(1+g)^{(b-1)}} + \frac{e^{-T\beta_1} \left(1 - e^{-T\beta_2}\right)}{(1+g)^b}$$

Develop

$$\frac{Know_{T,b}}{K_b} = e^{-\beta_1(T-b)} - e^{-(\beta_1+\beta_2)(T-b)} + \frac{e^{-\beta_1(T-b+1)} - e^{-(\beta_1+\beta_2)(T-b+1)}}{(1+g)} + \dots +$$

$$\frac{e^{-\beta_1(T-1)} - e^{-(\beta_1+\beta_2)(T-1)}}{(1+g)^{(b-1)}} + \frac{e^{-T\beta_1} - e^{-T(\beta_1+\beta_2)}}{(1+g)^b}$$

and simplify

$$\frac{Know_{T,b}}{K_b} = \left( e^{-\beta_1(T-b)} + \frac{e^{-\beta_1(T-b+1)}}{(1+g)} + \dots + \frac{e^{-\beta_1(T-1)}}{(1+g)^{(b-1)}} + \frac{e^{-T\beta_1}}{(1+g)^b} \right) -$$

$$\left( e^{-(\beta_1+\beta_2)(T-b)} + \frac{e^{-(\beta_1+\beta_2)(T-b+1)}}{(1+g)} + \dots + \frac{e^{-(\beta_1+\beta_2)(T-1)}}{(1+g)^{(b-1)}} + \frac{e^{-T(\beta_1+\beta_2)}}{(1+g)^b} \right)$$

Rearrange in order to highlight

$$\frac{Know_{T,b}}{K_b} = e^{-\beta_1(T-b)} \left( 1 + \frac{e^{-\beta_1(1)}}{(1+g)} + \dots + \frac{e^{-\beta_1(b-1)}}{(1+g)^{(b-1)}} + \frac{e^{-b\beta_1}}{(1+g)^b} \right) -$$

$$e^{-(\beta_1+\beta_2)(T-b)} \left( 1 + \frac{e^{-(\beta_1+\beta_2)(1)}}{(1+g)} + \dots + \frac{e^{-(\beta_1+\beta_2)(b-1)}}{(1+g)^{(b-1)}} + \frac{e^{-b(\beta_1+\beta_2)}}{(1+g)^b} \right)$$

Finally we organize the right hand side



$$\frac{Know_{T,b}}{K_b} = e^{-\beta_1(T-b)} \left( \frac{1 - \frac{e^{-(b+1)\beta_1}}{(1+g)^{(b+1)}}}{1 - \frac{e^{-\beta_1}}{(1+g)}} \right) - e^{-(\beta_1+\beta_2)(T-b)} \left( \frac{1 - \frac{e^{-(b+1)(\beta_1+\beta_2)}}{(1+g)^{(b+1)}}}{1 - \frac{e^{-(\beta_1+\beta_2)}}{(1+g)}} \right)$$

And rearrange to get the knowledge stock at the base point:

$$\frac{Know_{T,b}}{K_b} = e^{-\beta_1(T-b)} \left[ \left( \frac{1+g}{(1+g) - e^{-\beta_1}} \right) - \left( 1 - e^{-\beta_2(T-b)} \right) \left( \frac{1+g}{(1+g) - e^{-(\beta_1+\beta_2)}} \right) \right]$$

4.7.5 Patents figures

(19) RÉPUBLIQUE FRANÇAISE  
 INSTITUT NATIONAL  
 DE LA PROPRIÉTÉ INDUSTRIELLE  
 PARIS

(11) N° de publication : **2 924 141**  
 (à n'utiliser que pour les  
 commandes de reproduction)

(21) N° d'enregistrement national : **07 59348**

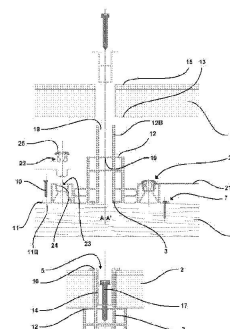
(51) Int Cl<sup>8</sup> : **E 04 D 13/18 (2006.01), H 01 L 31/042**

(12) **DEMANDE DE BREVET D'INVENTION** **A1**

<p>(22) Date de dépôt : 27.11.07.</p> <p>(30) Priorité :</p> <p>(43) Date de mise à la disposition du public de la demande : 29.05.09 Bulletin 09/22.</p> <p>(56) Liste des documents cités dans le rapport de recherche préliminaire : <i>Se reporter à la fin du présent fascicule</i></p> <p>(60) Références à d'autres documents nationaux apparentés :</p>	<p>(71) Demandeur(s) : <i>QUALITELEC Société à responsabilité limitée — FR.</i></p> <p>(72) Inventeur(s) : CANET PATRICE et DELPEYROU JOSE.</p> <p>(73) Titulaire(s) :</p> <p>(74) Mandataire(s) : CABINET BREV&amp;SUD.</p>
---	--

(54) DISPOSITIF SUPPORT DE PANNEAU DE COUVERTURE DE TOITURE.

(57) La présente invention concerne un dispositif (1) support d'au moins un panneau (2) de couverture de toiture et une telle toiture. Ce dispositif (1) est constitué d'au moins un rail (3) pourvu en partie haute (4) de moyens (5) supports dudit panneau (2) et en partie basse (6) de moyens (7) de fixation à au moins un élément (8) de ladite toiture, caractérisé par le fait que ledit rail (3) comprend des moyens (20) de blocage d'un film (21) de protection de ladite toiture.



FR 2 924 141 - A1



Figure 4.7: P1; A first patent application at the INPI (FR)

RÉPUBLIQUE FRANÇAISE

2924141



**RAPPORT DE RECHERCHE  
PRÉLIMINAIRE**

N° d'enregistrement  
national



établi sur la base des dernières revendications  
déposées avant le commencement de la recherche

FA 703531  
FR 0759348

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
X	DE 20 2007 005995 U1 (IDEEMATIC DEUTSCHLAND GMBH [DE]) 28 juin 2007 (2007-06-28) * alinéas [0002], [0027] - [0037]; revendication 1; figures 1-6 *	1,2,8	E04D13/18 H01L31/042
X	WO 02/16707 A (POWERLIGHT CORP [US]; DINWOODIE THOMAS L [US]; JACKAWAY ADAM D [US]; M) 28 février 2002 (2002-02-28) * page 3, ligne 30 - page 8, ligne 9; figures 3-11 *	1,2,8	
A	EP 1 777 357 A (MITSUBISHI ELECTRIC CORP [JP]) 25 avril 2007 (2007-04-25) * le document en entier *	1,8	
A	EP 0 858 115 A (CANON KK [JP]) 12 août 1998 (1998-08-12) * le document en entier *	1,8	
A	EP 0 828 035 A (CANON KK [JP]) 11 mars 1998 (1998-03-11)		
			DOMAINES TECHNIQUES RECHERCHÉS (IPC)
			H01L F24J
Date d'achèvement de la recherche		Examineur	
15 juillet 2008		Visentin, Alberto	
<p>CATÉGORIE DES DOCUMENTS CITÉS</p> <p>X : particulièrement pertinent à lui seul                      Y : particulièrement pertinent en combinaison avec un                      autre document de la même catégorie                      A : arrière-plan technologique                      O : divulgation non-écrite                      P : document intercalaire</p> <p>T : théorie ou principe à la base de l'invention                      E : document de brevet bénéficiant d'une date antérieure                      à la date de dépôt et qui n'a été publié qu'à cette date                      de dépôt ou qu'à une date postérieure.                      D : cité dans la demande                      L : cité pour d'autres raisons                      &amp; : membre de la même famille, document correspondant</p>			

EPO FORM 1503 12.98 (P4C14)

Figure 4.8: P1 search report ; Citations made by the patent P1

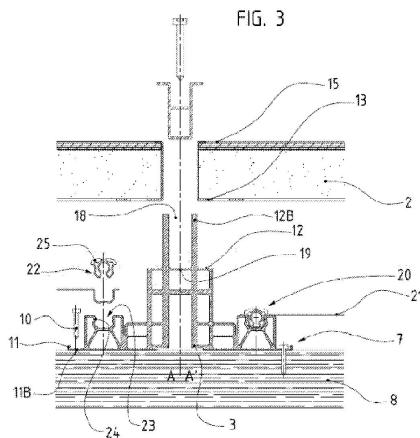
<p>(19)  <b>Europäisches Patentamt</b> <b>European Patent Office</b> <b>Office européen des brevets</b></p>	
<p>(11) <b>EP 2 065 944 A1</b></p>	
<p>(12) <b>DEMANDE DE BREVET EUROPEEN</b></p>	
<p>(43) Date de publication: <b>03.06.2009 Bulletin 2009/23</b></p>	<p>(51) Int Cl.: <b>H01L 31/048<sup>(2006.01)</sup> H01L 31/042<sup>(2006.01)</sup></b> <b>E04D 13/18<sup>(2006.01)</sup></b></p>
<p>(21) Numéro de dépôt: <b>08300284.0</b></p>	
<p>(22) Date de dépôt: <b>27.11.2008</b></p>	
<p>(84) Etats contractants désignés: <b>AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MT NL NO PL PT RO SE SI SK TR</b> Etats d'extension désignés: <b>AL BA MK RS</b></p>	<p>(72) Inventeurs: • <b>Canet, Patrice</b> <b>11570, PALAJA (FR)</b> • <b>Delpeyrou, José</b> <b>11000, CARCASSONNE (FR)</b></p>
<p>(30) Priorité: <b>27.11.2007 FR 0759348</b></p>	
<p>(74) Mandataire: <b>Rhein, Alain</b> <b>Cabinet BREV&amp;SUD</b> <b>2460, avenue Albert Einstein</b> <b>34000 Montpellier (FR)</b></p>	
<p>(71) Demandeur: <b>Graphisol (Société à Responsabilité Limitée à Associé unique)</b> <b>11800 Laure-Minervois (FR)</b></p>	

(54) **Dispositif support de panneaux de couverture de toiture**

(57) La présente invention concerne un dispositif (1) support d'au moins un panneau (2) de couverture de toiture et une telle toiture. Ce dispositif (1) est constitué d'au moins un rail (3) pourvu en partie haute (4) de moyens (5) supports dudit panneau (2) et en partie basse (6) de moyens (7) de fixation à au moins un élément (8) de ladite toiture, ledit rail (3) comprenant encore des moyens (20) de blocage d'un film (21) de protection de ladite toi-

ture, lesdits moyens de blocage (20) se situant supérieurement auxdits moyens de fixation (7) de sorte que ces derniers soient couverts par ledit film (21), caractérisé par le fait que lesdits moyens de blocage (20) comprennent au moins une partie mâle (22) de forme complémentaire avec au moins une partie femelle (23) de manière à coopérer par emboîtement, bloquant une extrémité dudit film (21).

EP 2 065 944 A1



Printed by Jouve, 75001 PARIS (FR)

Figure 4.9: P12; A second application of P1 at the EPO

EP 2 065 944 A1



RAPPORT DE RECHERCHE EUROPEENNE

Numéro de la demande  
EP 08 30 0284

DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	Revendication concernée	CLASSEMENT DE LA DEMANDE (IPC)
A	DE 20 2007 005995 U1 (IDEEMATIC DEUTSCHLAND GMBH [DE]) 28 juin 2007 (2007-06-28) * alinéas [0002], [0027] - [0037]; revendication 1; figures 1-6 *	1,7,8	INV. H01L31/048 H01L31/042 E04D13/18
A	WO 02/16707 A (POWERLIGHT CORP [US]; DINWOODIE THOMAS L [US]; JACKAWAY ADAM D [US]; M) 28 février 2002 (2002-02-28) * page 3, ligne 30 - page 8, ligne 9; figures 3-11 *	1,7,8	
A	EP 1 777 357 A (MITSUBISHI ELECTRIC CORP [JP]) 25 avril 2007 (2007-04-25) * le document en entier *	1,7,8	
A	EP 0 858 115 A (CANON KK [JP]) 12 août 1998 (1998-08-12) * le document en entier *	1,7,8	
A	EP 0 828 035 A (CANON KK [JP]) 11 mars 1998 (1998-03-11)		DOMAINES TECHNIQUES RECHERCHES (IPC) H01L F24J
Le présent rapport a été établi pour toutes les revendications			
Lieu de la recherche Berlin		Date d'achèvement de la recherche 11 mars 2009	Examineur Visentin, Alberto
<p>CATEGORIE DES DOCUMENTS CITES</p> <p>X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : arrière-plan technologique O : divulgation non-écrite P : document intermédiaire</p> <p>T : théorie ou principe à la base de l'invention E : document de brevet antérieur, mais publié à la date de dépôt ou après cette date D : cité dans la demande L : cité pour d'autres raisons &amp; : membre de la même famille, document correspondant</p>			

EPO FORM 1503 (03.02) (FR) (A02)

Figure 4.10: P12 search report ; Citations made by the patent P12

(19) 日本国特許庁 (J P)      (12) 公開特許公報 (A)      (11) 特許出願公開番号  
**特開平10-219950**  
 (43) 公開日 平成10年(1998) 8月18日

(51) Int.Cl. <sup>8</sup>	識別記号	F I
E 0 4 D 13/18		E 0 4 D 13/18
	3/40	3/40
		W
H 0 1 L 31/042		H 0 1 L 31/04
		R

審査請求 未請求 請求項の数36 O L (全 14 頁)

(21) 出願番号	特願平9-22613	(71) 出願人	000001007 キヤノン株式会社 東京都大田区下丸子3丁目30番2号
(22) 出願日	平成9年(1997) 2月5日	(72) 発明者	大塚 崇志 東京都大田区下丸子3丁目30番2号キヤノン株式会社内
		(72) 発明者	森 昌宏 東京都大田区下丸子3丁目30番2号キヤノン株式会社内
		(72) 発明者	三村 敏彦 東京都大田区下丸子3丁目30番2号キヤノン株式会社内
		(74) 代理人	弁理士 丸島 儀一

最終頁に続く

(54) 【発明の名称】 屋根材及びその施工方法

(57) 【要約】

【課題】 設置施工が容易で、外観と電気安全性に優れた屋根材を提供する。

【解決手段】 本発明の屋根材は、太陽電池素子と金属補強材を持つ太陽電池一体型屋根材が、屋根に固定される屋根材であり、太陽電池一体型屋根材の下側に金属部材を設け、あるいは、隣り合う太陽電池一体型屋根材の隣接部に沿って金属部材を設け、該金属部材と、複数の太陽電池一体型屋根材の金属補強材とが電氣的に導通され、かつ金属部材が電氣的に接地されていることを特徴とする屋根材である。

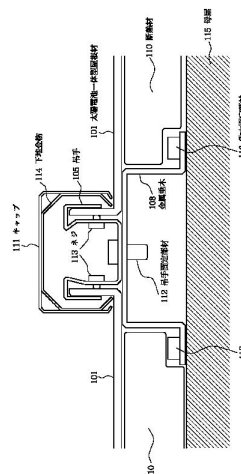


Figure 4.11: P2; A first application at the JPO

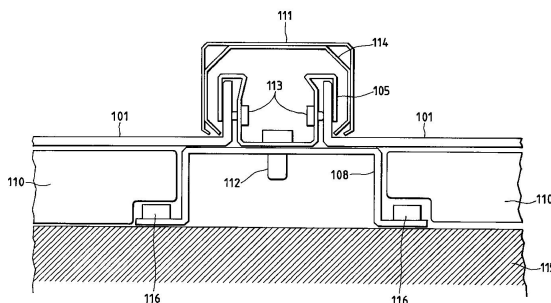
	Europäisches Patentamt European Patent Office Office européen des brevets	
(19)		(11) <b>EP 0 858 115 A2</b>
(12) <b>EUROPEAN PATENT APPLICATION</b>		
(43) Date of publication: 12.08.1998 Bulletin 1998/33	(51) Int. Cl. <sup>6</sup> : <b>H01L 31/042</b> , H01L 31/048, E04D 13/18	
(21) Application number: <b>98101909.4</b>		
(22) Date of filing: <b>04.02.1998</b>		
(84) Designated Contracting States: <b>AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE</b> Designated Extension States: <b>AL LT LV MK RO SI</b>	<ul style="list-style-type: none"> <li>• Fukae, Kimitoshi Ohta-ku, Tokyo (JP)</li> <li>• Mimura, Toshihiko Ohta-ku, Tokyo (JP)</li> <li>• Mori, Masahiro Ohta-ku, Tokyo (JP)</li> </ul>	
(30) Priority: <b>05.02.1997 JP 22613/97</b>	(74) Representative: <b>Pellmann, Hans-Bernd, Dipl.-Ing. et al</b> Patentanwaltsbüro Tiedtke-Bühling-Kinne & Partner Bavariaring 4 80336 München (DE)	
(71) Applicant: <b>CANON KABUSHIKI KAISHA</b> Tokyo (JP)		
(72) Inventors: • Ohtsuka, Takashi Ohta-ku, Tokyo (JP)		

(54) **Roof member comprising a solar cell and mounting method thereof**

(57) Roof members of the present invention are those fixed to the roof, each roof member being a combination solar cell and roof member having a solar cell element and a metal reinforcing member, wherein a metal member is provided below the combination solar cell and roof member or a metal member is provided along an adjacent portion between adjacent combination solar cell and roof members, wherein the metal

member is electrically conductive to metal reinforcing members of plural combination solar cell and roof members and wherein the metal member is electrically grounded. Provided based on this structure are the roof members easy to install and excellent in the external view and electric safety.

FIG. 1



EP 0 858 115 A2

Printed by Xerox (UK) Business Services  
2.16.3/3.4

Figure 4.12: P22; A second application of P2 at the EPO

## Chapter 5

# Statistical Lessons from the PATSTAT database

The EPO holds one of the largest digital repositories of public knowledge in the world  
[D. Guellec]

### 5.1 Introduction

Knowledge is one of the most important factors in economic growth, but many difficulties had hindered the measurement of knowledge stocks and flows. Previous chapter has assessed the contribution of patents as an indicator of innovative activity, but at the exception of USPTO and the EPO Patent Data<sup>1</sup>, the lack of user friendly databases has kept researchers away from exploiting this rich and valuable data mine. Moreover, the presence of the “home advantage effect” identified in these databases restricts the range of analysis to domestic or regional applicants. This effect makes the study of international aspects of knowledge problematic, excluding the possibility of measuring international knowledge spillovers.

Fortunately, the EPO has recently launched the construction of a world patent database called Patstat that covers more than 81 patent offices. In this chapter, we take advantage of the opportunity provided by this database, offering such unique data coverage, that make possible the construction of a sub-Patstat database dedicated to the analysis and the measurement of international knowledge flows

---

<sup>1</sup>USPTO and EPO are respectively the US and the European patent office.



between the G5 inventors.

The analytical framework used follows the patent citation's literature discussed in the previous chapter. This exercise is all the more interesting since Patstat data have not been much exploited because they became available only recently, per consequence it allows the presentation of several interesting statistical results that have never been presented before. To clarify, there exists a literature gap concerning practices in patent offices, other than the EPO and the USPTO, such as in the DPMA, the INPI, the IPO or at the JPO<sup>2</sup>. Here we propose to fill this gap by a quantitative analysis that goes with the qualitative review of patent procedures done in the previous chapter.

Furthermore, understanding patent citations at a European national level becomes particularly relevant with the arrival of national data with easier access (Patstat database), especially since almost all EPO patent applications are subsequent applications from national prior applications. But our analysis is not limited to European offices, we also investigate the JPO citations pattern. The present analysis surveys the different facets of patent citations : how many citations contains a search report, what is the origin of the cited patents and to what extend the search reports are similar between offices considered. For the last point, we investigate how citations are similar between the national patent application (filed at the DPMA, INPI, IPO or the JPO) and the EPO subsequent one.

First, the results show that the number of citations per patent is increasing over time for national offices whereas it is constant for regional offices. Second, patent citations suffer from a home bias toward patents belonging to the same office. And finally, patent search reports are highly similar between INPI, JPO and EPO, meaning that these citations can be useful into the measurement of international knowledge flows. Concerning the DPMA and IPO citations, further analysis need to be carried out. Before turning to the reminder of the chapter, we want to caution readers that the extraction of data from the Patstat Database is not a simple matter because of the large size of the base that requires expertises in relational databases and SQL language<sup>3</sup> to manage them.

The chapter is organized as follows. First, we will present in detail the Patstat Database (section 5.2). The next section will shed some light on the Patstat data management and the different steps toward the construction of a sub-database from which citation data are extracted (section 5.3). Finally, the last section will be dedicated to the quantitative analysis of citation between the different offices (section 5.4).

---

<sup>2</sup>DPMA, INPI, IPO and JPO are respectively the German, the French, the UK and the Japan patent office.

<sup>3</sup>SQL (Structured Query Language) is a database computer language designed for managing data in relational database management systems (RDBMS), and originally based upon Relational Algebra.

### 5.2 What is PATSTAT ?

The European Patent Office has, on behalf of the OECD Task-force on Patent Statistics, prepared a database designed to assist in statistical research into patent information. This database is called the EPO Worldwide Patent Statistical Database, also known as PATSTAT. It is an off-line database distributed on DVDs in CSV<sup>4</sup> format from the Vienna sub-office of the EPO by the Publication department. The first complete version was issued in April 2006 and has been updated twice a year. It was, until October 2008, available to all users committing to non commercial use and no further dissemination of the data. Since October 2008, the EPO has revised the distribution policy of its database: PATSTAT is now available without restrictions to anybody who wishes to purchase it. It was developed specifically with the needs of policy makers, academics, analysts and Intellectual Property institutions in mind. Researchers working on patent statistics have previously had to assemble data sets from various and disparate sources and were obliged to perform extensive "cleaning" of the data at considerable cost and time. The PATSTAT data set addresses these issues, efficiently harmonizing data, resolving issues over family members and addressing problems like applications from one applicant appearing under several different names. The database also contains related information on citations, procedural information and legal status, which are all of interest to statisticians, Oddo (2008) [130]. PATSTAT is unique in that it includes patent data from 81 offices world-wide (see Table 5.7), since beginning of the 20th century for certain offices and post-grant data from about 40 offices. The September 2008 version contains almost 60 millions of patent documents<sup>5</sup>. But such coverage involves high sizes -the indexed database sizes about 100 Gg- and the relational structure requires SQL knowledge to make queries, per consequence Patstat is neither a "plug & play" product nor an easy access database.

**Variables and framework** Much of the raw data are extracted from: the EPO's master bibliographic database DocDB, the EPO Patent Information Resource; the PRS, the Patent register for legal data; and from the EPASYS, database for EP patent grant procedure data. In addition to the basic bibliometric and legal data, the database also includes description and harmonized citation data. The database provides all information relating to patent, namely :

- Application information (dates, numbers...),

---

<sup>4</sup>Comma-Separated Values

<sup>5</sup>There are exactly 61,497,371 patent applications documented in the Sept 08 version

- Inventors & Applicants information (name, address...),
- Priorities information,
- IPC classes information & National patent classes' information,
- Publications information,
- References (citations) information.

In order to be thorough and to present the names of the database variables we will use thereafter, all elements included in the Patstat database are presented in the Appendix, Table 5.8. For a more detailed technical review of the database PASTAT, the reader may refer to Robinson & Lunga (2008) [148]. In the following sections we will keep the names of variables given by the EPO for our purposes.

The PATSTAT database is constructed with a relational structure, containing 15 different tables (see the physical model in Appendix, Figure 5.12). Tables are linked together with an unique surrogate key. The structure of the database can be easily understood with Figure 5.1. The tables are centralized around the *application* table and are linked according to the application number except the *citations* table that is linked to the *publication* table with the publication number.

An example of such relational structure is extracted from Patstat and is presented in Table 5.2. Starting from the central table *tls201*, all relevant information can be found. This application is an EPO application (Application Authority = EP) but is not a prior application, the priority number is provided in the *prior application* table (table *tls204*). The search report is published 18 months after the application (publication kind A1 in table *tls212*) and contains references to previous documents (table *tls211*). The technological field is identified according to the IPC classes in table *tls209*, and finally, information about inventor(s) and applicant(s) can be found in table *tls206* & *tls207*. In the present case, the patent was developed by an US inventor and applied by “Telcordia Technologies Inc”.

### 5.3 The sub-database construction

This section presents in detail the technical issues and the different steps toward the construction of our own database (we also refer to it as the 'sub-Patstat database'). The database we create is almost exclusively directed by the specification of the citation function presented in the previous chapter. Nevertheless, in order to proceed to statistical analysis, the data in our database are selected with less restrictive conditions than those applied to the citation function. Because we want to study citations

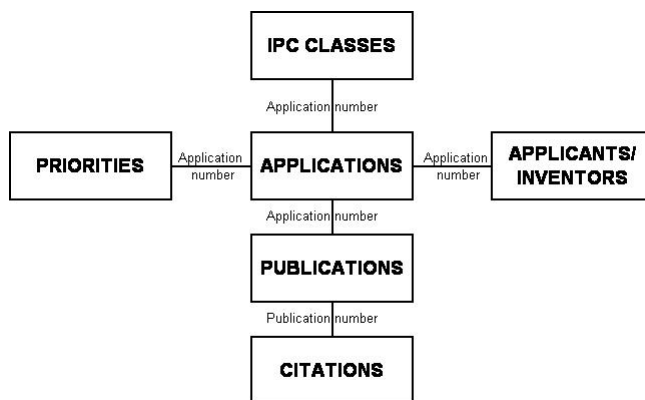


Figure 5.1: Summary of the conceptual model

between different cohorts, we will create two main tables, named *citing* and *cited* tables. The *citing* table corresponds to all first patent applications made in the major patent offices (DPMA, EPO, INPI, IPO, JPO, USPTO, WIPO<sup>6</sup>) and the *cited* table embodies all patents published with at least one patent of the patent family published in the previous selected offices.

### 5.3.1 The Citing Table

The *citing* table embodies information about patents which make citations. For the construction of this table we retain only the first filing, i.e. an application that do not claime the priority of any other application. A first filing can be identified by its absence in the table *table204* (i.e. in the absence of the application number in the field: *appln\_id* of *tls204* table). We collect all patents applied in the DPMA, EPO, INPI, IPO, JPO and USPTO offices after 1970 according to the following points:

- First, we exclude utility models and design applications. Utility models have a shorter duration than regular patents and do not require the same inventive step. Registered designs protect only the appearance of products.
- Second, we collect information on the inventor’s residence country. If an application has more than one inventor, the application is assigned to the first inventor, thus avoiding double counting. This could cause some problems if the application is jointly invented by inventors residing in different countries<sup>7</sup>. If the first inventor has no country code assigned in the database we will use code of subsequent inventors.

---

<sup>6</sup>WIPO is the World Intellectual Property Organization and manages the PCT applications

<sup>7</sup>Nevertheless this phenomena is not widespread and concerns less than 6% of patents

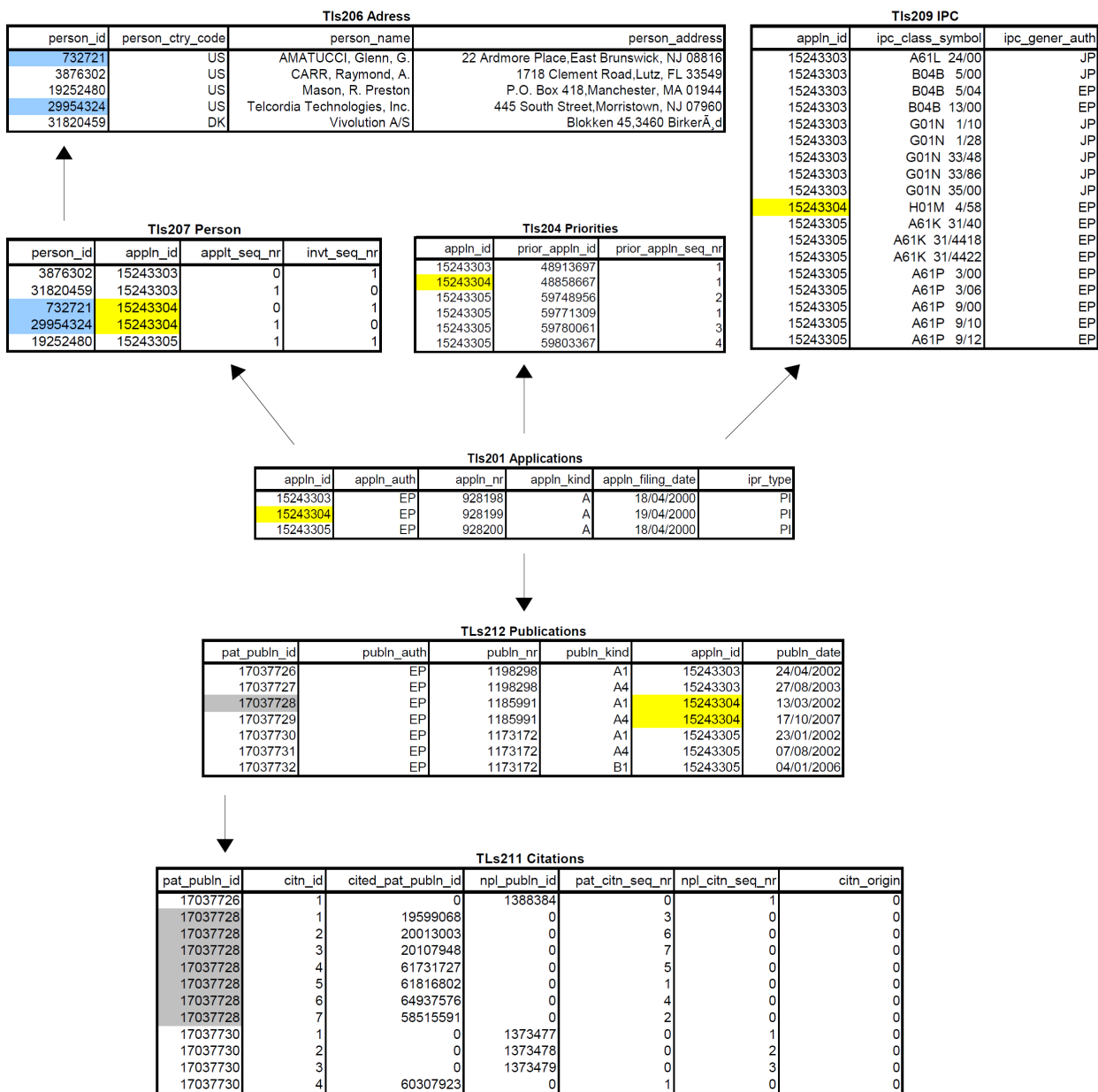


Figure 5.2: An example extracted from the PASTATS database

- Third and about the technological classification, we use IPC data in order to assign patent to a particular technology field (we will come back on this point in the next chapter).

The summary of these requests are provided in Figure 5.3. Finally, the *Citing* table contains 15 200 097 patents applied in the six offices from 1970 to 2008. Table 5.1 presents descriptive statistics related to the *Citing* Table, but only over the period 1978-2006. We reduce the analysis period to 1978-2006 in order to have comparative statistics between offices (EPO was created in 1978) and because some patent applications are not reported after 2006 in the database. The main drawback is the percentage of missing country code for the UK and the JPO patents that are respectively about 76 % and 98.5 % (the column “ratio” is the ratio between the second and first column; that holds for next tables). The over-representation of Japan applications is due to the particularity of Japanese patent procedure: a high rate of application is never published. For instance, in 1999 there were 357,485 JPO patent applications, 117,032 were examined and only 28,966 granted (Goto & Motohashi 2007 [55]).

Office	application	application with ctry code	ratio
DPMA	1 006 489	881 931	87.62%
EPO	160 013	108 427	67.76%
INPI	339 270	324 071	95.52%
IPO	528 734	127 936	24.20%
JPO	8 424 689	124 155	1.47%
USPTO	2 934 421	1 917 465	65.34%

Table 5.1: Citing table statistics - First patent application (1978-2006)

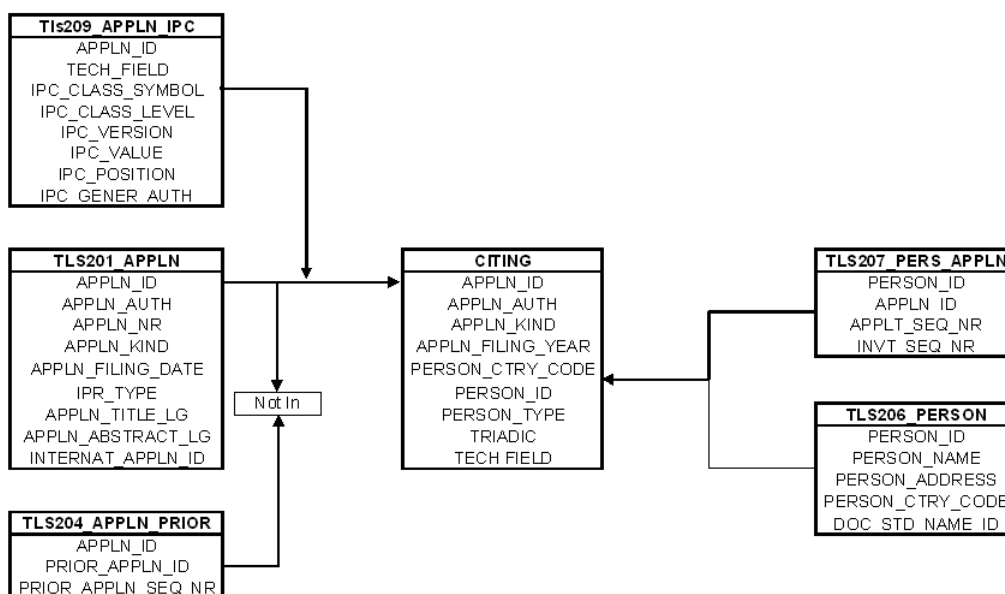


Figure 5.3: The Citing Table

### 5.3.2 The Citing publication Table

The second step is the inclusion of information related to the patent publication. We create a new table based on the *Citing* and the *Tls211* tables. We link the application number of the *Citing* table to the publication number of *tls212* table. In most cases, it gives several links because offices publish documents at more than one procedural step during the processing of an application. These documents can either be : the search report, the patent specification, the amended claims... Previous documents are each defined by a unique publication number, but all are linked to the same application number. To avoid multiple counting we retain only one publication number per patent application number. Not all application numbers have a correspondence with a publication number but all publication numbers have one with an application number. We retain only the publication number which corresponds to the documents that are likely to provide information on patent citations<sup>8</sup>. These publications are in most cases, the search report, with the main exception that USPTO provides citation information only once the patent is granted. Thus for USPTO patent application, we retain only the patent specification

<sup>8</sup>For the DPMA, we retain patent publications which correspond to one of the following patent publication DocDB code: 'A1', 'A5', 'A8', 'B1', 'B2', 'B3' or 'B4'; For the INPI : 'A1', 'A2' or 'A3'; For the EPO : 'A1', 'A2'; For the USPTO : 'A', 'B2', 'B1', 'E', 'E1', 'H'; For the IPO : 'A'; For the JPO : 'B2', 'B' and For the WIPO : 'A1' or 'A2'. Patent Publication Kind Codes are accessible on line on the EPO website (<http://www.epo.org/patents/patent-information/raw-data/useful-tables.html>).

(this issue is detailed in the previous chapter).

Figure 5.5 resumes procedures followed for the creation of the *Citing\_publn* table. This table contains 6 204 213 patents applied between 1970 and 2006 and published by the seven major offices (the former six plus the WIPO).

Table 5.2 presents descriptive statistics related to the *Citing\_publn* according to patent application years. The number of publication without country code is substantively lower than those reported in Table 5.1 at the exception of JPO patents ratio which is constant (1.47 % & 1.55 %).

First, it seems that patents, that are subsequently published, contain higher quality data, at least considering information on inventors. Second, the number of publications is inferior to the number of applications: some applications never get published or never get granted. Since the search report is published at different steps of patent procedures -that vary according to patent office- the signification of the difference between the number of applications and the number of publications depends on them. For the USPTO, this difference is mainly explained by the fact that we retain only granted patents. In the USPTO, grant procedure can take a long time, therefore, a part of patents applied after 2000 are not published in 2006. For the JPO and the IPO, the difference is mainly due to the share of applicants that do not request a subsequent examination. In these offices, applicants have to request a first examination leading up to the patent publication (requests need to be done within 12 months after the first filing for the IPO and 3 years for the JPO). Per consequence a large part of applicant do not request for subsequent examination and the difference between the number of publications and applications for the IPO and the JPO are important: only 25% of all applications are published.

The creation of this table provides interesting statistical information that we have already used in the previous chapter, namely: the number of first patent applications per office and the percentage of inventors who first file in their domestic offices (see Table 4.1). Figure 5.4 breaks down these results into the different offices in a dynamic way. As it has been already indicated, there is a strong dominance of domestic routes for all national offices. For the DPMA, the INPI and the JPO, almost all prior applications are made by domestic inventors, the lowest rate is for the INPI, where only 94% of all inventors who first file their patents are French. We find similar results for the IPO and the USPTO, but at a lower level, around 80%. The common result in all the offices is the reduction in the proportion of national inventors in their own offices between 1990 and 2000. This reduction ranges from 1 point for the JPO and DPMA to 8 points for the USPTO. Concerning regional offices, there is a predominance of German inventors representing one third of all inventors who first file at the EPO,



and one half at the WIPO. Finally, in addition to the G5 inventors, we have added the category 'OT' corresponding to all patents applied by inventors located in other countries than in the G5. We are witnessing, in these offices, an increase in the proportion of inventors located in other countries than the G5 (15 points at the EPO and 7 points at the WIPO).

In order to be symmetric with the results presented in the previous chapter, we provide figures (see in Appendix figure 5.13) corresponding to the first line of Chapter 4 Table 4.1, i.e. corresponding to “the proportion of inventors who first file at their national offices”. Figure 5.13 breaks down these statistics between the different selected offices where inventors first file their inventions. This figure highlights the dominance of the USPTO, nearly two third of inventors located in other countries than in the G5 first file their inventions in the USPTO. This position is strengthened in the recent years with 3 more points between 1990 and 2000. The proportion of these inventors, OT, also increases in the EPO, reaching the second place. Regarding the G5 inventors, all domestic offices correspond to their first choice. The remark made in the previous chapter on JPO data quality still holds for this figure. Subsequently, to address this issue, the Japan country code will be directly associated to inventors who first file their inventions to the JPO. Around 99% of patents which are first filed in the JPO are developed by Japanese inventors, per consequence this approximation does not induce serious bias.

Office	publication	publication with ctry code	ratio
DPMA	889 153	858 102	96.51%
EPO	99 844	99 816	99.97%
INPI	332 434	321 517	96.72%
IPO	133 084	110 631	83.13%
JPO	2 117 400	32 912	1.55%
USPTO	1 557 972	1 556 042	99.88%

Table 5.2: Citing publication table statistics (1978-2006)

### 5.3.3 The Citing citation Table

Once this table built, next step was to develop a table with citations data, the *citing\_citation* table. This task was done thanks to the table *tls12* that links citing and cited patents. Citing patents are those presented in table *citing\_publn* and are identified in table *tls212* according to the publication number. Once again, a patent can make multiple citations or, on the opposite, it can make none. We exclude citations made to non patent literature (NPL). The summary of these requests are provided in Figure 5.5 and descriptive statistics in Table 5.3. The first column of Table 5.3 presents the number of

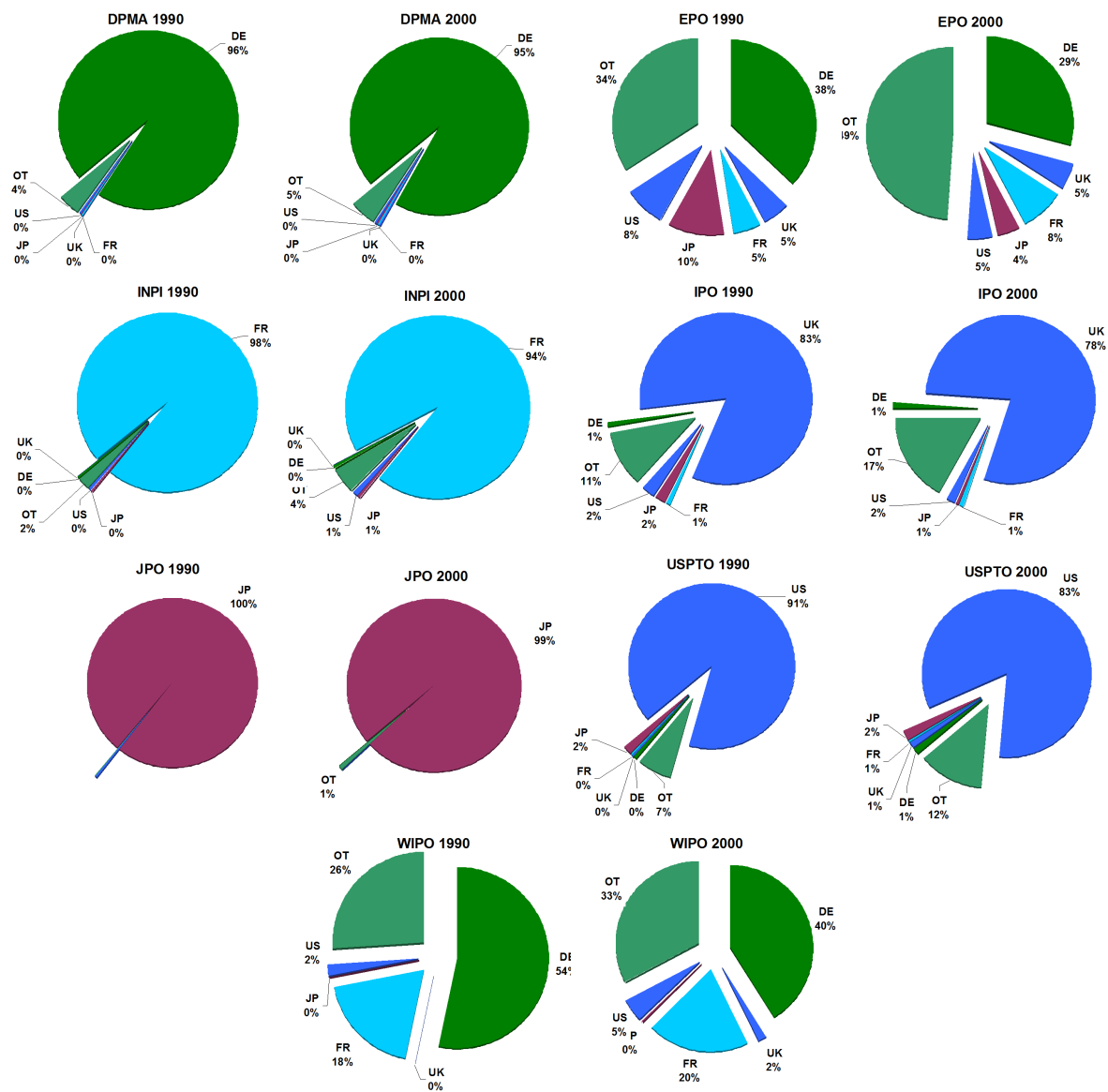


Figure 5.4: Inventors routes

citations made by all patents contained in our sample, i.e. included in table *citing\_publn*. The second column is the number of patents that encompass at least one backward citation. For instance, on 99 844 EPO documents that are likely to cite previous patent, 96 623 do it and totalise 455 582 citations. Not all patents make citations to previous patents, but almost all of them do, at the exception of DPMA and JPO patents with respectively only 55.80 % and 22.52 % of patents provide citations to previous patents. This result is quite unexpected for DPMA patents but when we compare the number of patent citations with the number of patent applications, we find a value in line with previous studies (see next section). Finally, the ratio of patent documents which includes information on the country of the inventor is once again rising.

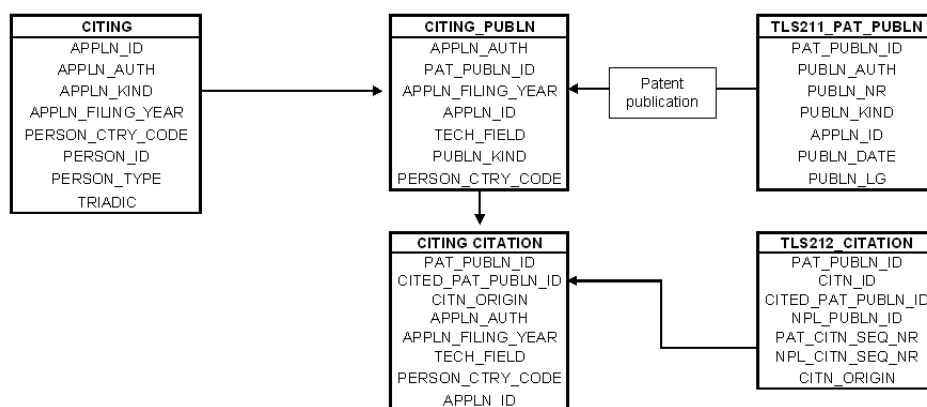


Figure 5.5: The Citing Publication and the Citing Citation Tables

Office	citation	publication with references	publication with references and ctry code	ratio	rate of patent with citations
DPMA	2 062 456	496 124	492 899	99.35%	55.80%
EPO	455 582	96 347	96 323	99.98%	96.50%
INPI	1 258 571	290 167	282 591	97.39%	87.29%
IPO	397 693	96 623	94 738	98.05%	72.60%
JPO	1 130 054	476 779	4 939	1.04%	22.52%
USPTO	21 411 981	1 532 581	1 532 082	99.97%	98.37%

Table 5.3: Citing citation table statistics (1978-2006)

Before turning to the table dedicated to the cohort of patents that are likely to be cited, we provide in Figure 5.6 the breakdown by year of application of all the statistics related to the citing cohort. 'Application' bars correspond to the values given in table *Citing*, 'Publication' corresponds to the *Citing publn* values and 'Citation' is taken from *citing\_citation* and refers to the number of documents including information on patent citations, i.e. including at least one citation. Concerning the European

offices, relations between all variables remained stable over the period, with the exception of IPO in which the citation data are available only since 1982. Regarding the USPTO, there is a divergence between the number of patent applications and the number of publications in the last years. Remember that for USPTO we keep patent specification because citation data are available only once the patent is granted. This means that some patents for which applications were submitted in the last years have not been granted yet and are therefore not in the database. Concerning the JPO, citation data are not available after 1996 (this issue has been documented in the EPO document provided with the PATSTAT database, “European Patent Office – World Wide Database Coverage Report January 2008” [135]). This limitation will reduce our sample for the JPO patents but not in a serious way since we still have 18 years covered.

### 5.3.4 The Cited Table

Once the first task completed, we turn into the construction of the table designed to encompass patents that are likely to be cited by the citing cohort. Here, we start directly with the publication table *Tls211*. We begin with patents that have not claimed priorities. Secondly, we insert patents that are subsequent publications. For patents belonging to the last group, we assign to them the information related to their prior applications. We applied criteria of selection close to those we have applied for the *citing* table, namely :

- not patents with a publication year before 1970 and after 2006,
- at least one of the applications (prior or the second) has to be made either in the G5 patent offices, either in the WIPO or in the EPO,
- and finally, we also exclude Utility models and Design applications.

We control that the persons address is the same for the prior and for the subsequent application. The date that matters for the cited patent is the date at which information embodied in these patents are publicly available. Information concerning the invention protected by the patent are made public when the patent is published (generally 18 months after the application), thus we retain the date of publication of these patents. The summary of these requests are provided in Figure 5.7.

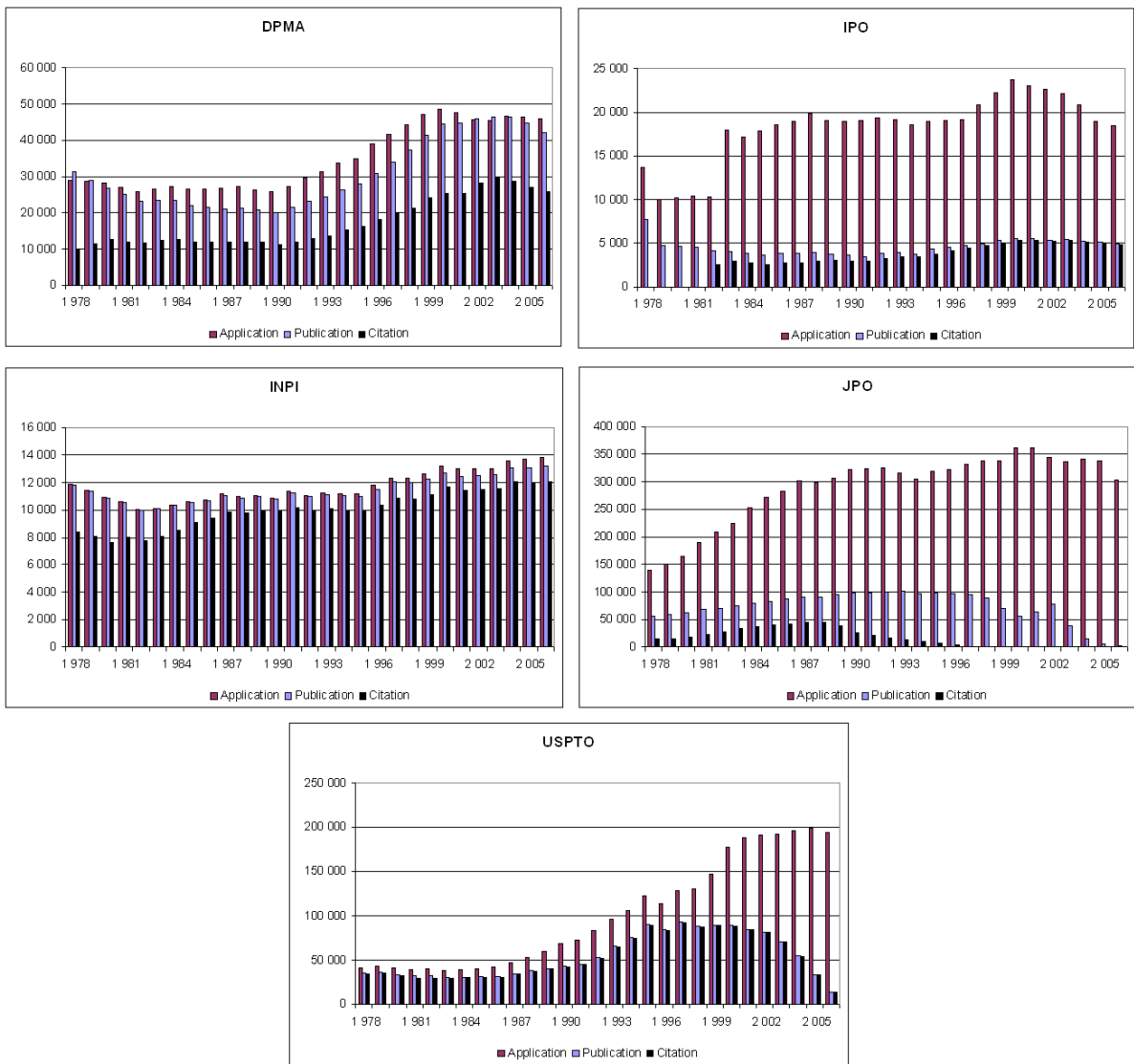


Figure 5.6: Citing cohort statistics summary

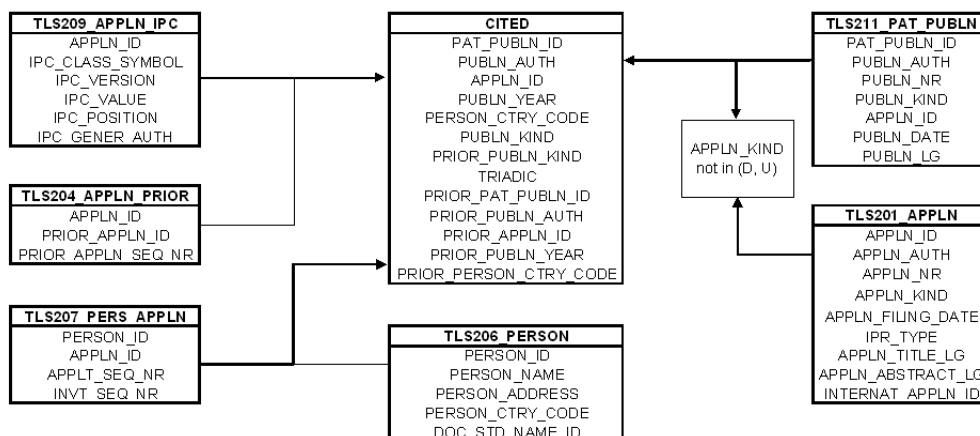


Figure 5.7: The Cited Table

The statistics of *Cited* table are provided in Table 5.4. The first column values represent the number of publications without priority claimed. As expected, values are closed to those provided in Table 5.2. We do not report the ratio of documents with a person country code available because these values are similar to those already provided in Table 5.2. New results lie on the part of first patent applications in selected offices. As expected, the EPO presents the lowest rate, only 5% of publications are outcomes from first applications. We have already noted that in the three European national offices, most candidates first file in their national office before going to the regional level. This behavior seems to be also accurate for offices other than the European ones. JPO is the office with the highest rate of first application, only a small part of second filing are directed toward this office. Finally, European countries and the USPTO present a rate in the same order of magnitude, lying between 30 % and 60 %.

Office	Publn with no priority claimed	Total publn	Rate of first application
DPMA	940 908	2 540 901	37.03%
EPO	84 857	1 735 863	4.89%
INPI	351 323	618 359	56.82%
IPO	151 631	534 730	28.36%
JPO	2 227 790	2 648 296	84.12%
USPTO	1 908 996	4 043 096	47.22%

Table 5.4: Cited table statistics (1976-2006)

## 5.4 Empirical investigations of patent citations among the different offices

### 5.4.1 Analysis & statistical overview

To start off with, there is a fundamental issue concerning the use of citations made by examiners from the major patent offices for the study of knowledge flows. We have previously established the possibility of using USPTO and EPO citations data for the analysis of knowledge flows. Although it is obvious for inventor added citations, it is not straightforward for examiner added citations. This point was extensively discussed in the previous chapter. To be brief, our argumentation was done in two steps. First, Breschi & Lissoni (2004) [23] have shown that from a theoretical point of view, there is no reason to exclude citations drawn up by examiners. Second, we have reported several empirical studies which have demonstrated the correlation between knowledge flows and EPO citations which are almost examiner added citations. While EPO and USPTO citation data were extensively studied in the literature, there exist a gap concerning citations set up at the DPMA, the INPI, the IPO and the JPO offices. We have already discussed the process of patent procedures for each office concerned, stressing the similarities between EPO and national European offices in the citations process in one hand and between the EPO and JPO in the other hand. This section goes further into the analysis of patent citations by looking into citation data from these offices. In a first time, we will investigate how much and what is exactly cited by the various patent offices. In a second time, we will compare citations made by examiners of European national and JPO offices to those made by the EPO examiners for patents sharing the same priority number.

### 5.4.2 Number of citations per patent

The number of citations that a patent makes (backward citations), varies substantively according to the offices. Table 5.5 gives an overview of the average number of patent citations that have been broken down by patent office for the publication year. The values are calculated as the ratio of the first column of Table 5.3 by the first column of Table 5.2. Concerning European national office (i.e. DPMA, INPI, IPO), the average number of citations are of the same order of magnitude compared to the EPO's one. Nevertheless, these values are lower for national offices than for the EPO. The average number of citations is 4.5 for the EPO, 3.5 for the IPO and the INPI and 2.5 for the DPMA. As noted

in the previous chapter, different rules lead to substantial differences between US and European patent offices. Due to the “duty of candor” (see chapter 4), US patent references are three times larger than those of the European offices. These results are in line with a previous study conducted by Michel & Bettels (2001) [122] except the value concerning the DPMA. Authors found a value one and a half higher than our calculation<sup>9</sup>. Nevertheless, a recent European technical report [51], partly based on the work of Cremers (2009) [35], found similar results in the order of magnitude compared to ours (they report an average of 2.4 for the DPMA patents).

Figure 5.8 presents the evolution of the number of citations per patent application over the period. We have already discussed the problem of JPO citations data availability. Nonetheless, behind the great number of applications in the JPO, our database contains more than one million of patent references that could be useful in the investigation of knowledge flows. We do not report the number of patent applications for the JPO in Table 5.5 because the pattern of citation data availability, in Figure 5.6, is similar to the pattern of citations number in Figure 5.8. This situation would lead obviously to misleading results. The national European offices present upward trends, especially for the DPMA where the number of citations has nearly doubled over the period. The most important increase concerns the USPTO where this number has been multiplied by four. Finally, regional offices (the EPO and the WIPO) present constant patterns.

Office	DPMA	EPO	INPI	IPO	USPTO	WIPO
Period 1990-1999 Source: Michel and Bettels, 2001	3.98	4.37	n.a.	3.87	12.96	n.a.
Period 1978-2006 Source: Own calculation from national database (Patstat)	2.27	4.62	3.77	3.49	12.52	4.73

Table 5.5: Average number of patent citations per application

### 5.4.3 Origin of references

Search reports are generally worldwide, but there exists practical limitations which are principally due to linguistic barriers. A westerner’s examiner would be unlikely to access database which can only be searched in Japanese characters (Michel & Bettels 2001 [122]). In the other hand, examiners should preferably reference documents in the language of the application and/or belonging to the same

---

<sup>9</sup>Michel and Bettels (2001) use the EPOQUE database while I use only the first patent applications from to the PATSTAT database



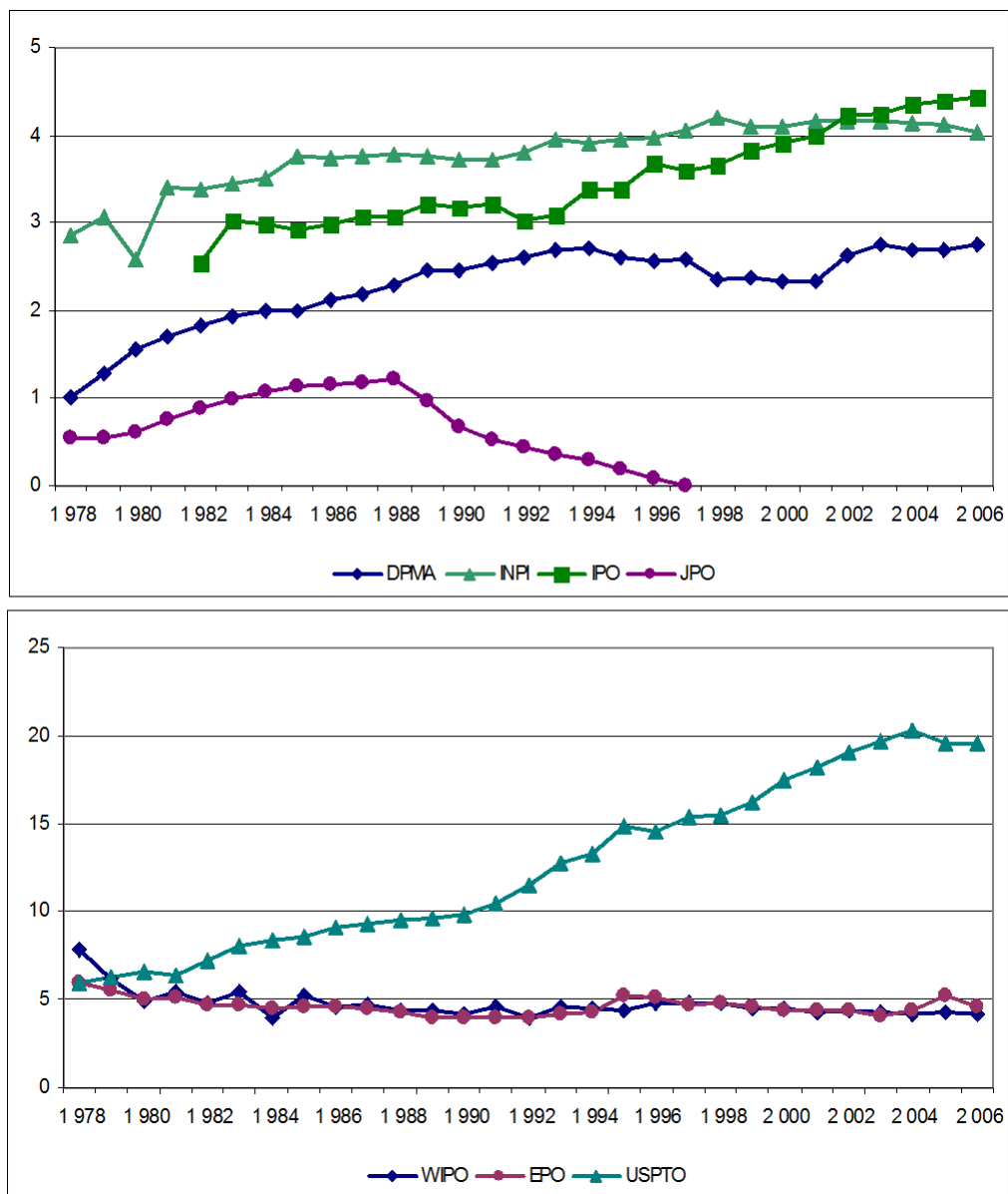


Figure 5.8: Number of citations per patent application

office, inducing an over-estimation of the influence of the applicant's home country and a distortion in citation count. This situation can be illustrated by patent documents provided in the previous chapter. Reconsider search report of Patent P1 (Figure 4.8 in Appendix of chapter 4), Patent P1 provides citations to previous patents that are related to this innovation. For instance, this patent cites another patent P22 (Figure 4.12), which was developed by Japanese inventors. But this patent is not a first application and a corrected reference by prior equivalent will refer to its prior application P2 (Figure 4.11). In order to have a global view of who cites what and to have an idea of the influence of the applicant home country on citation process (the home advantage effect) we will investigate the origin of citations for the selected patent offices both at a direct and indirect degree.

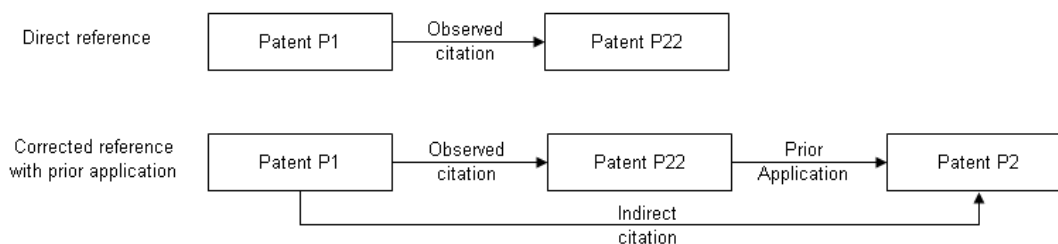


Figure 5.9: Direct and Corrected references by prior application

**Statistical investigation** Figure 5.10 gives a breakdown of the origin of cited patents at selected citing patent offices (we do not take into account references toward Non Patent Literature). Let's start with the direct reference mechanism as it is represented in Figure 5.9. Direct reference looks exclusively on citations as they are drawn up in the search report, relating to previous example, it is the case when the INPI patent P1 cites the EPO patent P22 (remember that, in this chapter, the citing patents cohort are exclusively based on prior applications, this is why we refer to the INPI search report and not to the EPO search report as that could be done in the previous chapter). The most striking finding is the dominant proportion of domestic citations for the JPO and USPTO where more than 90% of all citations are done toward domestic's references. In all cases, at the exception of INPI and WIPO, direct citations are dominated by references to patents belonging to the same office. The French national office set out the lower rate of own national citations, with a value of about 22%. More unexpected, INPI patents cites more USPTO patents than its own national patents with a rate of about 29%. All offices, at the exception of JPO, cite largely patents published by the USPTO suggesting an important transfer of knowledge from the US to other countries. On the contrary, OT proportion is very low, on average less than 3 % of citations are directed toward other offices than the

seven selected patent offices (OT holds for other patent offices and represents all the 74 other patent offices, see Table 5.7).

Results of interest happen when we turn to indirect citations figures. Indirect citations correspond to citations corrected by prior application for the patent cited (if cited patent is a first application we keep the direct references). In the previous example, this situation corresponds to the case where the INPI patent P1 cites indirectly the JPO patent P2, see Figure 5.9. In all cases (at the exception of the INPI), the part of own national citation decreases. JPO that was previously absent of cited offices, now gets a substantial part of references, more than 10% in all cases. Second feature is the part of the other offices, OT, that also gains in proportion, reaching almost 10% in most cases. On the contrary, the own part of EPO and WIPO collapses because of the low proportion of first application in these offices. More interesting is the similar pattern of citations between these two offices both before and after the correction for prior equivalences. Nevertheless, even after the correction process, the JPO, the USPTO and the DPMA still stay in a situation close to exclusive own citation practice, with a respectively rate of 90%, 70% and 60%. Figure 5.10 shows that examiner tends largely to cite patents belonging to the same office and overwhelming the proportion of own national citation, maybe at the exception of INPI. By considering indirect citations, the pattern of origin is more widespread and gives place to patents arising from other offices than the seven selected.

**Date issue** The over-estimation of own domestic or regional references could induce misleading measures of the citation process especially if we use a global database as PATSTAT. We know that the date that matter for the cited patent is the date at which the information becomes available, i.e. the publication date. The actual date on which the invention characteristics become publicly available is the publication date of the prior application. Thus, considering only direct citations to the patents invented by foreigners, a homogeneity date problem arises. For instance, consider the case of patents invented by Japanese that are likely to be cited by INPI patents. We have shown that, considering only direct citations from INPI patents, we have over-estimated the citations toward EPO or WIPO patents on the expense of JPO patents. Thereby, if a citing cohort cites both an EPO and a JPO Japan invented patent with the same priority year, the publication date of the subsequent EPO application will be older than the one of the JPO application (all things being equal, i.e. with a same delay for requesting examination at the JPO...). Per consequence, at least one part of the difference between these inventions in term of diffusion time pattern will be due to this issue, resulting in a shorted

diffusion for the invention bearing an EPO application.

Of course we must bear in mind that time diffusion is not only dependent on the publication date and may also result from intrinsic features of the invention. Nevertheless, we face a homogeneity problem that echoes to the discussion held in the previous chapter. Do we assume that it is the time at which the patent is readable by another, which makes the information available as it is considered by Jaffe et al. (2000) [88], or do we assume that is the time at which the information is legally available anywhere in the world and thus diffuses by social chains, as it is argued by Breschi & Lissoni (2004) ? In this dissertation we have clearly side with the latter. Moreover, considering only the publication date of the direct cited patent, even if it is not a prior application, we can face a problem of double counting, because of the global dimension of the database. This case can occur if a citing patent cohort cites two patents belonging to the same patent family<sup>10</sup>, cited patents represent the same invention, but with obvious different dates of publication.

### 5.4.4 Similarity of citation between EPO and European national offices

The next topic to be explored is the similarity of family search reports among the different offices, i.e. what is the proportion of citations that match between the different applications of a same invention? The application of the same technical invention in different patent office rises to a “patent family” and the different search reports derived from these applications rises to a “search report family”. This issue is particularly relevant for European countries since almost all their patents are first filed in a national office. In the other hand, we have established that the EPO citations are a good indicator of knowledge flows, therefore, high similarities between national and EPO citations means that the national citations are, per consequence also a good indicator of knowledge flows. Here, we propose to investigate the degree of similarity between citations made by DPMA, INPI and IPO examiners with those added by EPO examiners for patents that are first filed in one of the three national European offices. As highlighted before, it is obvious that a large part of citations does not match between the national and regional office for the origin of citations differs significantly according to the office considered, see Figure 5.10. But conversely we can expect that an EPO and DPMA examiner drawing up a search report about the same invention (first file at the DPMA and after at the EPO) will cite

---

<sup>10</sup>A patent family is a set of individual patents granted by various countries. The patent family is all the equivalent patent applications corresponding to a single invention, covering different geographical regions. Patent family size is a measure of the geographical breadth for which protection of the invention is sought.

patents belonging to the same patent family. For instance, we can expect that a DPMA examiner cites a German prior application and the EPO examiner its second application made at the EPO. In order to investigate this issue we proceed in two steps, as is illustrated in Figure 5.11.

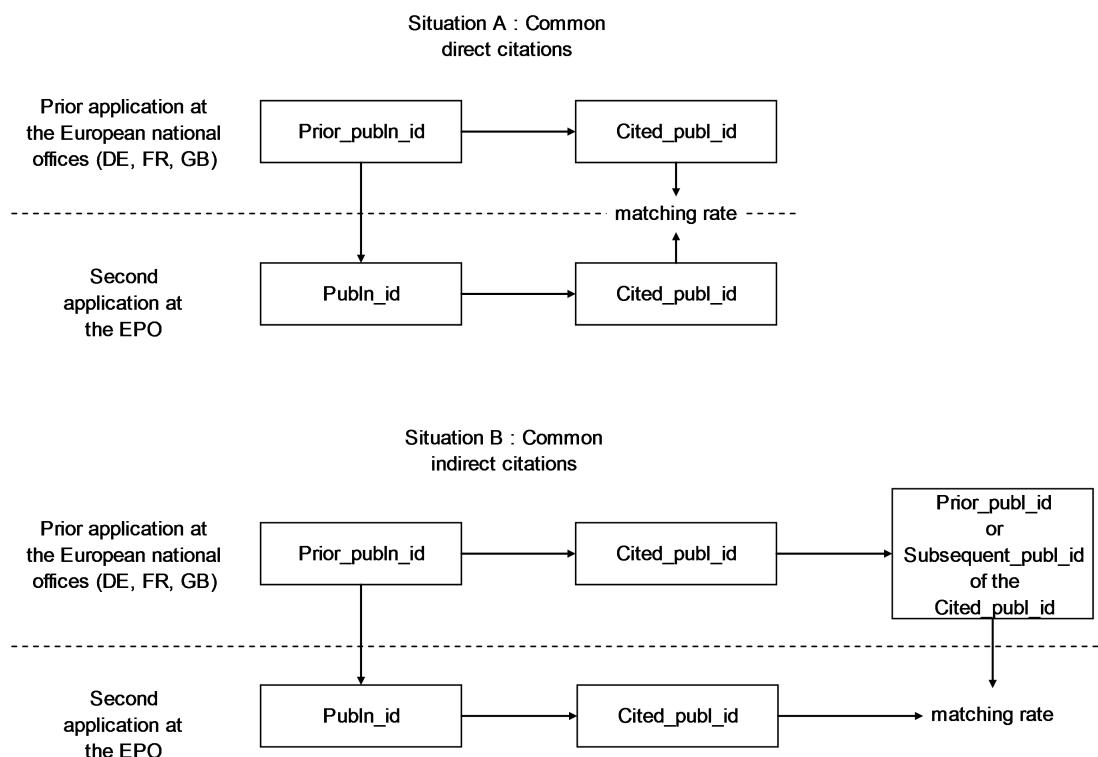


Figure 5.11: The matching citations structure

We start with comparing the publication number of the cited patents as illustrated by the situation A. Then, we turn to the comparison of indirect citations as illustrated by situation B. In situation B, we identify patents cited by the application made at the European national offices and the subsequent or prior applications of these cited patents. Then, we match these indirect cited patents with the ones cited by the subsequent EPO application. Table 5.6 provides results of our matching procedure. We investigate citations matching rate since the creation of the EPO in 1978 through 2006 for the INPI and the DPMA and since 1982 for the IPO because citation data are not available before for this office (cf. Table 5.8). Before turning to citations matching results, the procedure gives us some information on the rate of patents that are subsequently filed to the EPO. INPI patents present the highest rate of subsequent filing to the EPO, about 35% of all patents first filed at the INPI are subsequently filed at the EPO (recall that these patents are almost exclusively developed by French inventors, at 92 %).

	<b>DPMA</b>	<b>INPI</b>	<b>IPO</b>	<b>JPO</b>
Period	1978-2006	1978-2006	1982-2006	1978-1996
Prior application (A)	889 153	332 434	111 542	1 603 566
Second appl to the EPO (B)	252 431	117 391	12 359	35 959
Ratio (B/A)	28.39%	35.31%	11.08%	2.24%
Nbr of citations of (B)	568 141	473 076	36 895	22 524
Common direct citations	63 172	280 364	3 618	857
Matching rate	11.12%	59.26%	9.81%	3.80%
Common indirect citations	186 000	132 135	10 795	12 411
Matching rate	32.74%	27.93%	29.26%	55.10%
Direct + Indirect	249 172	412 499	14 413	13 268
Total matching rate	43.86%	87.20%	39.06%	58.91%

Table 5.6: Matching rate of citations

DPMA patent rate is not so far with a value of about 28%. In contrast, only 11 % of IPO patents are subsequently filed to the EPO (but only 80% of British inventors first file to the IPO).

Our database allows us to have a sample size relatively high for all offices selected even for the IPO which contains over 36 000 citations from patents that were subsequently filing to the EPO. Concerning direct citations, INPI presents the highest rate, nearly two third of direct citations are similar. This position was expected since INPI is the office that cites the more EPO patents through direct citations, but the extent was not. INPI refers 50% more to EPO patents than its counterpart European office, about 15% against 10%, see Figure 5.10, that is not so much in regard of the difference in citation similarities, the latter is 6 times higher for “Common direct citations”, see Table 5.6. Since we do not expect to have high matching rates for direct citation, we turn to indirect citations. Between the different offices, the rates are similar, about less than one third of indirect citations are similar. This rate seems normal for the IPO and DPMA since these offices have a high own national citation practices. Finally, overall results show that French search reports are very similar to EPO ones, with a rate of common citations close to 90%. For the other two offices, rates are twice below the INPI, with values about 40%, which means that less than half of national examiner added citations refers to the same patent family as those cited by the EPO examiners. In conclusion, we can argue that the INPI citations are a good indicator of knowledge flows through the transition of the argumentation made at the EPO level. Concerning citations made by the DPMA and the IPO examiners, we can not conclude on the quality of citations as a knowledge flows indicator. Conclusions on this topic require to wait for

further analysis and empirical tests that we will implement in the next chapter.

### 5.4.5 Similarity of citation between EPO and JPO

Goto and Motohashi (2007) have compared the JPO citations with the ones from USPTO and EPO within the OECD patent family database<sup>11</sup>. On the 113 490 JPO citations contained in their sample, 23 904 match with USPTO and 9 046 with EPO citations (respectively 20% and 8%). Authors explain their low matching rates by their limited sample data. They dispose only of citations to domestic patents for the JPO and USPTO data. The second limitation is the home country bias towards their own patents, as shown in Figure 5.10.

We take advantage from information that Goto and Motohashi have not and we implement the same process that we have used to investigate the citation similarities between European offices and EPO. We limit our investigation with EPO because, as discussed in previous chapter, the JPO citation process is closer to the EPO than to the USPTO. Results are provided in Table 5.6. The direct matching rate is two times smaller than those found by Goto and Motohashi, this value can be explained by the difference of the sample considered. First, they retain a period from 1978 to 2002 while we retain a reduced one: 1978-1996. Second, they take into account all patent applications filed at the JPO while we retain only first patent applications at the JPO, thus we withdraw all second applications made to the JPO by foreigners. As suggested in the previous section (Origin of References), the matching rate increases substantially up to 55% when we turn to indirect citations rate. Finally, the direct and indirect rate reached together a value close to 60% which is higher than for the IPO and the DPMA, meaning that the JPO citations may be used as an effective proxy for tracing knowledge flows.

## 5.5 Conclusion

In this chapter, it has been attempted to present the Patstat database in detail. Patstat is one of the richest sources of data on inventions and certainly the one with the widest coverage that the economists can access. Nevertheless this database is quite difficult to handle and has not been exploited so much until now. We have tried to explicit the construction of a sub database directed to the extraction and computation of data needed for the estimation of the patent citation function in an international

---

<sup>11</sup>The OECD patent family database is comprised of patents on patented inventions, at least two of the following three jurisdiction: Europe, Japan and the US (Dernis and Khan, 2004)

way. Through the construction of this database we have been able to present some interesting facts on patent procedures. This effort was as much as important that there exists a gap on patent literature concerning patent offices other than the USPTO and EPO. One focus of this chapter was deeply done on citations process over the DPMA, the EPO, the INPI, the IPO, the JPO, the USPTO and the WIPO office. Analysis reveals an evidence of domestic bias effect of citations toward patents belonging to the same office, but. this effect can be in part straightened by a prior equivalence correction. The main contribution of this chapter results in the comparative analysis of citation structures among the different offices considered. Figures highlight the similarity of citation structures among offices when we take into account reference to patent family. In a citation point of a view, INPI is the closer office to the EPO (citations between these two offices match at a rate of 87%), meaning that INPI citations can be used as an effective tool for analyzing knowledge flows and the cumulative nature of innovation by transition of EPO citation properties. Concerning DPMA and IPO, the matching rate is quite lower with a respective value of 44 % and 40 %. These results do not mean that we cannot use these citations to the study of knowledge flows, but simply that we need to go further into the analysis of citations process. Finally we have also investigated the similarity between JPO and EPO citations. Result is unexpectly fair since we reach a matching rate close to 60 %.



## 5.6 Appendix

### 5.6.1 Tables and figures

---



---

1 Albania (AL)	42 Kenya (KE)
2 ARIPO (AP)	43 Korea (South) (KR)
3 Argentina (AR)	44 Liechtenstein (LI)
4 Austria (AT)	45 Lithuania (LT)
5 Australia (AU)	46 Luxembourg (LU)
6 Bosnia and Herzegovina (BA)	47 Latvia (LV)
7 Belgium (BE)	48 Morocco (MA)
8 Bulgaria (BG)	49 Monaco (MC)
9 Brazil (BR)	50 Moldova (MD)
10 Canada (CA)	51 Former Yugoslav Republic of Macedonia (MK)
11 Switzerland (CH)	52 Mongolia (MN)
12 China (CN)	53 Malta (MT)
13 Czechoslovakia (CS)	54 Malawi (MW)
14 Cyprus (CY)	55 Mexico (MX)
15 Czech Republic (CZ)	56 Malaysia (MY)
16 Cuba (CU)	57 Netherlands (NL)
17 German Democratic Republic (DD)	58 Norway (NO)
18 Germany (DE)	59 New Zealand (NZ)
19 Denmark (DK)	60 OAPI (OA)
20 Algeria (DZ)	61 The Philippines (PH)
21 Eurasia (EA)	62 Poland (PL)
22 Estonia (EE)	63 Portugal (PT)
23 Egypt (EG)	64 Romania (RO)
24 European Patent Office (EP)	65 Russia (RU)
25 Spain (ES)	66 Sweden (SE)
26 Finland (FI)	67 Singapore (SG)
27 France (FR)	68 Slovenia (SI)
28 Great Britain (GB)	69 Slovakia (SK)
29 Gulf Cooperation Council (GC)	70 Soviet Union (SU)
30 Georgia (GE)	71 Tajikistan (TJ)
31 Greece (GR)	72 Turkey (TR)
32 Hong Kong S.A.R. (HK)	73 Taiwan (TW)
33 Croatia (HR)	74 Ukraine (UA)
34 Hungary (HU)	75 United States of America (US)
35 Indonesia (ID)	76 Viet Nam (VN)
36 Ireland (IE)	77 World Intellectual Property Organization (WO)
37 Israel (IL)	78 Former Serbia and Montenegro (YU)
38 India (IN)	79 South Africa (ZA)
39 Iceland (IS)	80 Zambia (ZM)
40 Italy (IT)	81 Zimbabwe (ZW)
41 Japan (JP)	

---



---

Table 5.7: PATSTAT data coverage

DB name	Element
APPLN_ABSTRACT	Abstract of application
APPLN_ABSTRACT_LG	Language of abstract of application
APPLN_AUTH	Application Authority
APPLN_FILING_DATE	Filing date of application
APPLN_ID	Application identification
APPLN_KIND	Kind of application
APPLN_NR	Application number
APPLN_TITLE	Title of application
APPLN_TITLE_LG	Language of title of application
APPLT_SEQ_NR	Sequence number of the applicant
CITED_PAT_PUBLN_ID	Identification of cited patent publication
CITN_CATEG	Category of the citation
CITN_ID	Citation identification
CITN_ORIGIN	Origin of the citation
CONTN_TYPE	Continuation type
DOCDB_FAMILY_ID	DocDB simple family id
DOC_STD_NAME	DocDB standard name
DOC_STD_NAME_ID	DocDB standard name identification
ICO_CLASS_SYMBOL	EPO InComputerOnly classification symbol
INPADOC_FAMILY_ID	INPADOC extended priority family
INTERNAT_APPLN_ID	Application identification of designating international application
INVT_SEQ_NR	Sequence number of the inventor
IPC_GENER_AUTH	IPC generating authority
IPC_CLASS_SYMBOL	IPC classification symbol
IPC_CLASS_LEVEL	IPC advanced / core flag
IPC_POSITION	First or later position of symbol
IPC_VALUE	Classification value
IPC_VERSION	IPC version
IPR_TYPE	Type of Intellectual Property Right
NAT_CLASS_SYMBOL	Classification symbol of National classification
NPL_BIBLIO	Non-Patent Literature bibliography
NPL_CITN_SEQ_NR	Sequence number of the NPL citation
NPL_PUBLN_ID	Non-Patent Literature publication identification
PARENT_APPLN_ID	Application identification of parent application
PAT_CITN_SEQ_NR	Sequence number of the patent citation
PAT_PUBLN_ID	Patent publication identification
PERSON_ADDRESS	Applicant address
PERSON_CTRY_CODE	Applicant country code
PERSON_ID	Applicant identification
PERSON_NAME	Applicant name
PRIOR_APPLN_ID	Application identification of claimed priority
PRIOR_APPLN_SEQ_NR	Sequence number of claimed priority
PUBLN_AUTH	Publication Authority
PUBLN_DATE	Publication date
PUBLN_KIND	Publication kind
PUBLN_LG	Publication language
PUBLN_NR	Publication number
TECH_REL_APPLN_ID	Application identification of the technically related application

Table 5.8: PATSTAT elements

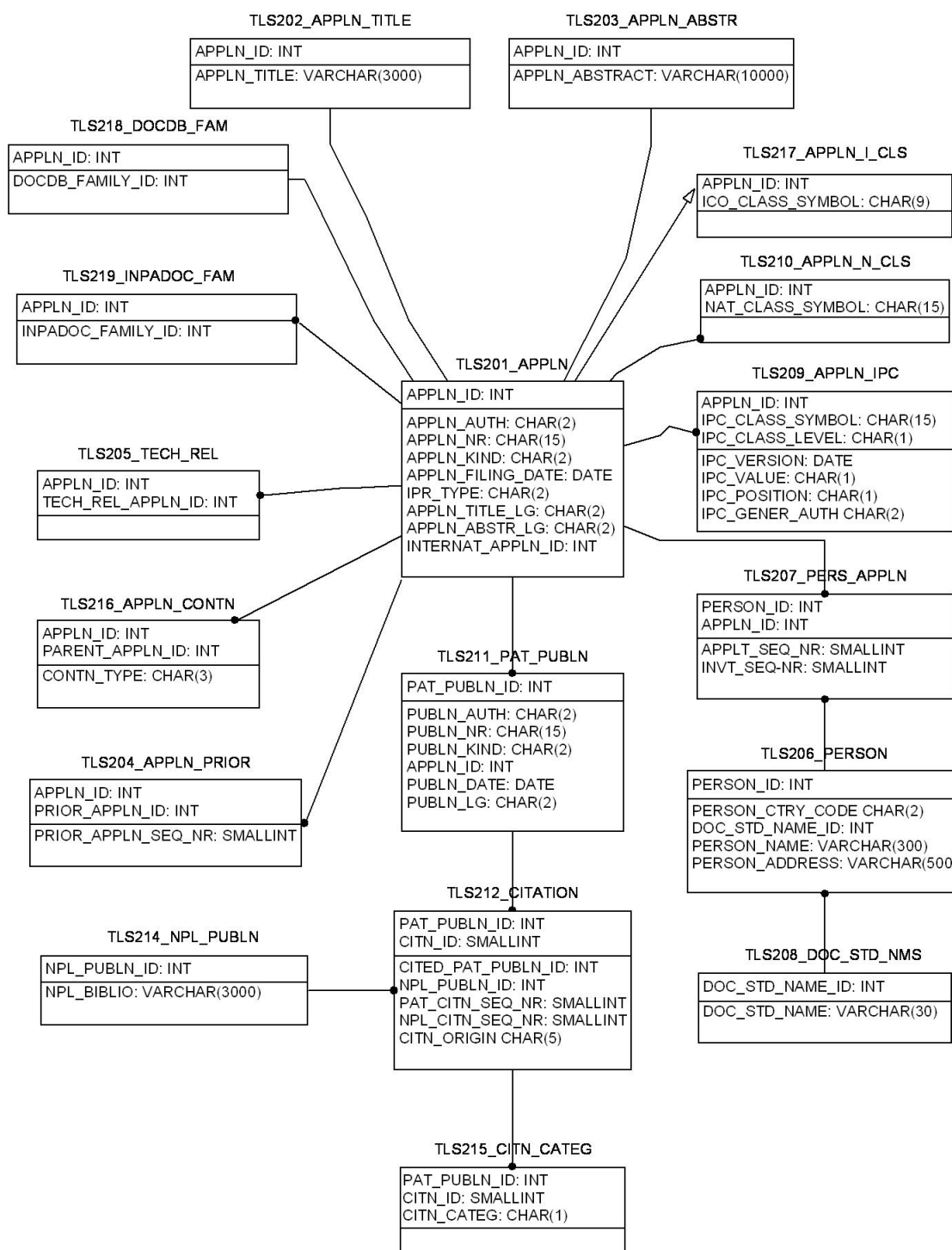


Figure 5.12: The PATSTAT physical model

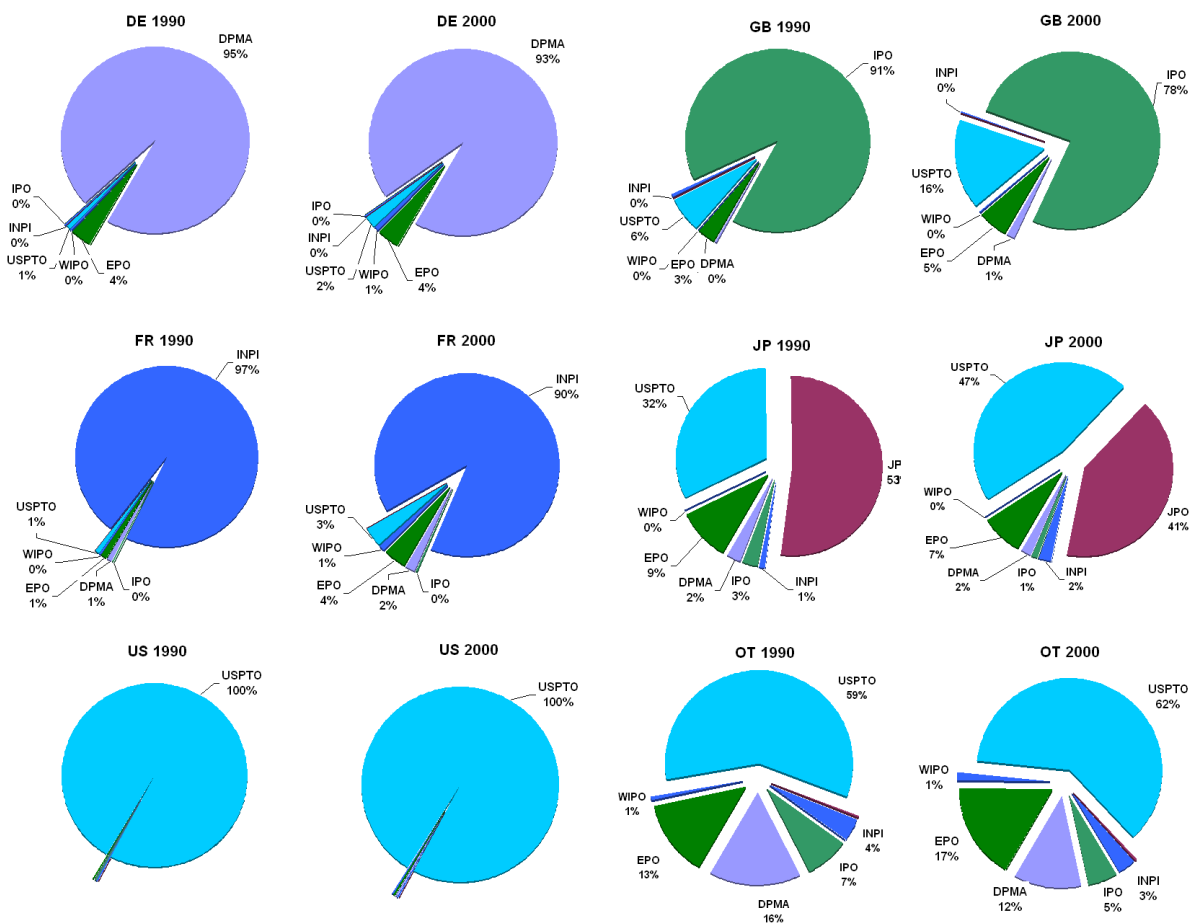


Figure 5.13: Inventors first applications according to the selected offices

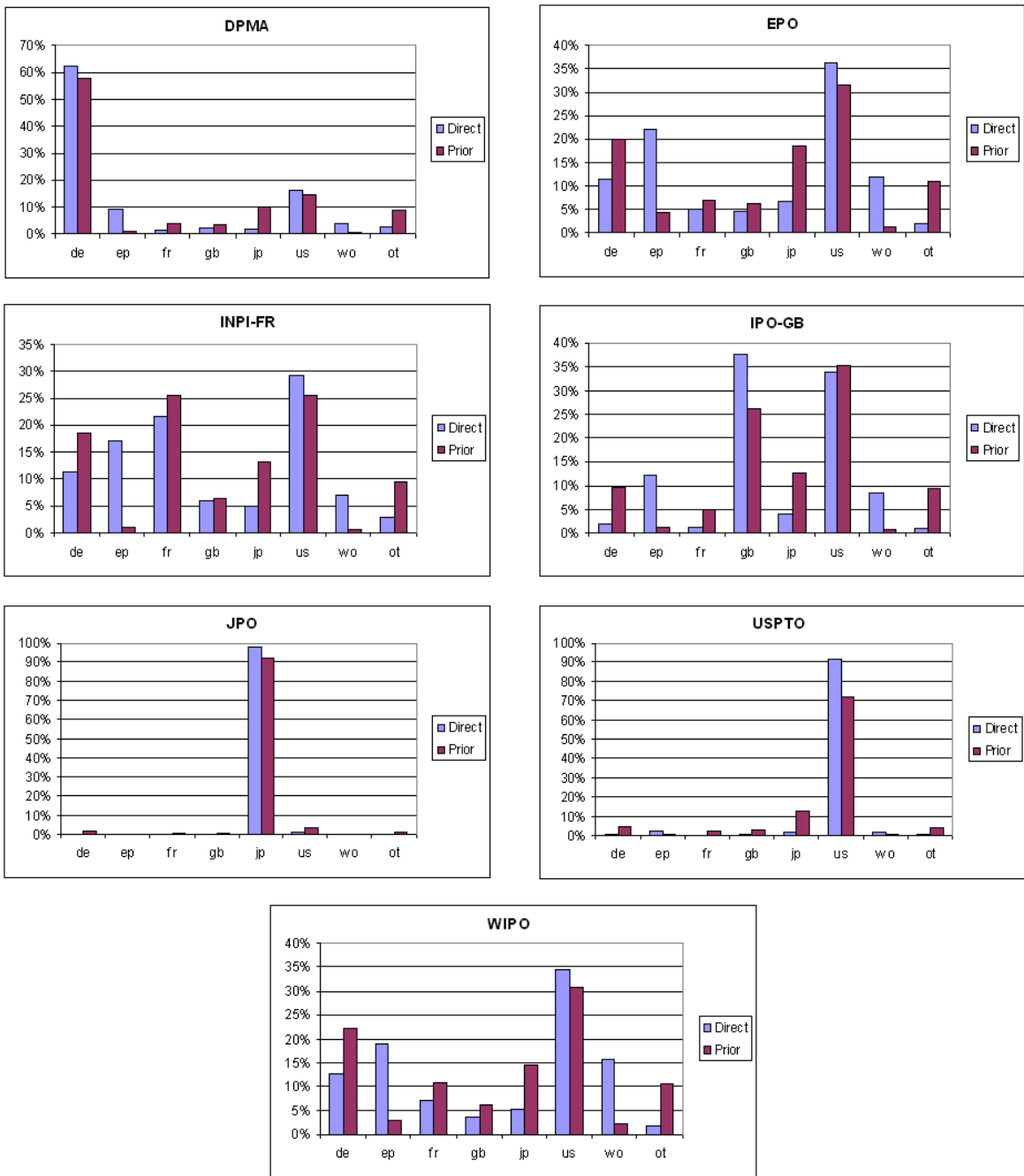


Figure 5.10: Origin of references

## Chapter 6

# A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

### 6.1 Introduction

In chapter 4 we have assessed the possibilities offer by patent and patent citations data into the exploration of knowledge diffusion process. These possibilities have emerged with the recent literature, led by Jaffe & Trajtenberg [87], which provides a structural framework for analyzing patent citations to previous patents hence opening a window on the understanding of knowledge flows process. The riches of patent data have thus been used as a way for tracking the influence of past inventions across time, space, technological fields and institutions. These studies were mainly carried using the USPTO and EPO database, but unfortunately, they suffer from the lack of international dimensions of these databases. Indeed, patent data, taken from these offices, are strongly biased toward domestic inventors inducing the well known “home advantage effect”, see Criscuolo (2006) [37]. For instance, studying international knowledge diffusion with the USPTO database will induce a positive bias toward US inventors. Fortunately, the EPO has spent effort on the construction of a worldwide database, PATSTAT, which regroups patent and patent citation data from almost all patent offices dispersed around the world.

The utilization of citations data taken from different offices, assumes that all of them are a good

indicator of knowledge flows. Previous studies have emphasized the possibilities of using EPO and USPTO citation data as a source for the measurement of knowledge flows but any study has investigated this possibilities for citation data originating from other offices (with the exception of Goto & Motohashi 2007 [55] for JPO data). Validate this issue is fundamental to the use of the PATSTAT database in the analysis of international knowledge flows. The last part of chapter 5 has tried to fill this gap by investigating the proximity between the INPI, IPO, JPO and DPMA patent citation data with the EPO ones. The main result is that strong similarities were found between JPO and INPI with EPO citations, but although we find a correlation of about 50% between the EPO data and other offices, some doubts remain about the DPMA and the IPO patent citations. Because Patstat is one of the largest digital repositories of public knowledge in the world, the validation of the possibilities offered by such data source is crucial to understand the institutional, geographic and temporal dimensions of the diffusion of newly created knowledge.

On the other hand, the importance of knowledge as a key factor of economic growth and as a fundamental input of production functions is now well recognized. Hence, the development of related indicator becomes essential for modeling and policy making purposes. Therefore, many researchers have strived to measure knowledge stock by using R&D expenditures data. However, these data represent only the input side of knowledge, thereby making us over value the stock of knowledge. Moreover R&D expenditures data are often not available over long period, neither at a sectoral, and even less, at a technological level.

These reasons promote the research of new indicators in which patents seem to take a fundamental part. But patent technological values are skewed, and their utilization in the construction of a knowledge input based indicator imposes to control changes in the technological opportunities of knowledge they embody. Popp [146] provides an interesting advance in this way, by developing an indicator based on the patent citation function (extensively detailed in the chapter 4) for US energy technologies. In his formulation, patent productivities depend on the number of citations received. But this indicator, based only on domestics knowledge, withdraws advances made by the international knowledge spillovers literature that set up the importance of foreign knowledge in national productivity growth<sup>1</sup>. Popp was limited by data availability during his research period, and thus restricts his analysis to domestic inventors within the USPTO database. In chapter 4, we have laid the foundation for the

---

<sup>1</sup>A meta analysis of such impact was implemented in the chapter 3.

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

---

development of a new indicator based on the work of Popp. We have extended his formulation in a way that takes international knowledge spillovers into account, thanks to international patent citation data, avoiding bias toward a particular country. Nonetheless, before setting up this indicator as an alternative of traditional indicator, we need to validate it empirically.

This chapter tries to fill the gaps highlighted above by implementing such indicator for inventors located in the G5 countries : France, Germany, Japan, the United Kingdom, and the United States for 13 industrial sectors and for a large time period : 1978 to 2003. This indicator is based on patent data taken from the seven major patent offices around the world : the DPMA, the EPO, the INPI, the IPO, the JPO, the USPTO and the WIPO. The indicator created, quantifies the knowledge used by an inventor within a specific space, time and technology dimension, depending on the geographical origin and publication year of this knowledge. Chapter 4 has highlighted homogeneity problems with data aggregated from different offices. However these problems can be avoided by :

- considering only a citing patent cohort by office,
- controlling cited patent cohort by the breadth of patent according to the office considered and
- identifying explicitly the prior applications

The contributions of this chapter can be summarized into two main objectives. (i) The first is to investigate econometrically the pattern of citations between inventors of the G5 countries with the citation function. For the first time, we explore citations made in the INPI, the DPMA, the IPO and the JPO that gives a richer picture of the geographic dimension of citations among these different offices. (ii) The second is to create a new knowledge stock indicator based on these patents and to validate it empirically, by testing its similarities with traditional indicators (R&D expenditures based indicators) and by testing its explanatory power on economic growth. This validation process leads to three important results.

- First it validates the use of patent citations, from the different offices selected, as a measure of international knowledge spillovers.
- Second, it validates the use of the model created by Popp for the creation of knowledge stock that could compensate the lack of R&D data generally not available at this range of detail.
- And finally, the indicator proposed takes into account international knowledge flows which can be used for the search of international spillovers, allowing a new perspective on this subject.



This chapter is organized in five sections as follows: the first section reminds the presentation and the construction of our indicator (Section 6.2), the second presents the data used (Section 7.4) and the third section describes the knowledge diffusion model and presents its estimated results (Section 6.4). The fourth section is dedicated to the empirical construction of the indicator (Section 6.5) and the last section validates it empirically (Section 6.6).

## 6.2 The input-based patent citation indicator

The patent citations indicator (our input-based patent citation indicator of knowledge stock, see chapter 4), is based on the formulation of knowledge stock developed by Caballero & Jaffe (1993) [25] and on the recent work of Popp (2002) [146] in which knowledge flows are proxied by patents. Nevertheless, the utilization of patents as an input indicator of knowledge suffers from a problem of technological value heterogeneity. Popp (2002) has demonstrated that when one wants to consider patents as an input indicator, it is necessary to take into account changes in the productivity or in the fertility of knowledge embodied in these patents. A solution is provided by the patent citation function that provides a framework to the estimation of these patents productivities. The underlying assumption is that a cohort providing useful knowledge is more likely to be cited by subsequent cohorts. The second interest of this function is that it allows the measurement of the two fundamental forces intrinsic to knowledge flows, namely the obsolescence and the diffusion of knowledge. As already mentioned, we extend previous methodology in a way that takes international knowledge spillovers into account, the resulting stock of knowledge will be split up into two components: the domestic and foreign stock of knowledge.

**Knowledge stock** Knowledge flows are composed by patents that are weighted by their technological opportunities to ascertain the usefulness of these patents to future inventors. The extent of technological opportunity that a patent can offer to future inventor depends on the relative number of citations received. Knowledge flows are affected by a combination of two exponential processes, one for the knowledge diffusion and the other for the natural process of obsolescence. The domestic stock of knowledge, *Knowd* that an inventor located in country  $L$  at year  $T$  could take advantage from, is defined by :

$$Knowd_{L,T} = \sum_{t=0}^{T-1} \alpha_{L,L,t} \cdot N_{L,t} \cdot e^{-\beta_1 \cdot L \cdot (T-t)} \left( 1 - e^{-\beta_2 (T-t)} \right) \quad (6.1)$$

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

---

and the corresponding foreign stock of knowledge,  $Knowf$ , is defined by :

$$Knowf_{L,T} = \sum_{l \neq L} \sum_{t=0}^{T-1} \alpha_{L,l,t} \cdot N_{l,t} \cdot e^{-\beta_1 \cdot L \cdot (T-t)} \left(1 - e^{-\beta_2(T-t)}\right) \quad (6.2)$$

Where  $N_{l,t}$  is the flow of patents published in year  $t$  by inventor located in country  $l$ ,  $\alpha$  is the productivity of patent i.e. the extent to which knowledge embodied in that patent provide technological opportunity for subsequent inventors,  $\beta_1$  is the rate of obsolescence and  $\beta_2$  is the rate of diffusion. Domestic and foreign knowledge stocks are calculated using the Perpetual Inventory Method (PIM) that gives for the domestic component<sup>2</sup>:

$$Knowd_{L,T} = \sum_{t=b+1}^T \alpha_{L,L,t} \cdot N_{l,t} \cdot e^{-\beta_1(T-t)} \left(1 - e^{-\beta_2(T-t)}\right) + e^{-\beta_1(T-b)} \left( \frac{1+g}{(1+g) - e^{-\beta_1}} - e^{-\beta_2(T-b)} \frac{1+g}{(1+g) - e^{-(\beta_1+\beta_2)}} \right) \alpha_{L,L,0} \cdot N_{L,0} (1+g)^b \quad (6.3)$$

The particularity of such specification is that the productivity of patent cohort,  $\alpha_{L,l,t}$ , depends, for each patent cohort published at year  $t$ , on both characteristics of the cited  $l$  and the citing  $L$  inventors. Thus, the knowledge embodied in a specific patent cohort will not have necessary the same impact in different countries, and this difference is not only explained by specific country pair effect. As we have noticed in the introduction, we will breakdown the previous specification into different technological fields  $i^3$ . But to avoid a too heavy formulation, we have withdrawn the technological index  $i$  in previous equations; there is no implication since we do not allow inter-technological citations.

**Citation function** Parameters of equation 6.1 and 6.2 are estimated with the citation function. We estimate the process underlying the citations with a quasi-structural approach, as detailed in chapter 4 (the reader can also refer to Hall et al. 2001 [71]). The citation function can be generally defined by :

$$p(k, K) = \alpha(k, K) \cdot e^{-\beta_1(T-t)} (1 - e^{-\beta_2(T-t)}) \quad (6.4)$$

Where  $p(k, K)$  is the likelihood that any particular patent cohort  $k$ , published at time  $t$ , is cited by

---

<sup>2</sup>Calculus are symmetric for the foreign component.

<sup>3</sup>Although we have discussed sectoral classification in the introduction, we begin the study of knowledge flows in terms of technological classification that we will convert later into a sectoral classification.

some particular patent cohort  $K$ , filed at year  $T$ . The coefficient  $\alpha$  can be seen as the fixed effect in a simple linear model, thus it is allowed to vary with attributes of the citing  $K$  and the cited  $k$  patents. The estimates of a particular  $\alpha(k, K)$ , indicates to what extent a patent  $k$  is more or less likely to be cited by a patent  $K$ , relatively to another cohort. The adequate definition of this parameter allows us to estimate the probability of a particular patent cohort to be cited by another and thus to estimates, all things being equal, the technological productivity of a particular cohort in function of the citing one:  $\alpha_{L,l,t}$ .

## 6.3 Data

### 6.3.1 Technology classification

First step is to define a technological classification  $i$  that will hold for our indicator. Patent technical content is classified according to the International Patent Classification (IPC) that divides technology into eight sections with approximately 60,000 subdivisions<sup>4</sup>. Each subdivision has a symbol consisting of Arabic numerals and letters of the Latin alphabet. Because one aim of this chapter is to test our indicator with traditional measures of knowledge which are expressed in sectoral classification, we need to translate our stock of knowledge expressed in the International Patent Classification into an ISIC-Revision classification. We use the concordance matrix developed by Schmoch et al. (2003) [161] who have established a reliable concordance table between 44 technology fields (based on IPC class) and 44 industry sectors (based on ISIC-Revision 3). This concordance table is based on the empirical investigation of firm patent activities, and each of the 625 IPC subclass corresponds to only one technological field (see Table 6.8.3). Due to the database weight and to data management issues (see Chapter 5), we aggregate the 44 original technological fields into 13 technology classes that regroup the main IPC classes. Detail of this aggregation process is provided in Table 6.1, left side of the Table correspond to the classification used in the present chapter although the right one corresponds to the original ones defined by Schmoch et al. (2003).

---

<sup>4</sup>The complete IPC can be found on the website of the World Intellectual Property Organization (WIPO): <http://www.wipo.int/classifications/ipc/en>

**Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator**

<b>Field no</b>	<b>Description</b>	<b>Field no</b>	<b>Original Descriptions (Schmoch 2003)</b>
1	Food, Beverages and Tobacco	1	Food, beverage
		2	Tobacco products
2	Textile	3	Textiles
		4	Wearing apparel
		5	Leather articles
3	Wood	6	Wood products
4	Pulp, Paper and Publishing	7	Paper
		8	Publishing, printing
5	Petroleum and Nuclear	9	Petroleum products, nuclear fuel
6	Chemical	10	Basic chemical
		11	Pesticides, agro-chemical products
		12	Paints, varnishes
		13	Pharmaceuticals
		14	Soaps, detergent, toilet preparations
		15	Other chemicals
		16	Man-made fibres
7	Rubber and Plastic	17	Rubber and plastic products
8	Non-metallic	18	Non-metallic mineral products
9	Metal	19	Basic metals
		20	Fabricated metal product
10	Machinery	21	Energy machinery
		22	Non-specific purpose machinery
		23	Agricultural and forestry machinery
		24	Machine-tools
		25	Special purpose machinery
		26	Weapons and ammunition
		27	Domestic appliances
11	Electrical and Optical	28	Office machinery and computers
		29	Electrics motors, generators, transformer
		30	Electric distribution, control, wire, cable
		31	Accumulators, battery
		32	Lightening equipment
		33	Other electrical equipment
		34	Electronic components
		35	Signal transmission, telecommunications
		36	Television and radio receivers, audiovisual electronics
		37	Medical equipment
		38	Measuring instruments
		39	Industrial process control equipment
		40	Optical instruments
		41	Watches, clocks
12	Transport Equipment	42	Motor vehicles
		43	Other transport equipment
13	Furniture	44	Furniture, consumer goods

Table 6.1: Technological class based on IPC classification

### **6.3.2 Patents**

In the present chapter we use the PATSTAT database which has been extensively presented in the previous chapter and on which we have set up the two cohorts that will be used in the present chapter. The sub-database created is used to extract patent and patent citation data needed for the estimation of the citation function (see equation 6.4). This function analyzes the determinant of citations between cohorts that differ in a temporal, geographical and technological dimension. As already noticed, we choose to implement one estimation by citing country (i.e. by country of residence of the inventor). The data extracted has to be processed in order to avoid mismeasurement in the parameter estimates, hence regarding the citing cohort, the following assumptions has been made :

- In order to avoid double accounting in our sample and to identify precisely the patent closer to the invention act (in a temporal point of view), we retain only first patent application.
- We consider the citing and cited inventor of five countries: France, Germany, United Kingdom, Japan and the United States. So, we need to count the number of citations between the citing cohort and several cited cohorts, for instance, we count the number of citations that a French inventor makes to patents registered by French, German, English, Japan or US inventors.
- To solve the patents heterogeneity problem between offices, we consider only patents applied in the domestic patent office for the citing cohort. This choice is directed to avoid combining different rules of citation from different patent offices. If a French inventor applies in the USPTO, the citation rules will be very different compared to the INPI citation ones with the consequence that no aggregations are possible. This choice however, does not involve to many data restrictions, because almost all of patents are first filed at national office.

We extract all patents with a first application filed either at the DPMA, the INPI, the IPO, the JPO or the USPTO from 1978 to 2006. Citing patents are classified by geographic origin (the country of the inventor), time (year of [first] application) and technological field. The final selection contains 4 148 014 patents.

Concerning the cited cohort, we do not make as many as restrictions, firstly because, the cited patent is not always a first application, so we consider that the cited patent could belong to the seven selected offices (the G5 patent offices plus the WIPO and the EPO) and secondly because the homogeneity problem could be easily solved following the methodology provided by Dechezprete et al. (2009) [39]. Hence, concerning the cited cohort we process the data with the following assumptions :

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

Statistics for regression variables						
	Total	DE	FR	UK	JP	US
range of citing patent		1978-2006	1978-2006	1983-2006	1978-1996	1978-2006
range of cited patent		1978-2003	1978-2003	1983-2003	1978-1993	1978-2003
Potentially citing patents	4 148 014	807 361	302 058	83 090	1 603 566	1 351 939
Potentially cited patents	4 746 555	827 144	314 710	63 798	2 093 012	1 447 891
Total Citations	15 035 841	910 031	538 502	161 085	802 506	12 623 717
<b>Citing patent by field in %</b>						
Food, Beverages & Tobacco	1.30	0.83	0.78	1.61	1.56	1.15
Textile	1.05	1.05	1.31	1.53	0.91	1.16
Wood	0.55	0.58	0.83	0.66	0.72	0.23
Pulp, Paper & Publishing	0.97	0.73	1.07	0.87	1.17	0.83
Petroleum and Nuclear	0.63	0.45	0.62	0.93	0.39	1.04
Chemical	13.26	13.44	5.08	12.42	13.26	13.86
Rubber and Plastic	4.89	5.30	7.50	7.55	4.43	4.59
Non-metallic	3.14	3.09	3.42	3.44	3.82	2.03
Metal	6.58	7.47	9.91	8.64	6.77	5.09
Machinery	21.46	26.92	22.12	21.72	21.46	18.08
Electical and Optical	37.59	27.09	34.64	27.39	39.34	43.58
Transport Equipment	6.22	11.19	7.54	9.45	4.61	4.96
Furniture	2.36	1.87	5.18	3.79	1.58	3.39

Table 6.2: Sample characteristics

- Because the breadths of patents are not the same according to the different offices, we weight the cited patents by their respective weights calculated with patent family data<sup>5</sup>.
- In order to identify correctly patents that represent the state of the art upon which new patents are built, only the first application of a patent is considered. Indeed, if a US patent cites another US patent that is a second application of a German one, the relevant knowledge comes from Germany and not from the US. This is why for each patent that is not a first application, we consider information related to its prior application.
- Finally, we retain all patents belonging to a patent family with at least one patent of this family filed at one of the seven selected offices.

Taking all these criteria into account, we extract all patents published from 1978 to 2003 and filed by inventors of the G5 countries. Cited patents are also classified by geographic origin (the country of inventor), time (year of publication) and technological field. Finally the cited sample encompass 4 746 555 patents.

The number of citations between the cited and the citing cohort can be computed following the characteristics outlined above. An example of counting process is provided in Figure 6.1 for French

<sup>5</sup>Weights are provided in chapter 4.

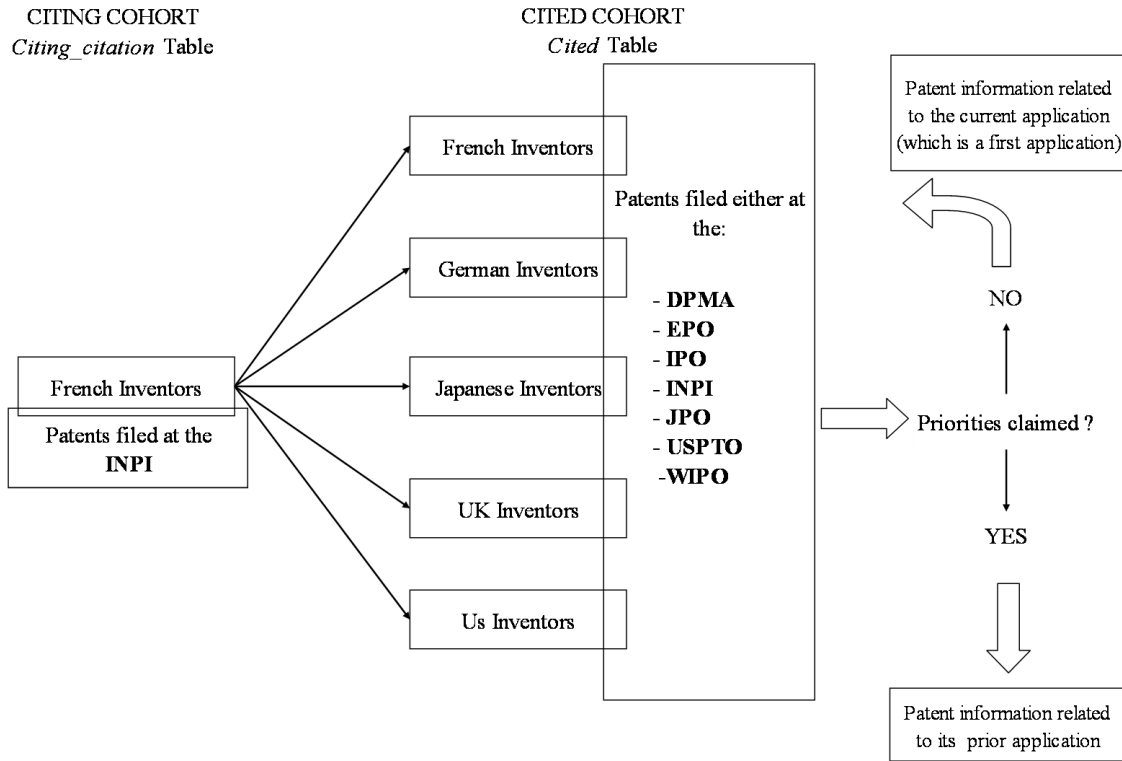


Figure 6.1: Counting citations process

citing inventors. The descriptive statistics of the 15 035 841 citations identified and used for regression are provided in Table 6.2 together with the citing and cited patent characteristics. The data consists in one observation for each feasible combination of  $l, i, t, T$  for each  $L$ . For the cited patents we have 26 years  $t$  (with the exception of the UK and Japan, for which, due to data constraint, we have respectively 21 and 19 years), 5 countries  $l$  and 13 technological field  $i$ . For the citing year  $T$  we have 29 years, except again for the UK (24 years) and Japan (19 years), 13 technological fields  $i$  and 1 country  $L$ . Hence the number of observations is  $n = 26 \times 13 \times 5 \times 29 = 49\,010$  for Germany, France and the US,  $n = 21 \times 13 \times 5 \times 24 = 32\,760$  for the United-Kingdom and  $n = 16 \times 13 \times 5 \times 19 = 19\,760$  for Japan. In each data set some patents are never cited, they represent a total of 105 208 observations (52.7 % of all feasible combinations).

## 6.4 Estimation of the patent citation function

### 6.4.1 The citation function

The required parameters for the construction of our knowledge stocks (equations (6.1) and (6.2)) are estimated through the citation function generalized in equation (6.4). The number of citations to a specific group of cited patents by a specific group of citing patents depends on the size of each cohort. It is thus necessary to control the size effect by looking at the probability of citation. The probability of citation is given by:

$$p(k, K) = \frac{C(k, K)}{N_k N_K}$$

With  $p(k, K)$ , the ratio of the number of citations  $C(k, K)$  to the product of the potential number of cited patents  $N_k$  and citing patents  $N_K$ . We assume that citations represent the flows of knowledge from inventors publishing the patent cohort  $k$  to inventors applying the patent cohort  $K$ . For the potentially cited patent in cohort  $k$  the considered attributes are:

- $i, l, t$ , the technology field, the location, and the publication year of a specific cited patent cohort.
- $l$ , the origin of the cited patents.
- $i$ , the technology class of the cited patents.

For the potentially citing patent in cohort  $K$  the considered attribute is:

- $T$ , the application year of the citing patents.

Two reasons explain that the only attributes of the citing cohort is the application years. At first as we implement only one estimation by citing inventor location, the inventor location  $L$  is fixed. Second, as we allow only intra-technological citations, the fixed effect  $i$  of the cited cohort  $k$  is enough to take differences in the composition of technological field into account. Hence, a treatable formulation of the model, in which the various different effects enter as multiplicative parameters is given by :

$$p_{i,l,t,T} = \frac{C_{i,l,t,T}}{N_{i,l,t} N_{i,T}} = \alpha_{i,l,t} \alpha_i \alpha_l \alpha_T \cdot e^{-(\beta_1) \beta_{1,i} \beta_{1,l} (T-t)} (1 - e^{-\beta_2 (T-t)}) \quad (6.5)$$

We have defined one citation function by country of citing inventor  $L$ , this is why  $L$  index doesn't



appear into the equation<sup>6</sup>. The variables entering into the equation are the following:

- $\alpha_{i,l,t}$  is the most important variable and represents the productivity of patent cohort  $k$  from  $K$  point of view, i.e. if a patent embodies useful knowledge we expect that it will be more likely cited by subsequent patent.
- $\alpha_T$  is the citing parameter that takes into account changes in the behaviors of citing agents.

Parameters  $\alpha_i$  and  $\alpha_l$  are multiplicative factors, like the fixed effects in a simple linear model and represent the extend to which a patent belonging to the technology group  $i$  or country  $l$  is more or less likely to be cited, with respect to the based group, by subsequent patents. Thus :

- $\alpha_i$  is the frequency of citation within each technology group, this variable captures the size effect of the technology group. Since we allow only citations to be intra-technological, the parameters  $\alpha_i$  represents only the composition effect of the technology group employed<sup>7</sup>.
- $\alpha_l$  is the frequency of citation from the country  $L$  to the country  $l$ , measuring the pattern of the geographic localization of patent citations. This parameter represents the closeness between two countries or the extent to which an inventor located in country  $L$  is likely to cite foreign inventors located in country  $l$ .
- $\beta_1$  is the rate of obsolescence
- $\beta_2$  is the diffusion parameter
- $\beta_{1,i}$  or  $\beta_{1,l}$  represent variations in  $\beta_1$  with respect to  $i$  and  $l$ , thus differences in the rate of decay across categories of cited patents. Higher value for  $\beta_{1,i}$  or  $\beta_{1,l}$ , with respect to the base group, means a higher obsolescence rate. Higher value leads to a faster speed of diffusion  $L_{Mod} \simeq \frac{1}{\beta_1}$ . We should expect that knowledge diffuse faster from domestic localization than from foreign localization.

The figure 6.2 summarizes the estimated parameters from the citation function (6.5).

---

<sup>6</sup>Indeed, the true specification of this ratio should be  $p_{L,i,l,t,T} = \frac{C_{L,i,l,t,T}}{N_{L,i,l,t}N_{L,i,T}}$ , but in order to avoid confusion we have omitted the citing location index  $L$ .

<sup>7</sup>Jaffe & Trajtenberg (1998) [86] have estimated the probability to cite a patent matching the same patent technology class, and find that the probability to cite a patent in the same technology class is 99 times higher than the one to cite a random patent in some other class. This means that a technology group based on few or close patent class are more likely to receive citations.

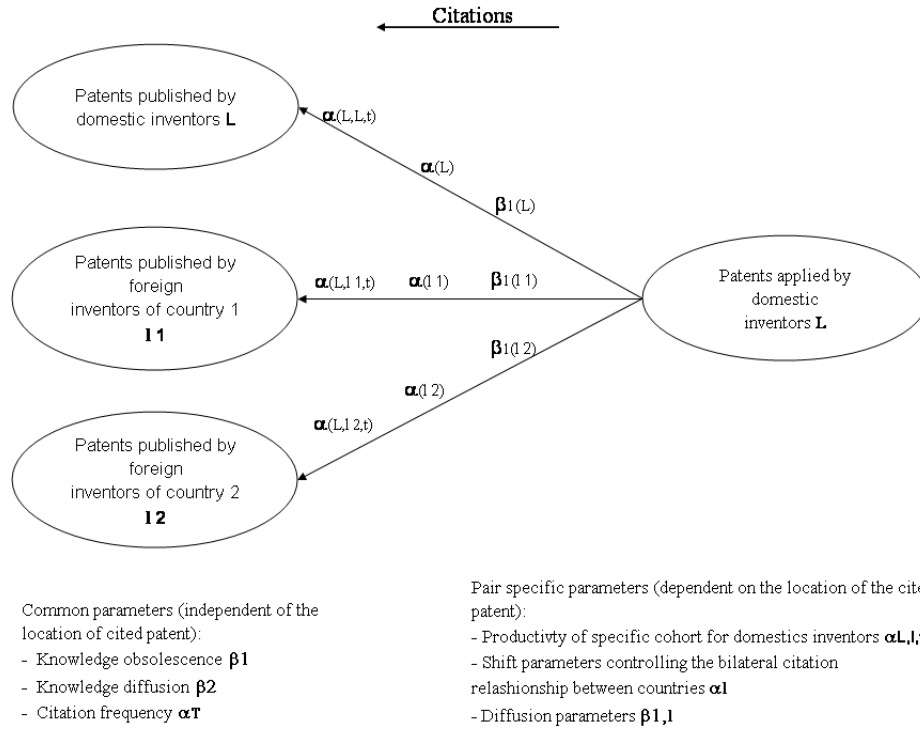


Figure 6.2: Estimated parameters from the citation function

### 6.4.2 Econometric issues and interpretations

The equation to be estimated is :

$$p_{i,l,t,T} = \alpha_{i,l,t} \alpha_i \alpha_l \alpha_T \cdot e^{-(\beta_1) \beta_{1,i} \beta_{1,l} (T-t)} (1 - e^{-\beta_2 (T-t)}) + \epsilon_{i,l,t,T} \quad (6.6)$$

Because the model can not converge with separate  $\alpha$  parameters for each cited  $t$  and citing  $T$  years, we consider four years interval for citing years  $T$ .  $\alpha(T)$  is thus constant over  $T$  within these periods, but allows to vary between each period.

The estimation of the citation function have to be carried out including base values for each attributes selected in order to deal with several identification problems. For each effect (relating to the attributes:  $i, l, T$  &  $i, l, t$ ) one group is omitted from estimation and the multiplicative parameter is constrained to unity. Hence, the estimates of each particular  $\alpha_k$ ,  $\alpha_K$  and  $\beta_{1,k}$  represents a proportionality factor measuring to what extent the considering patents are more or less likely to be cited over time relatively to patents belonging to the base category. These base values are the “food, beverages and tobacco” for technological field  $i$ , the domestic country for geographical location  $l$ , 1978 for the

publication year  $t$  of the cited patents and 1978-1980 for the filed period  $T$  of the citing patents.

For instance, if we consider the technology field of cited patent :  $\alpha_i$ , the estimate  $\alpha(i = Chemical) = 1.4$  means that the likelihood that a patent in the technological field “Chemical” receives a citation is 40% higher than for a patent in the base group, “Food, Beverages & Tobacco”. The mechanism is the same for the fixed effects related to location, for example, for the US citing cohort, the value  $\alpha(l = france) = 0.5$  means that the probability of a patent granted by French inventors to be cited by an US patents is 50% less than a patent filed by US inventors .

Concerning the obsolescence parameters a common value is estimated, represented by  $\beta_1$ , and the other parameters  $\beta_{1,i}$  or  $\beta_{1,l}$ , represent the deviation to the base groups, as for  $\alpha$  parameters, however their interpretations are not done in the same way. Hence, the case  $\beta_1(i = Chemical) = 1.29$  does not means that patent belonging to this field is 0.29% more likely to depreciate faster. A value of  $\beta_1(i = Chemical)$  higher than unity just means that patents in this field become obsolete faster and that its modal lag of citations is higher.

In this model, the null hypothesis of no effect corresponds to parameters values of unity rather than zero (except for  $\beta_1$  and  $\beta_2$ ). Equation (6.6) is estimated using a weighted non-linear least-squares procedure. Each observation is weighted by  $\sqrt{N_K N_k}$ . This weighting scheme should take care of the possible heteroskedasticity, since the observations correspond essentially to “grouped data”, that is, each observation is an average, computed by dividing the number of citations by  $N_K N_k$ . In general, this weighting scheme improves the fit of the model and get the right standard errors, but does not alter the parameter estimates materially.

### 6.4.3 Results

The results for main estimates of the five regressions, one for each selected country, are presented in Table 6.3 and 6.4 and all productivity parameters are provided in appendix<sup>8</sup>. Each model has 1 667 parameters<sup>9</sup> including a base value of  $\beta_1$  and  $\beta_2$ ; 1628 productivity parameters for  $\alpha$ ; 4 cited countries for  $\alpha$ ; 12 technological fields for  $\alpha$ ; 7 citing time period for  $\alpha$ ; 4 cited countries for  $\beta_1$ ; 12 technological fields for  $\beta_1$ . Globally, the models fit the data reasonably well, Adjusted R-square range from 0.64 for France to 0.90 for the US.

---

<sup>8</sup>Due to presentation issues the estimates are provided in graphical forms. The S.E. are available on request.

<sup>9</sup>With the exception of the United Kingdom and Japan.

Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

	Germany		France		United Kingdom	
Adjusted R-squared	0.791		0.617		0.666	
Std. error of regression	0.008		0.010		0.015	
Number of observations	23567		26199		18124	
	Estimates	S.E	Estimates	S.E	Estimates	S.E
<b>Citing year effect</b>						
1978-1979	1	n.a.	1	n.a.	.	.
1980-1983	0.863	0.012	0.941	0.038	1	n.a.
1984-1987	0.912	0.009	0.989	0.017	0.985	0.063
1988-1991	0.985	0.008	0.947	0.015	0.950	0.031
1992-1995	0.992	0.007	0.990	0.013	1.035	0.029
1996-1999	1.008	0.006	1.053	0.014	1.219	0.024
2000-2003	1.082	0.006	1.058	0.017	0.846	0.109
2004-2006	0.899	0.008	0.896	0.014	0.973	0.067
<b>Cited inventor country code</b>						
Germany	1	n.a.	0.199	0.006	0.058	0.003
France	0.071	0.003	1	n.a.	0.069	0.004
United Kingdom	0.045	0.003	0.238	0.006	1	n.a.
Japan	0.053	0.004	0.092	0.005	0.061	0.004
United States	0.069	0.003	0.198	0.006	0.138	0.007
<b>Technological field</b>						
Food, Beverages & Tobacco	1	n.a.	1	n.a.	1	n.a.
Textile	3.499	0.161	3.680	0.115	3.977	0.402
Wood	3.454	0.125	2.476	0.074	10.374	2.275
Pulp, Paper & Publishing	3.457	0.145	2.329	0.096	3.464	0.343
Petroleum and Nuclear	5.138	0.158	7.628	0.185	6.442	0.822
Chemical	4.573	0.190	6.784	0.172	2.763	0.226
Rubber and Plastic	4.055	0.126	3.012	0.100	4.522	0.339
Non-metallic	6.867	0.116	3.026	0.096	3.542	0.313
Metal	3.972	0.119	3.125	0.105	6.723	0.476
Machinery	2.917	0.065	1.650	0.056	1.864	0.100
Electrical and Optical	4.348	0.086	2.985	0.086	3.159	0.150
Transport Equipment	10.705	0.177	5.163	0.136	4.692	0.287
Furniture	2.142	0.126	3.112	0.117	3.673	0.224
<b>Beta 1</b>	0.302	0.004	0.274	0.002	0.244	0.005
<b>Beta 2</b>	0.014	4.12E-04	0.016	2.77E-04	0.011	4.13E-04
<b>Rate of obsolescence :</b>						
<i>by cited inventor country code</i>						
DE	1	n.a.	0.611	0.013	0.566	0.022
FR	0.452	0.014	1	n.a.	0.590	0.026
UK	0.494	0.025	0.545	0.010	1	n.a.
JP	0.522	0.023	0.475	0.017	0.622	0.022
US	0.510	0.016	0.572	0.010	0.618	0.018
<i>by technological field</i>						
Food, Beverages & Tobacco	1	n.a.	1	n.a.	1	n.a.
Textile	1.391	0.040	1.307	0.026	0.936	0.060
Wood	1.149	0.026	0.840	0.018	1.138	0.132
Pulp, Paper & Publishing	1.244	0.033	1.062	0.030	0.785	0.051
Petroleum and Nuclear	1.321	0.026	1.753	0.027	1.116	0.094
Chemical	1.604	0.040	1.800	0.028	1.306	0.059
Rubber and Plastic	1.193	0.023	1.226	0.026	1.201	0.048
Non-metallic	1.117	0.012	1.170	0.024	0.937	0.049
Metal	1.144	0.021	1.277	0.027	1.249	0.047
Machinery	1.226	0.017	1.347	0.029	1.128	0.033
Electrical and Optical	1.494	0.018	1.815	0.032	1.542	0.039
Transport Equipment	1.516	0.015	1.525	0.025	1.232	0.042
Furniture	1.178	0.043	1.361	0.033	0.967	0.036

Table 6.3: Regression results

Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

	Japan		United States	
Adjusted R-squared	0.891		0.853	
Std. error of regression	0.014		0.040	
Number of observations	7366		29952	
	Estimates	S.E	Estimates	S.E
<b>Citing year effect</b>				
1978-1979	1	n.a	1	n.a
1980-1983	0.938	0.018	0.758	0.014
1984-1987	1.061	0.017	0.803	0.010
1988-1991	0.952	0.017	0.838	0.007
1992-1995	0.776	0.017	1.080	0.005
1996-1999	0.318	0.017	1.119	0.005
2000-2003	.	.	1.121	0.005
2004-2006	.	.	0.668	0.006
<b>Cited inventor country code</b>				
Germany	0.011	0.012	0.053	0.005
France	0.011	0.014	0.049	0.005
United Kingdom	0.005	0.016	0.113	0.005
Japan	1	n.a	0.114	0.006
United States	0.021	0.014	1	n.a
<b>Technological field</b>				
Food, Beverages & Tobacco	1	n.a	1	n.a
Textile	2.645	0.259	3.455	0.087
Wood	2.073	0.307	1.324	0.088
Pulp, Paper & Publishing	14.104	0.409	4.063	0.134
Petroleum and Nuclear	2.403	0.262	5.905	0.169
Chemical	4.309	0.324	6.163	0.146
Rubber and Plastic	1.847	0.256	3.524	0.099
Non-metallic	3.009	0.294	3.112	0.114
Metal	2.981	0.297	2.742	0.105
Machinery	1.446	0.139	1.922	0.057
Electrical and Optical	3.315	0.151	6.440	0.093
Transport Equipment	4.956	0.366	4.824	0.143
Furniture	3.415	0.259	3.968	0.109
<b>Beta 1</b>	0.622	0.022	0.177	0.002
<b>Beta 2</b>	0.049	2.92E-03	0.043	1.19E-03
<b>Rate of obsolescence :</b>				
<i>by cited inventor country code</i>				
DE	0.561	0.427	0.599	0.041
FR	0.558	0.488	0.652	0.052
UK	0.405	1.093	0.578	0.019
JP	1	n.a	0.753	0.029
US	0.706	0.309	1	n.a
<i>by technological field</i>				
Food, Beverages & Tobacco	1	n.a	1	n.a
Textile	0.905	0.055	0.818	0.016
Wood	1.033	0.094	0.785	0.041
Pulp, Paper & Publishing	1.289	0.023	1.187	0.028
Petroleum and Nuclear	0.910	0.063	1.442	0.028
Chemical	1.078	0.050	1.295	0.021
Rubber and Plastic	0.903	0.078	0.931	0.019
Non-metallic	1.001	0.060	1.048	0.028
Metal	1.009	0.062	0.980	0.027
Machinery	0.962	0.057	1.056	0.023
Electrical and Optical	1.023	0.029	1.552	0.015
Transport Equipment	1.184	0.053	1.273	0.026
Furniture	0.908	0.043	1.017	0.020

Table 6.4: Regression results (continued)

**Domestic obsolescence rate** The base value of  $\beta_1$  represents the rate of decay of domestic knowledge for the technological field “Food, beverages and Tobacco”. One can note that estimated values are quite similar for the different European offices, ranging between 0.244 and 0.3, suggesting a median lag of about 2 or 3 years, while the US value is lower, leading to higher modal lag of about 5 years (this result is in line with previous study which uses USPTO citation data, see Jaffe & Trajtenberg 2002 [87]). Concerning Japan,  $\beta_1$  is quite high, around 0.6, corresponding to a modal lag of 1.6 years.

However, the global obsolescence rate has to be considered together with the other technological obsolescence values  $\beta_{1,i}$ , indeed most of them have a lower estimate than unity, reducing the base value of  $\beta_1$ , thus increasing the modal lag.

**Technological fields** Two types of variations relative to the technological fields are considered in the model : variation in the fixed effect and in the obsolescence parameter. The fixed effect parameter of technological field only reflects a composition effect, a technological field based on high number of IPC classes is more likely to be cited than a field based on few number of IPC classes. For instance, Chemical encompass 74 IPC class and Food, Beverages & Tobacco 22, this composition effect is reflected in the estimated results, Chemical field presents a value equal to  $\alpha(\text{chemical}) = 4.5$  for German case, means that patents in this field get on average five as many citations as those in the base field.

Concerning technology diffusion, it’s interesting to note the technologies which are represented by the larger part of patents present the higher diffusion rate, thus the three top technologies are Chemical, Electrical & Optical and Transport Equipment. Back to the German chemical field, the large initial citation advantage,  $\alpha(\text{chemical}) = 4.5$ , of this field fades rather quickly over time due the large value of its obsolescence rate  $\beta_1(\text{chemical}) = 1.604$ . Thus, the fixed effect and the obsolescence parameter have to be considered simultaneously. A high fixed effect associated with a high obsolescence means that the group considered is much more highly cited during the first years after publication, however, due to its faster obsolescence, it is actually less cited in later years than those in the base group. Figure 6.5 and 6.6 given in appendix, present graphically the effects of these parameter differences. Each figure on the left side presents the estimated citation functions for citing country, with the different lines within each figure corresponding to the different cited technology fields.

**Cited Countries** Concerning the country fixed effect, the based group is the domestic country. Domestic based value always dominates the other countries, meaning that there is a strong pattern toward domestic country. The gaps with respect to foreign countries are quite important in all regressions.

They are about 90% for all with the exception of France which presents gaps around 80% (recall that an estimates equal to  $\alpha(\text{France}) = 0.071$  for Germany regression means that a German inventor is 93 % more likely to cite a German inventor than a French one).

These results are in line with those of chapter 5, in which we found that a large majority of citations are done toward patents belonging to the same office. Here, we do not consider citations between offices but between inventors, since we have retained only first patent application there exists a strong similitude between these two concepts. Concerning European countries, Germany and France benefit the most from other European countries but the differences with other countries are very small. On the contrary, it seems that the United Kingdom does not present the same localization effect, knowledge coming from Japan is as important as the one coming from Germany. For these European countries, Japan is the country which generates the less flow of knowledge. On the opposite, the US seems to benefit more from Japan than from European countries. Finally Japan is an exception as it does not seem to benefit from knowledge coming form foreign countries.

These results provide a new look on the geographic pattern of knowledge flows, whereas the main-stream literature suggests that knowledge flows decrease with geographic distance (see chapter 2), this is clearly not the case here, as farthest countries create as much knowledge opportunity as neighboring countries. This finding can be explained by the control of knowledge productivity we have implemented. Less useful knowledge do not diffuse as far as useful knowledge, thus geographic distance can be seen as a kind of filter through which less useful knowledge is abandoned. In this specification, by controlling productivity (which is equivalent to retain only knowledge with equal productivity), we see that the geographic pattern disappear and the remaining differences between countries can be explained by country pair specific relationships.

Dominance of domestic knowledge is also evident in the  $\beta_1$  parameter. All parameters,  $\beta_{1,l}$ , are below the base value meaning that knowledge diffuses more slowly from abroad than from home location, the modal lag is thus higher for foreign knowledge. This effect is reinforced by the fact that domestic inventors are more likely to be cited by domestic inventors at every lag ( $\alpha_{l \neq L} < 1$ ). The figures 6.5 and 6.6 <sup>10</sup> highlight the supremacy of domestic country both in lag and in intensity. Japanese inventors cite almost exclusively JPO patent although French inventors are more likely to cite foreign patents.

---

<sup>10</sup>Graphs on the right side of the figures present the estimated citation functions for citing countries. The different lines within each figure corresponding to the different cited inventor location.

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

---

Modal Lag	Citing				
	GE	FR	UK	JP	US
Cited					
GE	3.3	6.0	7.2	2.9	9.5
FR	7.3	3.6	7.0	2.9	8.7
UK	6.7	6.7	4.1	4.0	9.8
JP	6.3	7.7	6.6	1.6	7.5
US	6.5	6.4	6.6	2.3	5.7

---

Table 6.5: Modal Lag of citations

We present the estimated modal lag in function of the citing and cited country in table 6.5. The diagonal elements are the smallest entry in each row and column, confirming previous figures. Modal citation lags are noticeably shorter for domestic citations, relatively to citations from others. There is a symmetric variation of estimated from the diagonal elements both in row and column for European countries. Lag of UK Cited patents by French is similar with Lag of cited FR patent by English. This is not true for Japan and the US, Japan cites more quickly than it been cited and that is the contrary for the US.

Our results are finally in line with those of Jaffe & Trajtenberg (1999) [86], the domestics pattern of citation decrease with time, as a combination of relative high  $\alpha$  and high  $\beta_1$  for domestic citation meaning that the initial probability is much higher, but it fades faster, letting the possibilities for foreign country to catch up. The results presented in table 6.6 investigates this issues and gives the probability of citation from various country relatively to domestic citation probability, for each cited country for the first and after ten years<sup>11</sup>. In all cases the relative citation frequency of the others countries is greater after 10 years than in the first year.

**Citing years** Citing year effect coefficients, serve principally as a control, they are relatively constant over time and do not present any particular trend for any countries (neither downward nor upward). Nevertheless, it seems that during the period 1996-2003 the number of citations reaches a pick.

**Productivity parameters** As noted before, we have estimated 1 628 parameters for each function with the exception of Japan where the sample stop in the year 1996 and the United Kingdom where it start, in 1982. The base group is the domestic country for the year 1978 for each technology (1983 for United Kingdom regression), thus parameter estimates greater than one means that this cohort is

---

<sup>11</sup>Jaffe and Trajtenberg retain the first and after the 20 years, our presents choice is directed by the fact that our modal lag is in mean lower than those find in their study. This difference arise because they only use the USPTO database.



Citing	Lag in years	Relative citation rate				
		GE-cited	FR-cited	UK-cited	JP-cited	US-cited
GE	1	1	0.09	0.06	0.07	0.09
	10	1	0.47	0.26	0.28	0.38
FR	1	0.23	1	0.29	0.11	0.24
	10	0.68	1	1.00	0.48	0.76
UK	1	0.07	0.08	1	0.07	0.16
	10	0.20	0.22	1	0.18	0.40
JP	1	0.02	0.02	0.01	1	0.03
	10	0.26	0.27	0.36	1	0.17
US	1	0.06	0.05	0.13	0.12	1
	10	0.12	0.10	0.27	0.19	1

Table 6.6: Pattern of domestic localization over time

more likely to be cited by future cohort than patent cohort published in 1978 by domestic inventors<sup>12</sup>. Complete set of productivity estimates are provided in appendix.

Figure 6.3 displays the results for productivity estimates for the Chemical technology field. For instance, if we focus on France, results show that UK patent productivity parameter in 1986 is higher than the base group, we assume that knowledge embodied in UK patents granted in 1986 embodies more useful knowledge for French inventors than in French patents published in 1978. It is interesting to notice that a same cohort  $k$  has very different impacts depending on the receiving country  $L$  considered. For instance, if we consider France and Germany as citing cohort, US cited cohorts have different impacts, indeed, for French inventors, the most useful cohort is the one granted in 1992 although it is the one granted in 1989 for German inventors. However, these two countries also share common characteristics, for instance the 1982 UK cohort is the less useful for inventors located in both countries.

Finally, it should be stressed that the estimates are fairly precise as standard errors are about more than one fifth for domestic patents and one third for foreign ones.

## 6.5 Construction of our knowledge stocks

We use parameters estimated below to construct our stocks of knowledge. In practice, we adapt equations (6.1) and (6.2) to the constraints raised by the estimations (about the specification of the

---

<sup>12</sup>Remember that the domestic productivity parameter was normalized to unity for  $t = 1978$  (for instance  $\alpha_{FR,1978} = 1$  in French regression, i.e. for  $L = Fr, l = Fr, t = 1978$  and  $i = Chemical$ ).

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator



Figure 6.3: Productivity parameters for Chemical technology

fixed effects). These constraints leads to the addition of the country fixed effects, whereas we have not added the technological fixed effects because citations are allowed to be international but no inter-technological. Specifically we use the following specification for domestic stocks:

$$Knowd_{L,T,i} = \sum_{t=0}^{T-1} \alpha_{i,L,t} \alpha_L . N_{L,t} e^{-(\beta_1)\beta_i\beta_L(T-t)} (1 - e^{-\beta_2(T-t)})$$

And the following one for foreign stocks :

$$Knowf_{L,T,i} = \sum_{l \neq L} \sum_{t=0}^{T-1} \alpha_{i,l,t} \alpha_l . N_{l,t} e^{-(\beta_1)\beta_i\beta_l(T-t)} (1 - e^{-\beta_2(T-t)})$$

For each country and each technology we construct separate stocks depending on the origin of knowledge, i.e. for each country we construct  $5 \times 13$  stocks (as many as the number productivity set estimated). Figure 6.4 plots the resulting stocks of knowledge from the G5 inventors in function of their origins. For European countries the importance of knowledge spillovers is highlighted by the relative high value of foreign stocks compared to their domestic stock. This is particularly true for France which appears to benefit as much from international knowledge than from its own knowledge while on the opposite, Japan and the US seem to benefit almost exclusively from their own knowledge and very few from international one. Concerning the foreign stocks, a common figure in all graphs is the importance of knowledge coming from the United States, indeed in all countries with the exception of Japan, US knowledge is one of the major sources for domestic inventors, suggesting that US plays the role of technological sourcing for European countries.

In the following we follow the spillovers literature by considering only one domestic and one foreign stock of knowledge, the foreign stock is just the sum of all stocks except the domestic one.

## 6.6 Empirical validations

In this section we turn on the main point of our analysis. We will, on one hand, establish the validity of the indicator elaborated for measuring knowledge stocks and on the other hand, we will validate the use of patent citations for the measurement of international knowledge spillovers. In order to test our indicator we implement two empirical validations :

- in a first time, we will explore the correlation of our indicator with the traditional input-based indicator: R&D stocks;

Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

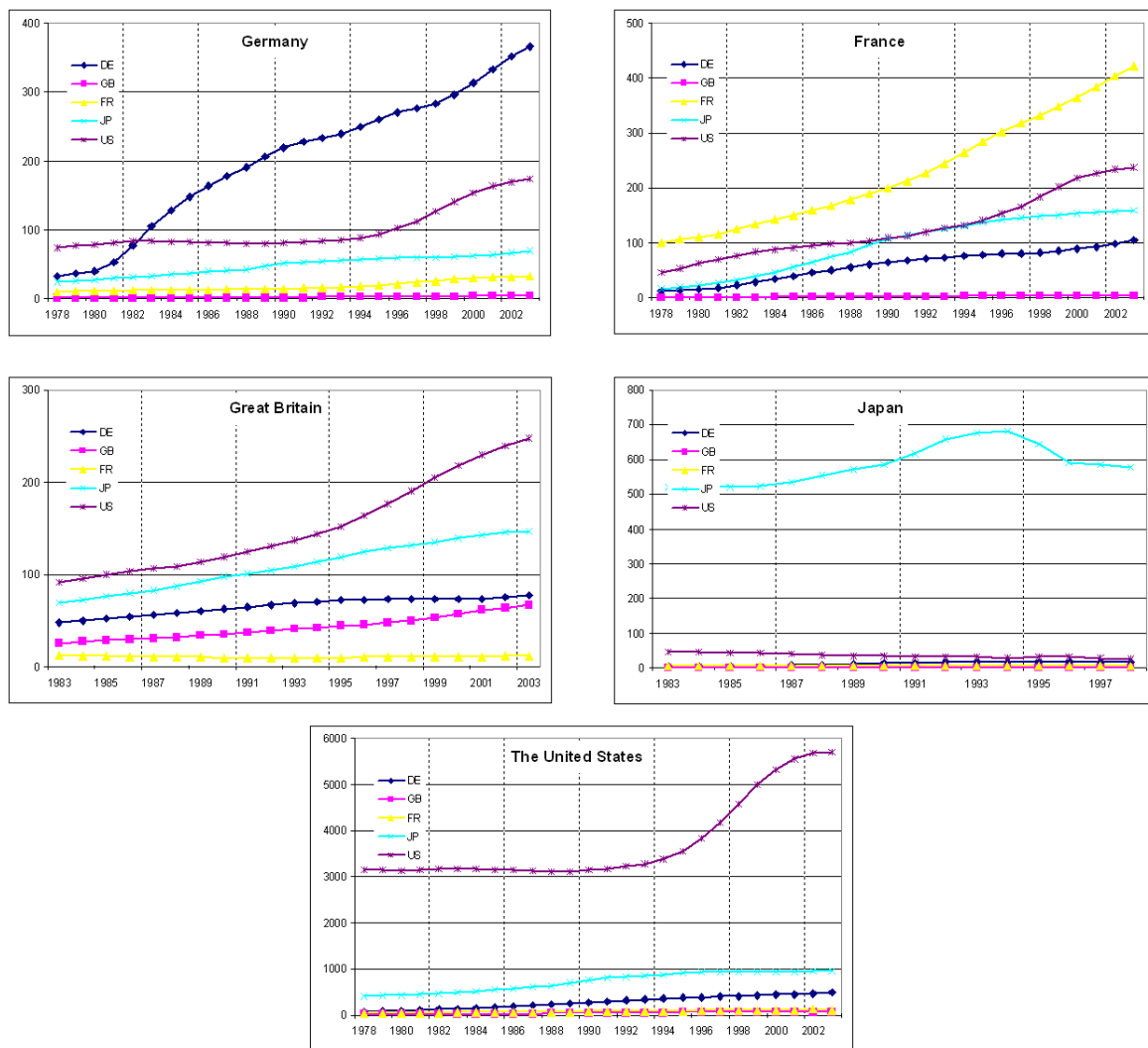


Figure 6.4: Detailed knowledge stocks for Chemical technology

- in a second time, we test the value of explanation of our indicator on productivity growth, see Park & Park (2006) [140] for the argumentation.

Before doing this, it is necessary to convert our knowledge stock based on IPC technological classification to the ISIC classification, which corresponds to the classification of R&D and TFP data. Since R&D and TFP data are not available for the 44 original sectors, we aggregate the 44 sectors into 13 sectors according to the availability of R&D and TFP data. The aggregation scheme is the same than the one used for technological classes (see Table 6.11).

### 6.6.1 Correlation with R&D expenditures data

In order to investigate the relationship between the traditional input based indicator and the one we have created, we performed a correlation analysis. R&D data are extracted from the OECD ANBERD (Analytical Business Enterprise Expenditure on R&D) database [134]. This data set provides industrial R&D data covering the period of 1987 up to 2006 for OECD member countries, however our analysis focuses on the G5 countries and the relevant data set is also restricted to these countries. Because our indicator is expressed in stock level, first step is to construct R&D stocks upon R&D expenditures. R&D stocks are constructed as usual way.

$$SRD_t = RD_t + (1 - \delta)SRD_{t-1}$$

Where the initial stock is given by the Perpetual Inventory Method (PIM) i.e.:

$$SRD_0 = RD_0 \frac{(1 + \lambda)}{(\delta + \lambda)}$$

with  $\lambda$  the growth rate of past R&D expenditures and  $\delta$  the obsolescence rate (we retain a rate of 15 %). Because we retain national R&D expenditures to build  $SRD_t$ , we only test the robustness of our domestic indicator i.e. *Knowd*.

To test the correlation between our full indicator and foreign R&D expenditures, we need to make some assumptions about the channel through which foreign R&D flow. Previous studies on this purpose are quite widespread and no consensus emerges, for this reason we do not test our full indicator with all R&D expenditures but only the domestic part. Finally, we have to keep in mind that R&D data concern only expenditures performed in the business enterprise sector although our indicator covers both public and business knowledge. Nevertheless, almost two third of the gross domestic expenditures on R&D

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

---

	Germany	United Kingdom	France	Japan	The United States
Correlation	0.713	0.595	0.770	0.852	0.703
Num Obs	169	187	221	91	187

---

Table 6.7: Correlation between R&D stocks and Patent stocks controlled by their productivities

is performed in business sector<sup>13</sup>. According this, results have to be interpreted in consequence, but they can yet give us a primary picture on our indicator.

Results are given in Table 6.7. For the five countries, all indicators seem to be highly correlated, the higher value being for Japan with a correlation of about 0.852 (but this country presents the lower number of observations), while the lowest is for the United Kingdom and is about 0.6, that is nevertheless quite reasonable.

### 6.6.2 Explanatory power on productivity growth

As mentioned before, one main reason for measuring the amount of knowledge is to analyze how knowledge contribute to the productivity growth or to the creation of new inventions. In this section, we test to what extend our indicator explains productivity growth at the industry level. Two specifications will be tested, one in Log-Level and one using an Error Correction Model, and for both of them we will test if taking international spillovers into account modifies globally the results or not. This set of estimations will allow us to test our global indicator (with domestic and foreign stocks).

If empirical studies on international spillovers differ on the role of the different spillovers channels, they agree to validate the role of international knowledge as an important factor of productivity growth. We follow the traditional approach for analyzing the contribution of knowledge to productivity growth by considering knowledge stocks as an additional factor of production. We based our estimated function on a simple Cobb-Douglas production function as defined in chapters 2 and 3. The model to be estimated is the following :

$$\log(Mfp_{L,i,t}) = \beta_d \log(Know_{L,i,t-1}) + \beta_f \log(Know_{F,L,i,t-1}) + \sigma_U U_t + \sigma_G G + \phi_L + \phi_i + \phi_t + \epsilon_{L,i,t} \quad (6.7)$$

- $\log(Mfp)$  is the logarithm of the total factor productivity,

---

<sup>13</sup>In 2008, rate are about 63.2 % for France, 69.9 for Germany, 77.9 for Japan, 64.1 for United Kingdom and 71.9 for the US.

- $\log(Know_d)$  is the logarithm of the domestic knowledge,
- $\log(Know_f)$  is the logarithm of the foreign knowledge,
- $U$  is equal to 1 minus the employment rate.  $U$  controls for changes in the business cycle effect and is considered as a proxy to the capacity rate of utilization,
- $G$  is a dummy that controls to the exogenous shock of the German reunification. Dummy takes 1 in 1991, and 0 otherwise,
- $\phi_i, \phi_L$  and  $\phi_t$  are respectively sector, country and time dummy,
- $L = FR, GE, JP, UK, US$  is a country index,
- $i = 1, \dots, 12^{14}$  is a sector index,
- $t = 1, \dots, 26$  is a time index ranging from  $t = 1978$  to 2003.

The parameters to be estimated are assumed to be constant across countries (Pooled panel form) and over time; they are defined as follow :  $\beta_d$  is the elasticity of MFP with respect to the domestic stock of knowledge,  $\beta_f$  is the elasticity of MFP with respect to the foreign stock of knowledge and  $\sigma_U$  the elasticity of MFP with respect to the capacity utilization rate. Data on MFP are taken from the EU-KLEMS database. MFP has been computed as the difference between the domestic output of industry and the weighted sum of the quantity of inputs (capital, labour and intermediate inputs), the weights are their respective share in the nominal value of output (see appendix 6.8.1 for detailed calculi). Data on unemployment rate are extracted from OECD Labour Force Statistics (Main Economic Indicator).

From the basic model, presented in equation (6.7), we derive an Error Correction Model that allows to separate short-term and long-term effects (Guellec & Pottelsberghe (2004) [69]). The resulting ECM equation is given by the following specification :

$$\begin{aligned} \Delta \log(Mfp_{L,i,t}) &= \alpha_d \Delta \log(Know_{L,i,t-1}) + \alpha_f \Delta \log(Know_{L,i,t-1}) + \theta \Delta \log(Mfp_{L,i,t-1}) \quad (6.8) \\ &+ \varphi \log(Mfp_{L,i,t-2}) + \beta_d \log(Know_{L,i,t-2}) + \beta_f \log(Know_{L,i,t-2}) + \sigma_U \Delta U_t + \phi_L + \phi_i + \phi_t + \epsilon_{L,i,t} \end{aligned}$$

---

<sup>14</sup>Previously 13 sectors were retained but due to data lacks on MFP, sample was reduced to 12 sectors.

**Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator**

Dependant variable :	Log MFP				
Regression	1	2	3	4	5
Log Knowd (-1)	0.071 *** (0.014)	0.121 *** (0.016)	0.131 *** (0.016)	0.029 * (0.012)	0.040 ** (0.020)
Log Knowf (-1)				0.250 *** (0.031)	0.242 *** (0.031)
German Reunification			0.007 (0.012)		0.006 (0.012)
Utilization capacity			0.626 *** (0.203)		0.471 ** (0.199)
Fixed effect	no	yes	yes	yes	yes
Adjusted R-squared	0.786	0.851	0.852	0.866	0.866
Number of observations	1177				

Table 6.8: Log Level estimation results

Where  $\Delta$  is the first difference operator. Long run elasticity with respect to domestic knowledge,  $Knowd_{L,i,t-1}$  is done thanks to the relation  $[-\beta_d/\varphi]$  (respectively  $[-\beta_f/\varphi]$  for the foreign stock). Estimation procedures are done in two steps. First we test our indicator in the log level form as specified in equation (6.7) and in a second step we estimate the error correction model that give us long run elasticities as specified in equation (6.8). Both sets of regressions are corrected for auto-correlation using the Cochrane-Orcutt transformation

**Log level** The results of estimating (6.7) are presented in Table 6.8. When international spillovers are withdrawn from the specification, the estimated coefficients of domestic knowledge range from 0.07 to 0.13 and are strongly significant, while this coefficient decreases with the introduction of foreign knowledge, then it is divided by 4. The estimated parameter of foreign knowledge is 0.24. The fact that foreign knowledge coefficient is higher than the domestic one is a classical result in international knowledge spillovers literature. The adjusted coefficient of determination, is relatively high and range from 0.79 to 0.87. Concerning the control variables, the dummy variable representing the German reunification does not appears with the expected sign but it is not significant. On the other side, the capacity utilization rate has a positive and significant impact on productivity confirming the assumption about the pro-cyclical characteristics of productivity.

**Error Correction Model** The results of the ECM specification are close to the log level ones (see Table 6.9). Both domestic and foreign knowledge effects are positives and significant. Long term elasticities of domestic knowledge range from 0.025 to 0.435 when domestic knowledge is the only



**Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator**

Dependant variable :	$\Delta\text{Log MFP}$				
Regression	1	2	3	4	5
$\Delta\text{Log Knowd} (-1)$	-0.014 (0.41)	0.088 *** (0.025)	0.139 *** (0.043)	0.084 * (0.045)	0.095 ** (0.046)
$\Delta\text{Log Knowf} (-1)$				0.124 (0.077)	0.130 * (0.079)
$\Delta\text{Log TFP}(-1)$	0.410 *** (0.041)	0.344 *** (0.046)	-0.350 *** (0.053)	-0.330 *** (0.056)	-0.374 *** (0.053)
$\text{Log TFP}(-2)$	-0.049 *** (0.006)	-0.057 *** (0.006)	-0.132 *** (0.019)	-0.142 *** (0.02)	-0.152 *** (0.022)
$\text{Log Knowd} (-2)$	0.001 ** (0.001)	0.026 *** (0.005)	0.057 *** (0.012)	0.034 *** (0.014)	0.034 ** (0.014)
$\text{Log Knowf} (-2)$				0.036 *** (0.013)	0.041 *** (0.014)
German reunification			0.013 (0.013)		0.014 (0.013)
$\Delta\text{Utilization capacity}$			0.595 *** (0.203)		0.650 *** (0.203)
Fixed effect	no	yes	yes	yes	yes
Adjusted R-squared	0.184	0.207	0.212	0.212	0.218
Number of observations	1089				
Long run elasticity					
Knowd	0.025	0.451	0.435	0.241	0.223
Knowf				0.255	0.270

Table 6.9: Error Correction Model estimations results

explicative parameter and from 0.22 to 0.24 when the foreign component is included in the regression. The foreign elasticity is quite stable on the two regressions with values of 0.25 and 0.27. These results are in line with the international literature on knowledge spillovers surveyed in the chapter 3 with the meta-analysis. Once again, foreign knowledge presents an elasticity higher than the domestic one but this difference is significantly lower than in log-level regression.

Regarding short term elasticities, values have a similar order of magnitude when compared to previous set of regression, but once again values are more homogeneous between both domestic and foreign components. Finally, the two control variables are very close to the log level regression, rate of utilization is positive and significant although it is not the case for the German reunification.

Globally, both specifications estimates (log-level and ECM) show that our indicator explains an important part of the productivity growth. Both national and international knowledge have a positive and significant impact on the productivity growth and the elasticities estimated are in line with the results find by the literature.

## **6.7 Conclusion**

In this chapter we create an input-based indicator of knowledge by using international patent and patent citation data. The indicator measures knowledge available to the inventor according to different geographic locations, so it takes into account international knowledge spillovers. The extent to which an inventor benefits from knowledge embodied in patents depends on the patents' technological opportunity, estimated with patent citation data. We used DPMA, EPO, INPI, IPO, JPO, USPTO and WIPO patent data from 1978 to 2006 to implement empirically our knowledge stock indicator for the G5 at 13 sectoral level.

One of the challenges of this chapter was to validate the use of patent citation data, from major patent offices, as an indicator of knowledge flows. This is particularly important because this issue has never been explored for patent citation data taken from DPMA, INPI, IPO and JPO. The validation was done by comparing our indicator with the R&D stock indicator, one of the most used, and by analyzing its explanatory power on productivity growth. Our indicator has passed both tests successfully, meaning that citations used are good proxies for knowledge flows. The indicator developed presents several advantages as it allows to address the following issues :

- First, the problem of patent heterogeneity in the construction of patent stocks.
- Second, the need for data at a more disaggregated level since it is based on patent data, available at very detailed levels.
- Third, it is a good measure of international spillovers, and it can be very useful in the search for knowledge spillovers.
- Finally, it gives some useful information on the time diffusion and the depreciation rate of domestic and foreign knowledge.

In this sense, this exercise allows the development of a clearer picture on the international diffusion of knowledge while avoiding the trap laid by the home advantage effect.

Regarding the patterns of international dissemination of knowledge, this chapter also provides some interesting results. First, domestic knowledge is predominant, it diffuses almost twice as fast as the foreign knowledge, although the probability of domestic citations compared with those from abroad decreases with time. Second, when foreign knowledge is controlled by its technological productivity, the geographic location effect seems to disappear.

Finally, one issue was not addressed in the present chapter but may nevertheless be relevant. In the search for international spillovers, the diffusion process estimated can be used as a weighted scheme of foreign R&D stocks. If R&D data are available we could just substitute our patent flows with R&D data.

## 6.8 Appendix

### 6.8.1 Multifactor productivity variables

MFP data are taken from the EU KLEMS database, construction of this indicator is extensively detailed in Timmer et al. (2007) [177] ; here the following paragraph we provide a summary of this section. Multifactor productivity measures: MFP, has been generated on value added basis according to a standard methodology developed by Jorgenson, Gollop and Fraumeni (1987) [94]. MFP is based on production possibility frontiers where industry gross output is a function of capital, labour, intermediate inputs and technology, which is indexed by time,  $t$ . Each industry, indexed by  $j$ , can produce a set of products indexed by  $i$  indicated by the production possibility set  $g$ . Each industry has its own production function and purchases a number of distinct intermediate inputs indexed by  $i$ , capital service inputs indexed by  $k$ , and labour inputs indexed by  $l$ . The production functions are assumed to be separable in these inputs, so that:

$$Y = g(Y_{ij}) = f(K_j, L_j, X_j, T)$$

where  $Y$  is value added,  $K$  is an index of capital service flow,  $L$  is an index of labour service flows and  $X$  is an index of intermediate inputs, which consists of the intermediate inputs purchased from the other domestic industries and imported products. Under the assumptions of constant returns to scale and competitive markets, the value of output is equal to the value of all inputs:

$$P_j^Y Y_j = P_j^K K_j + P_j^L L_j + P_j^X X_j$$

where  $P_j^Y$  denotes the price of output,  $P_j^X$  denotes the price of intermediate inputs,  $P_j^K$  denotes the price of capital services and  $P_j^L$  denotes the price of labour services. Under the standard assumption of profit maximizing behavior, competitive markets, such that factors are paid their marginal product, and constant returns to scale, we can define MFP growth ( $\Delta \log MPF$ ) as follows

$$\Delta \log MFP_j = \Delta \ln Y_j - \bar{v}_j^X \Delta \ln X_j - \bar{v}_j^K \Delta \ln K_j - \bar{v}_j^L \Delta \ln L_j \quad (6.9)$$

Growth of MFP is derived as the real growth of output minus a weighted growth of inputs where  $\Delta x = x_t - x_{t-1}$  denotes the change between year  $t-1$  and  $t$ , and  $\bar{v}_j$  with a bar denoting period averages and  $\bar{v}$  is the two period average share of the input in the nominal value of output. The value share of each input is defined as follows:

$$\bar{v}_j^X = \frac{P_j^X X_j}{P_j^Y Y_j}; \quad \bar{v}_j^K = \frac{P_j^K K_j}{P_j^Y Y_j}; \quad \bar{v}_j^L = \frac{P_j^L L_j}{P_j^Y Y_j}$$

The assumption of constant returns to scale implies  $v_j^X + v_j^K + v_j^L = 1$  and allows the observed input shares to be used in the estimation of MFP growth in equation (6.9). This assumption is common in the growth accounting literature (see e.g. Schreyer 2001).

### 6.8.2 R&D capital stocks

Data on R&D expenditure are taken from the OECD ANBERD (ed2009) database. This database presents data on research and development (R&D) expenditure performed in the business enterprise sector by industry according to the International Standard Industrial Classification (ISIC) revision 3.1 since 1987. R&D capital stocks,  $SRD_t$  are calculated following the perpetual inventory method. The stock at time  $t$  is equal to the new investment at time  $t$ ,  $RD_t$  plus the stock at time  $t-1$  minus depreciation  $\delta$ :

$$SRD_t = r_t + (1 - \delta)SRD_{t-1}$$

Then stock at time  $t$  can be expressed as the sum of past R&D investment:

$$SRD_t = RD_t + (1 - \delta)RD_{t-1} + (1 - \delta)^2RD_{t-2} + \dots + (1 - \delta)^tRD_0$$

Under the assumption of constant annual growth rate,  $\lambda$  of the past investment,

$$SRD_t = RD_t + (1 - \delta)\lambda RD_t + (1 - \delta)^2\lambda^2 RD_t + \dots + (1 - \delta)^t\lambda^t RD_t$$

Finally, if we reorganize the right side of previous equation, then we can get the knowledge stock at the base point  $t$  as follow :

$$SRD_t = \frac{1 + \lambda}{\delta + \lambda} RD_t$$

Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

6.8.3 Tables and figures

1	A01H	10	B09B	15	C06B	19	H01B	21	F16C	24	B23C	25	B44B	27	A45D	33	B60M	37	A61M	42	F01N	44	A63F
1	A21D	10	B09C	15	C06C	20	A01L	21	F16E	24	B23D	25	B65B	27	A47G	33	B61L	37	A61N	42	F01P	44	A63G
1	A23B	10	B29B	15	C06D	20	A44B	21	F16F	24	B23E	25	B65C	27	A47J	33	F21P	37	A62B	42	F02B	44	A63H
1	A23C	10	C01B	15	C08H	20	A47H	21	F16G	24	B23H	25	B65H	27	A47L	33	F21Q	37	B01L	42	F02E	44	A63J
1	A23D	10	C01C	15	C08G	20	A47I	21	F16H	24	B23I	25	B65I	27	B01B	33	G08B	37	B01B	42	F02F	44	A63K
1	A23E	10	C01D	15	C08H	20	B21K	21	F16M	24	B23P	25	B67C	27	D06F	33	G08G	37	C12M	42	F02G	44	A63L
1	A23J	10	C01F	15	C08J	20	B21L	21	F23K	24	B23Q	25	B68F	27	D06E	33	G10K	37	E06C	42	F02H	44	A63M
1	A23K	10	C01G	15	C10M	20	B21M	21	A62C	24	B24B	25	B69F	27	F23N	33	G21C	37	G21G	42	F02I	44	A63N
1	A23L	10	C02F	15	C11B	20	B25B	22	B01D	24	B24C	25	C13D	27	F24B	33	G21D	37	G21K	42	F02P	44	A63O
1	A23P	10	C05B	15	C11C	20	B25C	22	B04C	24	B25D	25	C13G	27	F24C	33	H01T	37	H05G	42	F16J	44	B68G
1	A23Q	10	C05C	15	C14C	20	B25E	22	B05B	24	B25J	25	C13H	27	F24D	33	H02H	38	F15C	42	G01P	44	C06F
1	C12C	10	C05D	15	C23F	20	B25G	22	B61B	24	B26F	25	C14B	27	F25C	33	H02M	38	G01B	42	G05D	44	F23Q
1	C12E	10	C05E	15	C23G	20	B25H	22	B66G	24	B27B	25	C23C	27	F25D	33	H05C	38	G01C	42	G05E	44	G10B
1	C12G	10	C05F	15	D01C	20	B26B	22	B66G	24	B27C	25	D01B	27	F25E	34	H05C	38	G01D	42	G05F	44	G10C
1	C12H	10	C07B	15	F42B	20	B27G	22	B66C	24	B27F	25	D01D	28	B47J	34	B81C	38	G01E	42	B60V	44	G10D
1	C12J	10	C07C	15	F42D	20	B44C	20	B66D	24	B27J	25	D01G	28	B41K	34	G11C	38	G01F	42	B60W	44	G10E
1	C13F	10	C07F	15	G03C	20	B65F	22	B66E	24	B28D	25	D01H	28	B43M	34	H01C	38	G01H	42	B60X	44	G10F
1	C13J	10	C07G	16	D01F	16	B82B	17	C10F	24	B30B	25	D02G	28	G02F	34	H01F	38	G01M	42	B60Y	44	G10G
1	C13K	10	C08B	17	A46C	20	C23D	22	C12L	24	E21C	25	D02H	28	G03G	34	H01G	38	G01N	42	B61F	44	G10H
2	A24B	10	C08C	17	B29C	20	C25D	22	F16G	25	A21C	25	D02J	28	G05F	34	H01J	38	G01O	42	B61H	44	G10I
2	A24D	10	C08E	17	B29D	20	E01D	22	F22D	25	A22B	25	D03C	28	G06C	34	H01L	38	G01S	42	B61J	44	G10J
2	A24F	10	C08G	17	B60C	20	E01F	22	F23B	25	A22C	25	D03D	28	G06D	35	G09B	38	G01W	42	B61K	44	G10K
3	D04D	10	C08J	17	B69D	20	E02G	22	F23C	25	A23N	25	D03J	28	G06E	35	G09B	38	G12B	42	B62C	43	B62C
3	D04G	10	C08K	17	B67D	20	E03B	22	F23D	25	A24C	25	D04B	28	G06F	35	H01P	39	G12B	42	B62D	43	B62D
3	D04H	10	C08L	17	E02B	20	E03C	22	F23E	25	A41H	25	D04C	28	G06G	35	H01Q	39	G01K	42	B62E	43	B62E
3	D06C	10	C09B	17	F16L	20	E03D	22	F23F	25	A42C	25	D05B	28	G06H	35	H01R	39	G05B	42	B62F	43	B62F
3	D06M	10	C09C	17	H02G	20	E05B	22	F23J	25	A43D	25	D05C	28	G06I	35	H02J	39	G08C	42	B62G	43	B62G
3	D06N	10	C09D	18	B24D	20	E05C	22	F23K	25	B01F	25	D06B	28	G06M	35	H03B	40	G02B	42	B62H	43	B62H
3	D06N	10	C09E	18	B28B	20	E05D	22	F23L	25	B02B	25	D06G	28	G06N	35	H03C	40	G02C	42	B62I	43	B62I
3	D06P	10	C10B	18	B28C	20	E05E	22	F23M	25	B02C	25	D06H	28	G06O	35	H03D	40	G02D	42	B62J	43	B62J
3	D06Q	10	C10C	18	B29B	20	E05F	22	F23N	25	B03B	25	D06I	28	G06P	35	H03E	40	G02E	42	B62K	43	B62K
4	A41B	10	C10D	18	C03B	20	E06B	22	F24H	25	B03C	25	D21B	28	G07C	35	H03F	40	G02F	42	B62L	43	B62L
4	A41C	10	C10E	18	C03C	20	E06B	22	F24I	25	B03C	25	D21C	28	G07C	35	H03G	40	G02F	42	B62M	43	B62M
4	A41D	10	C10F	18	C03B	20	F01K	20	F24J	25	B03D	25	D21D	28	G07D	35	H03H	40	G02F	42	B62N	43	B62N
4	A41F	10	C12S	18	E04B	20	F15D	22	F24K	25	B03D	25	D21E	28	G07E	35	H03H	40	G02F	42	B62O	43	B62O
4	A43B	10	C25B	18	E04C	20	F16B	22	F24L	25	B05C	25	D21G	28	G07F	35	H03M	41	G04B	42	B62P	43	B62P
5	A43C	10	F17C	18	E04D	20	F16P	22	F24M	25	B06B	25	E02D	28	G08D	35	H04J	41	G04D	42	B62Q	43	B62Q
5	A43C	10	F17C	18	E04E	20	F16S	22	F24N	25	B07B	25	E02E	28	G08D	35	H04K	41	G04E	42	B62R	43	B62R
5	B68B	10	F17D	18	E04F	20	F16T	22	F24O	25	B07C	25	E21B	28	G10L	35	H04L	41	G04G	42	B62S	43	B62S
5	B68B	10	F25J	18	G21B	20	F17B	22	F24P	25	B08B	25	E21C	28	G11B	35	H04M	42	B60R	42	B60R	43	F02C
6	B27D	10	G21F	19	B21C	20	F22B	22	G01G	25	B21B	25	E21F	28	H03K	35	H04Q	42	B60S	42	B60S	43	F02K
6	B27H	10	A01N	19	B21G	20	F22B	22	H06F	25	B21C	25	F04F	28	H03L	35	H04Q	42	B60T	42	B60T	43	F02K
6	B27M	12	B21K	19	B21H	20	F22C	22	H06G	25	B21D	25	F04G	28	H03L	35	H04Q	42	B60U	42	B60U	43	F02K
6	B27N	13	A61K	19	C21B	21	G21H	23	A01C	25	B31B	25	F16N	29	H02P	36	H03J	42	B60V	42	B60V	43	F02K
6	E04G	13	A61P	19	C21C	21	B23F	23	A01D	25	B31C	25	H05H	29	H02P	36	H04H	42	B60W	42	B60W	43	A44C
7	B41M	13	C07D	19	C21D	21	F01B	23	A01E	25	B31F	26	F41A	30	H01H	36	H04N	42	B60X	42	B60X	43	A44C
7	B42D	13	C07H	19	C22B	21	F01C	23	A01F	25	B31F	26	F41A	30	H01H	36	H04N	42	B60Y	42	B60Y	43	A44C
7	B42E	13	C07I	19	C22C	21	F01D	23	A01J	25	B41B	26	F41B	30	H02B	36	H04S	42	B60Z	42	B60Z	43	A44C
7	B42F	13	C07J	19	C22E	21	F03B	23	A01K	25	B41C	26	F41C	30	H01M	37	H04S	42	B60A	42	B60A	43	A44C
7	D21C	13	C12N	19	C25C	21	F03C	23	A01M	25	B41D	26	F41E	30	F21H	37	A61G	42	B60B	42	B60B	43	A44C
7	D21H	13	C12P	19	C25F	21	F03D	23	B27L	25	B41E	26	F41G	30	F21I	37	A61H	42	B60C	42	B60C	43	A44C
7	D21J	13	C12Q	19	C30B	21	F03G	24	B27M	25	B41F	26	F41H	30	F21J	37	A61I	42	B60D	42	B60D	43	A44C
9	C10G	14	C09F	19	D07B	21	F04B	24	B27N	25	B41G	26	F41I	30	F21K	37	A61J	42	B60E	42	B60E	43	A44C
9	C10L	14	C09G	19	D07B	21	F04C	24	B27O	25	B41H	26	F41J	30	F21L	37	A61K	42	B60F	42	B60F	43	A44C
9	D08L	14	E04H	19	E04H	21	F04D	24	B27P	25	B41I	26	F41K	30	F21M	37	A61L	42	B60G	42	B60G	43	A44C
9	G01V	14	D08L	19	E04H	21	F04E	24	B27Q	25	B41J	26	F41L	30	F21N	37	A61M	42	B60H	42	B60H	43	A44C
10	B01J	15	A62D	19	F27D	21	F15B	24	B27R	25	B42C	26	G21J	32	F21V	37	A61N	42	F01M	42	A63D	44	A63D

Table 6.10: Schmoch concordance table

Chapter 6. A measure of International Knowledge Flows through our Input-based  
Patent Citation Indicator

Field No	Descriptions	Isic Rev 3
1	Food, beverage	15
	Tobacco products	16
2	Textiles	17
	Wearing apparel	18
	Leather articles	19
3	Wood products	20
4	Paper	21
	Publishing, printing	22
5	Petroleum products, nuclear fuel	23
6	Basic chemical	241
	Pesticides, agro-chemical products	2421
	Paints, varnishes	2422
	Pharmaceuticals	2423
	Soaps, detergent, toilet preparations	2424
	Other chemicals	2429
7	Man-made fibres	243
	Rubber and plastic products	25
8	Non-metallic mineral products	26
9	Basic metals	27
	Fabricated metal product	28
10	Energy machinery	2911 2912 2913
	Non-specific purpose machinery	2914 2915 2919
	Agricultural and forestry machinery	2921
	Machine-tools	2922
	Special purpose machinery	2923 2923 2925 2926
	Weapons and ammunition	2927
	Domestic appliances	293
	Office machinery and computers	30
11	Electrics motors, generators, transformer	311
	Electric distribution, control, wire, cable	312 131
	Accumulators, battery	314
	Lightening equipment	315
	Other electrical equipment	319
	Electronic components	321
	Signal transmission, telecommunications	322
	Television and radio receivers, audiovisual electronics	323
	Medical equipment	3311
	Measuring instruments	3312
	Industrial process control equipment	3313
Optical instruments	332	
Watches, clocks	333	
12	Motor vehicles	34
	Other transport equipment	35 353
13	Furniture, consumer goods	36

Table 6.11: Definition of the 44 sectoral field based on ISIC codes

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

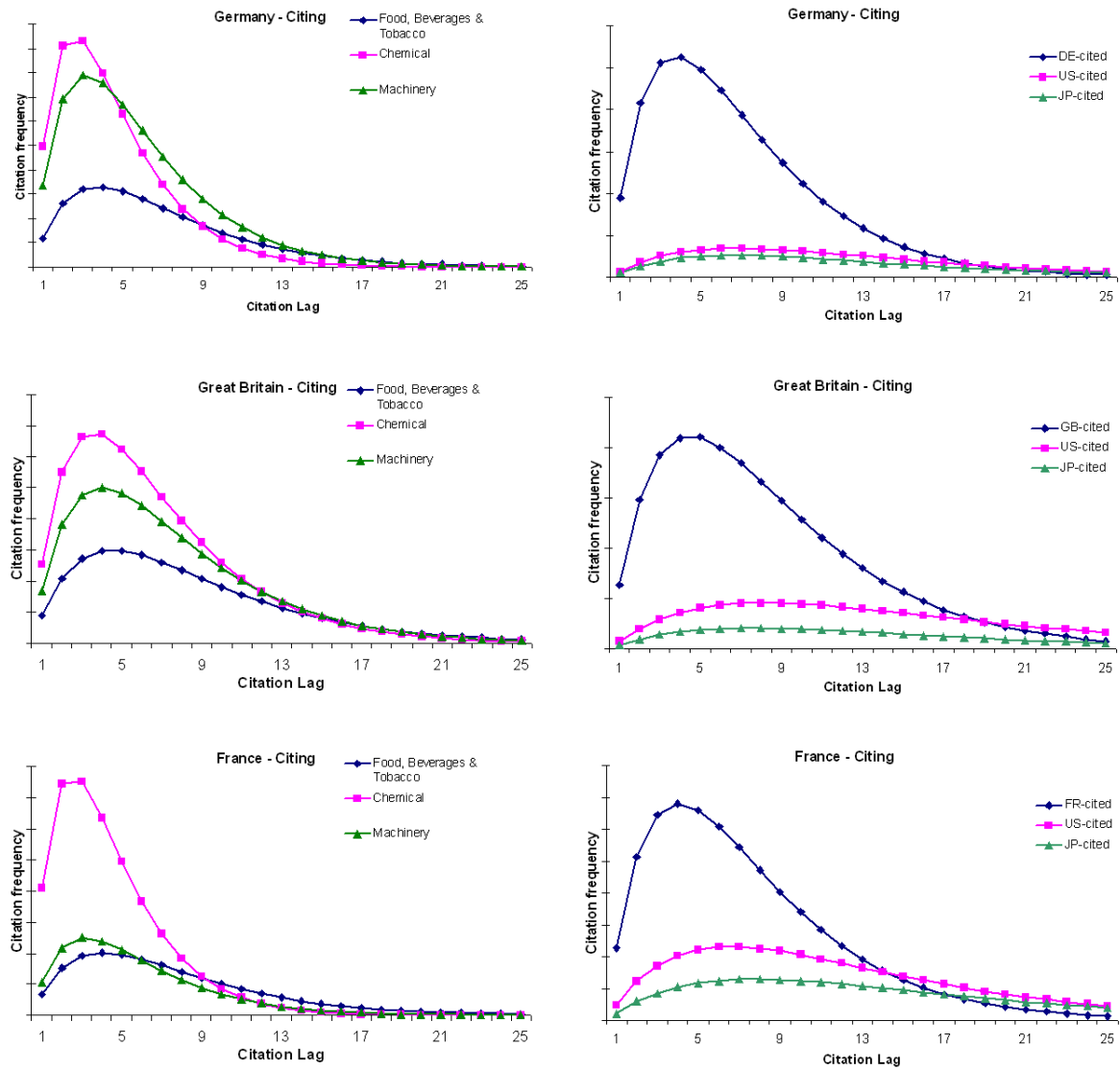


Figure 6.5: Germany, United Kingdom and France estimated citation functions



## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

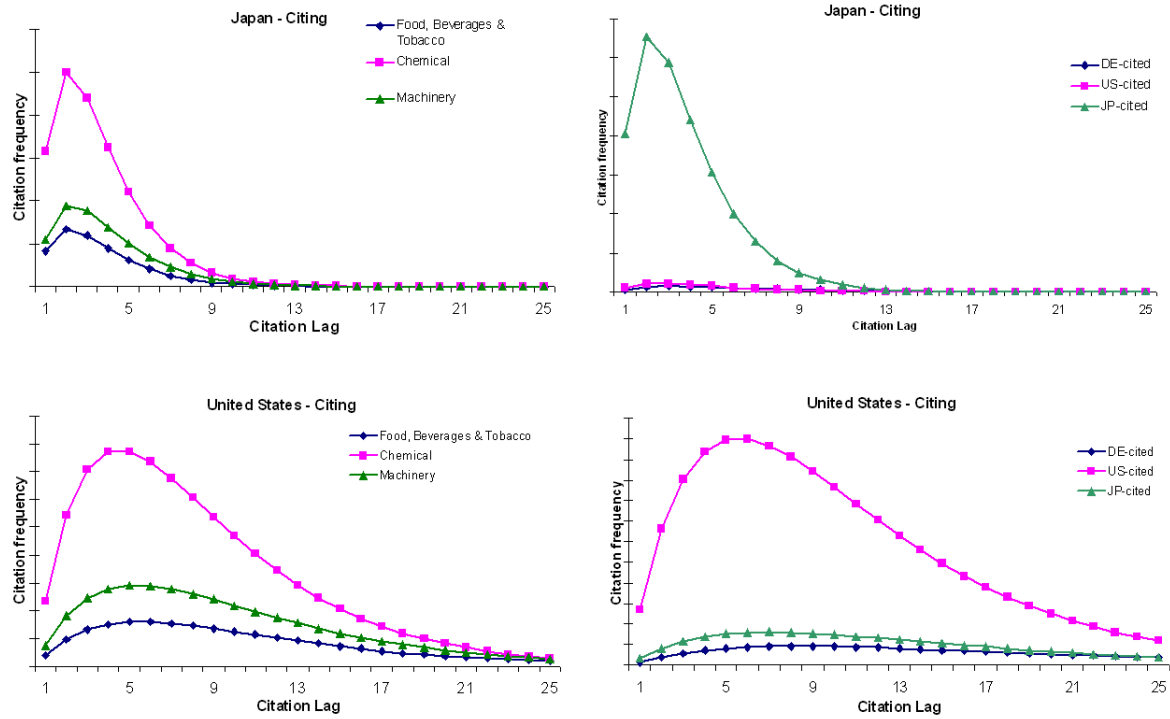


Figure 6.6: Japan and US estimated citation functions

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

### 6.8.4 Productivity estimates

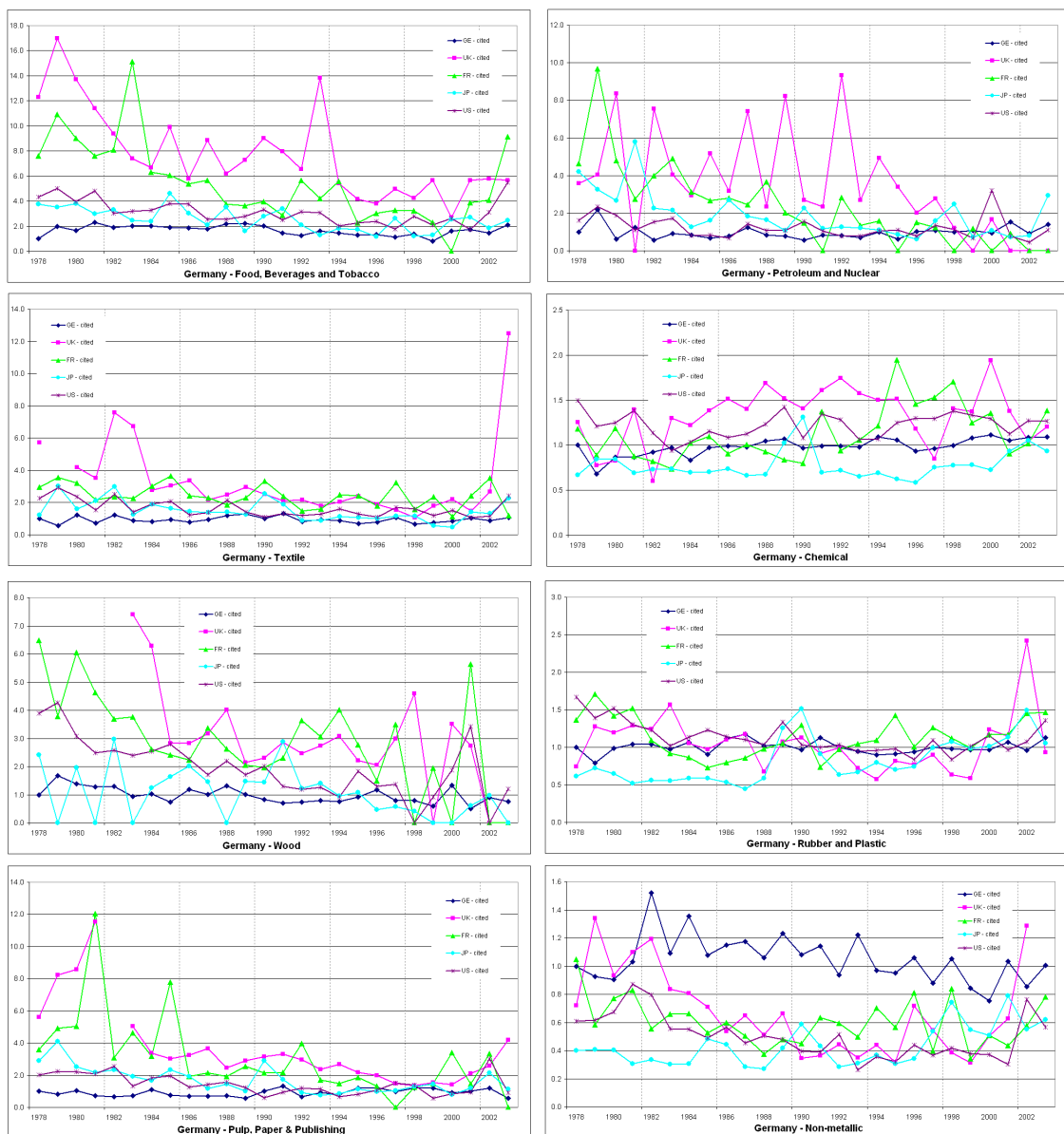


Figure 6.7: German productivity parameters - sector 1-8

# Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

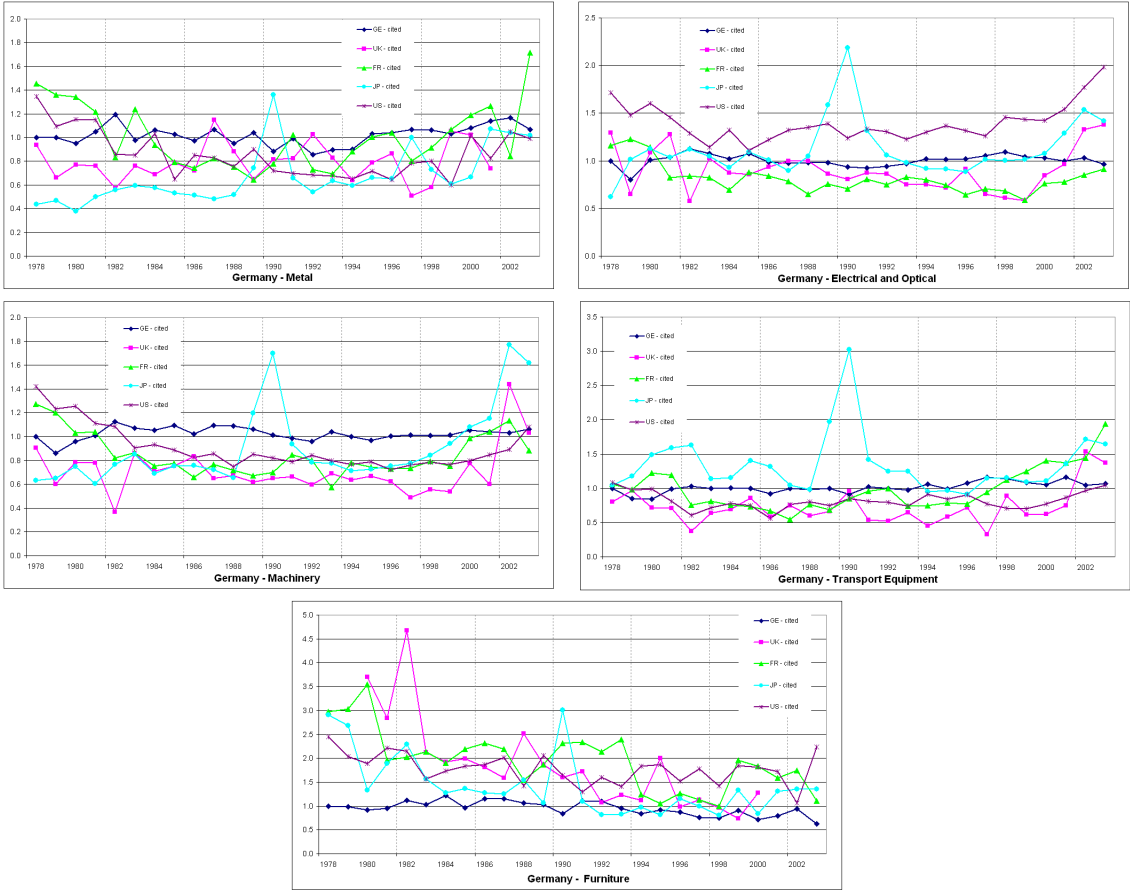


Figure 6.8: German productivity parameters - sector 9-13

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

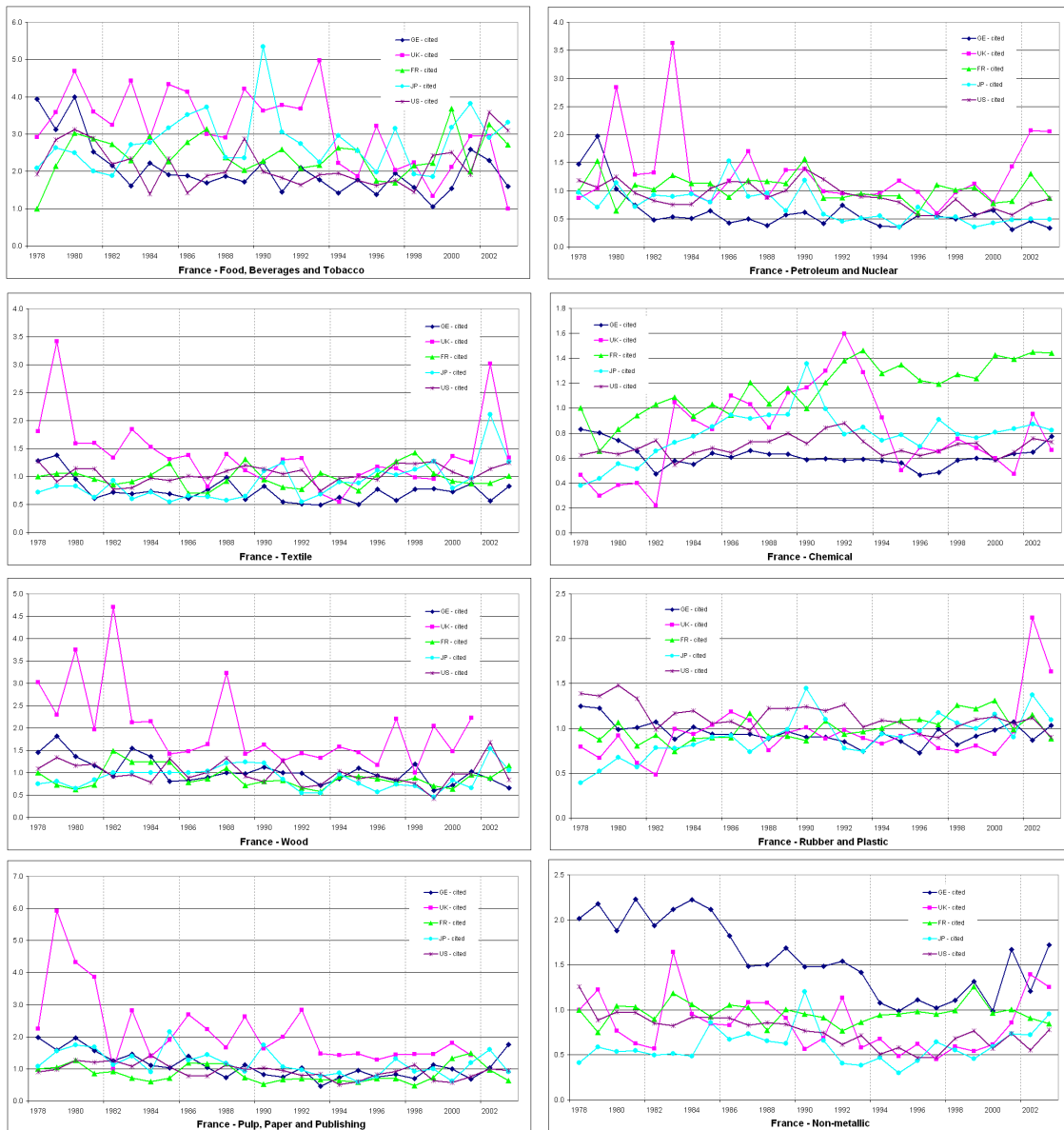


Figure 6.9: French productivity parameters - sector 1-8

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

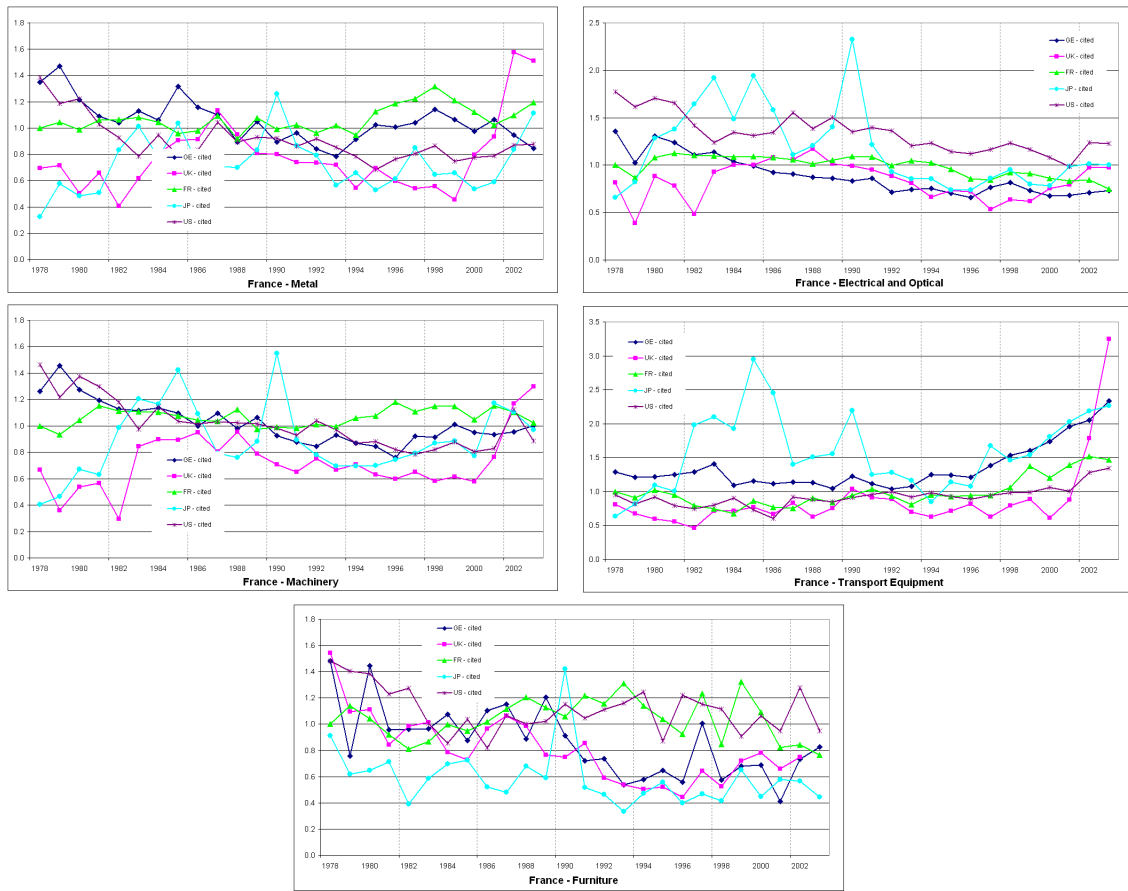


Figure 6.10: French productivity parameters - sector 9-13

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

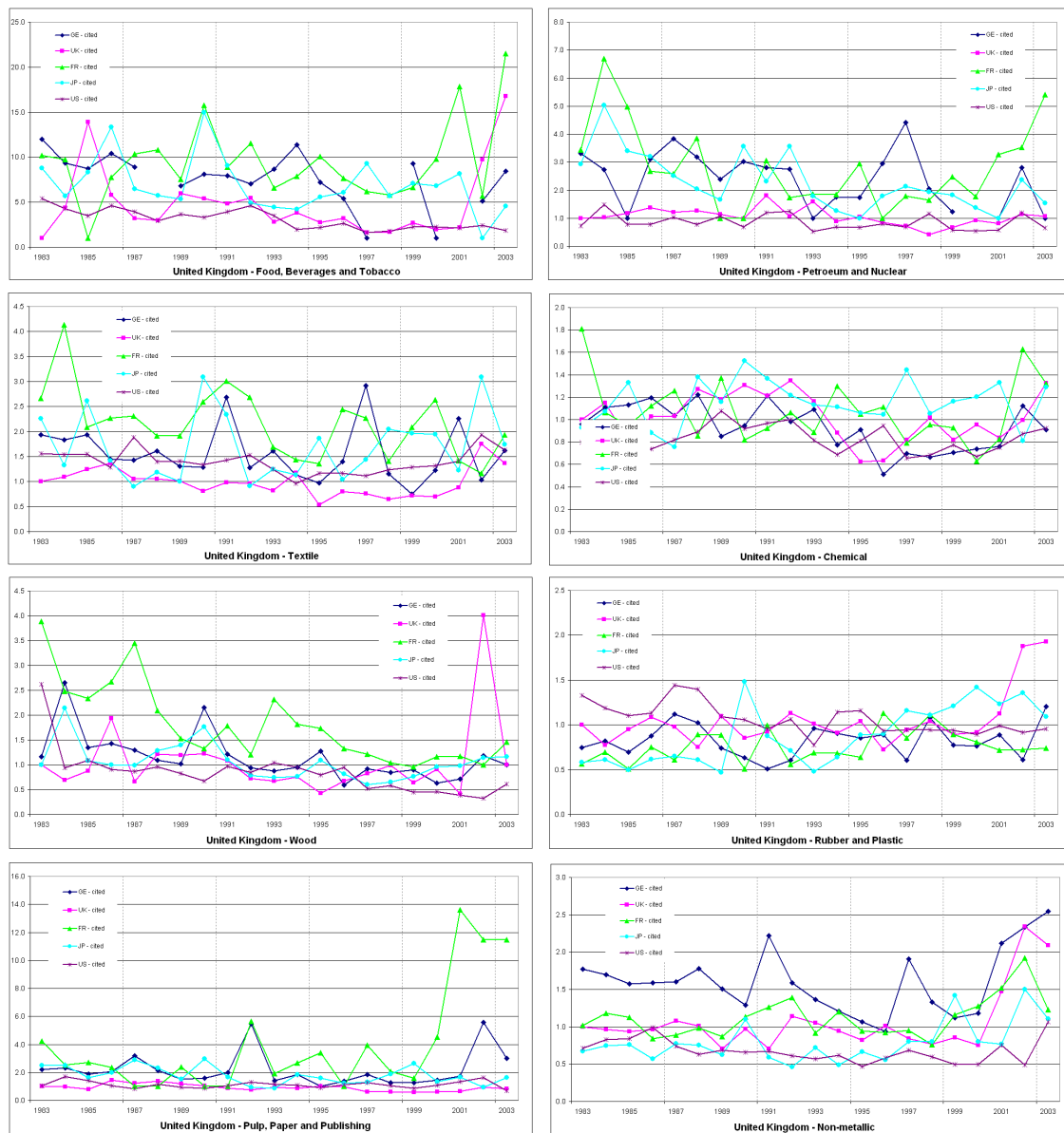


Figure 6.11: UK productivity parameters - sector 1-8

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator



Figure 6.12: UK productivity parameters - sector 9-13

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

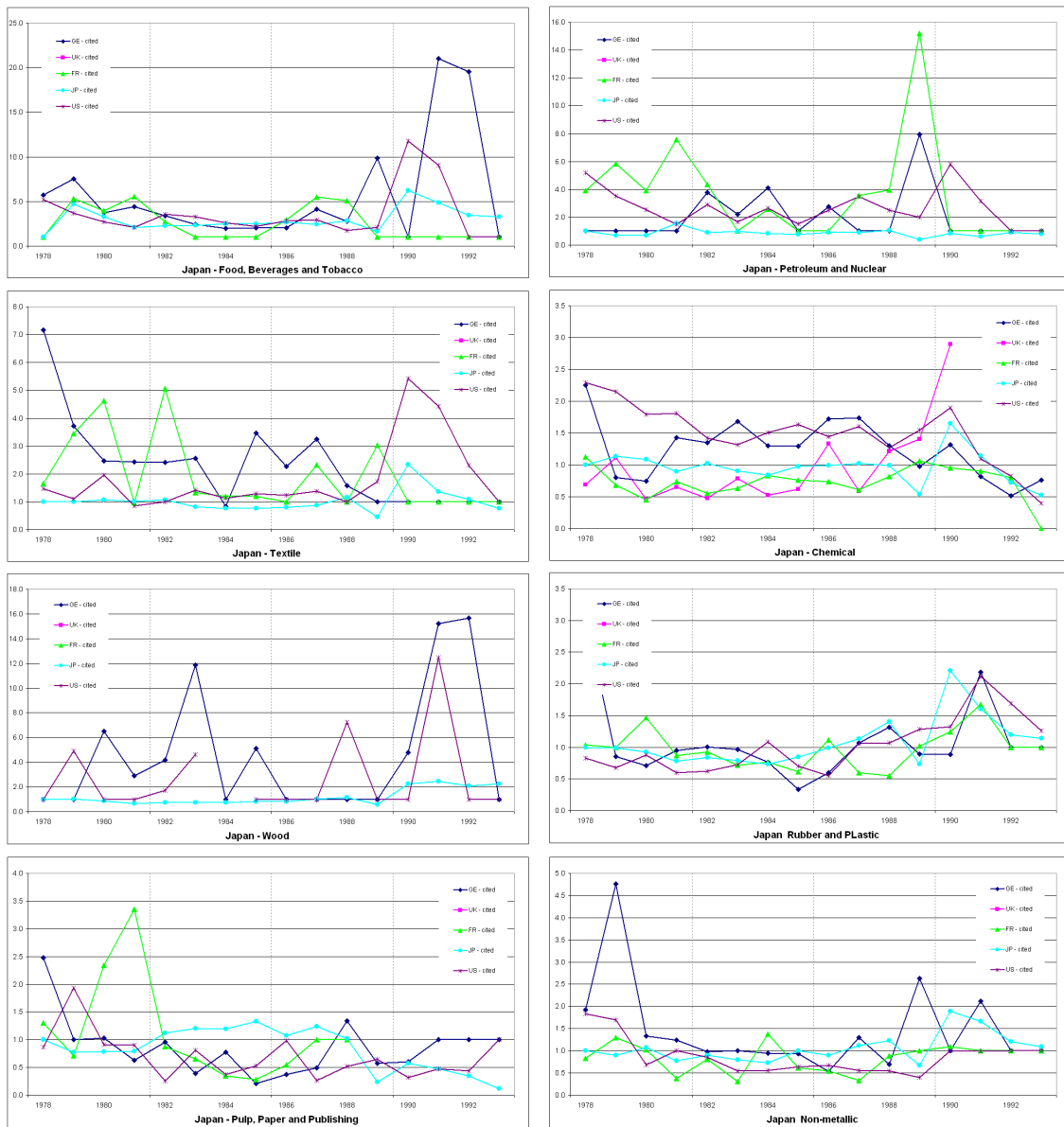


Figure 6.13: Japan productivity parameters - sector 1-8



## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

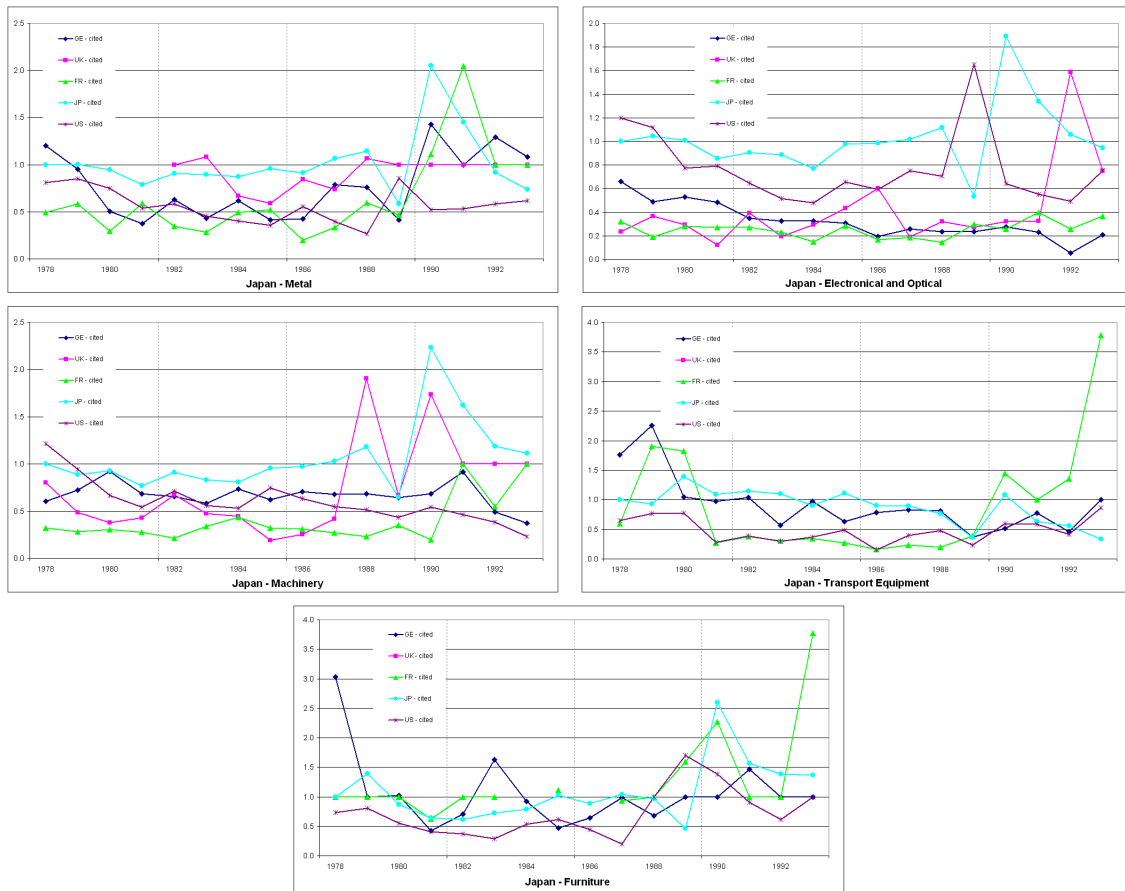


Figure 6.14: Japan productivity parameters - sector 9-13

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

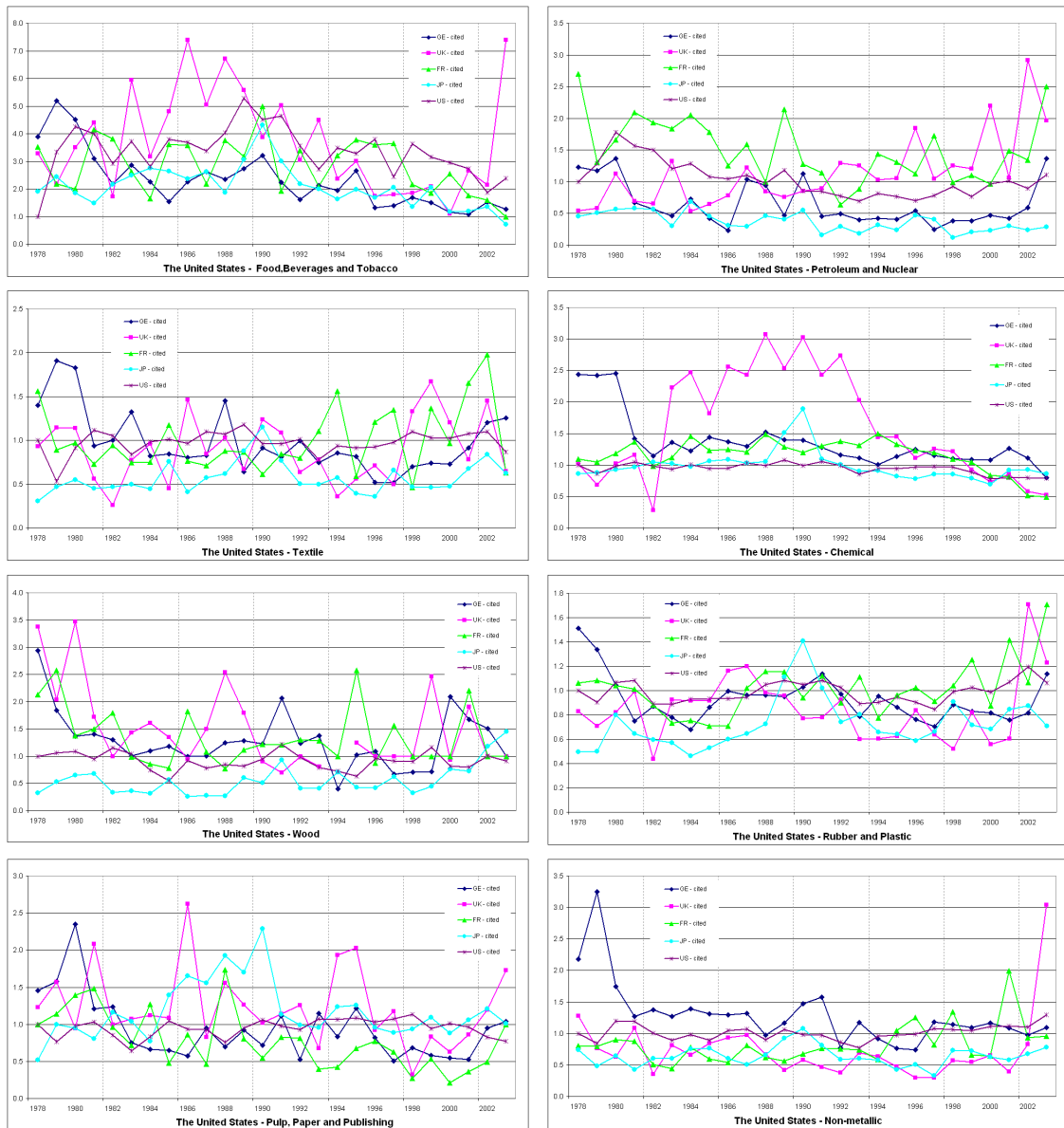


Figure 6.15: US productivity parameters - sector 1-8

## Chapter 6. A measure of International Knowledge Flows through our Input-based Patent Citation Indicator

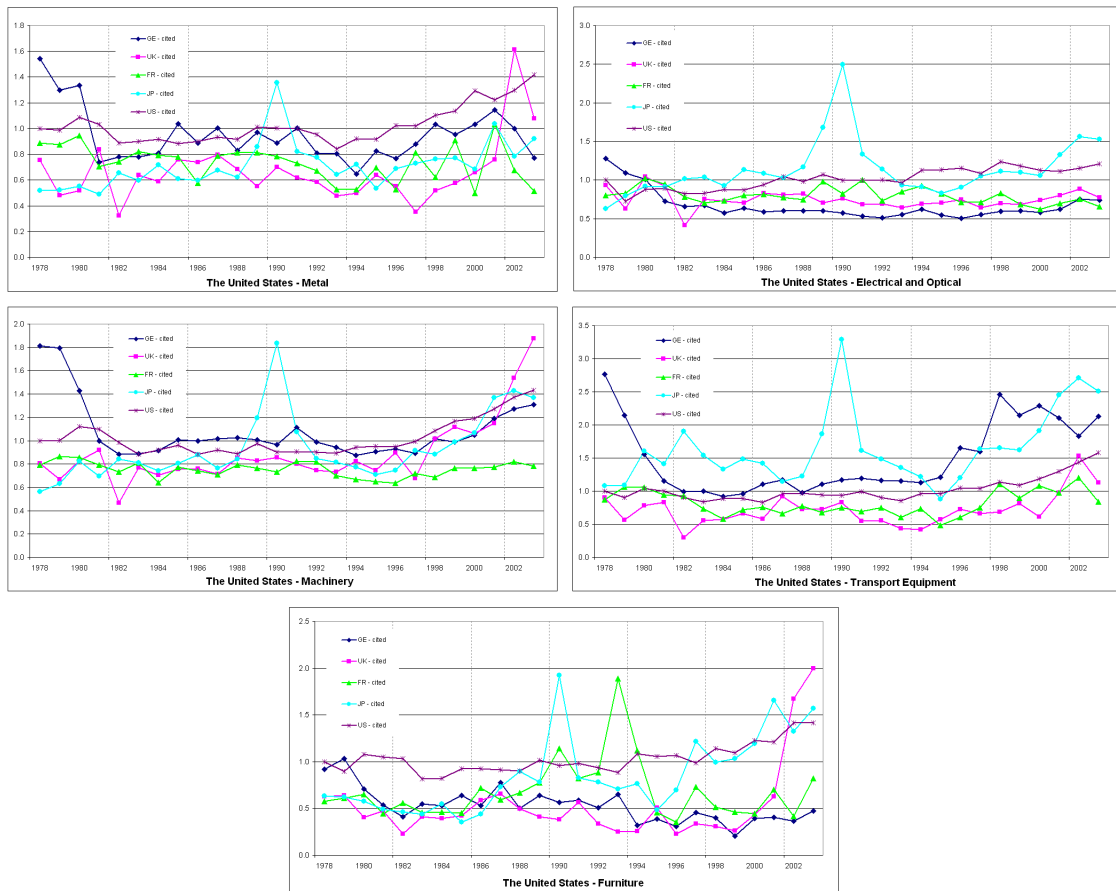


Figure 6.16: US productivity parameters - sector 9-13

## Chapter 7

# Demand-Pull and Technological-Push mechanisms in Energy-Efficient Technologies

### 7.1 Introduction

Today in a context of global warming, search for energy security and increasing world energy consumption, technological change stands to play a crucial role from environmental and economical point of view. Reducing carbon emissions without dramatic reductions in output and consumption requires the use of new technologies. These may be as simple as improvements in energy efficiency, or in advanced technologies for generating electricity from solar power or capturing and storing carbon emissions from coal combustion. Thereby, the potential of technical change is related to concerns for energy supply and to the complexity of energy systems. Understanding the process of technological change is central to evaluate public policies intended to increase the overall potential solutions to environmental and energy challenges without compromising economic growth.

Literature identifies two main determinants in the process of technical change: demand-side and supply-side factors. Demand-side factors are derived from demand-pull theories and explain new inventions by the market demand for them. The most relevant demand pull theory for energy saving technologies is the induced innovation theory. Supply-side factors are derived from technology push theories and explain inventions by the technological opportunities that inventors could benefit from, i.e. focusing on the role of existing knowledge (expressed in term of stocks of knowledge).

Innovation in energy field takes principally two forms: new product and new process. The first

group includes innovations designed to increase the supply of available energy by developing new sources of energy, particularly renewable sources (biomass, fuel cell, geothermal, hydrogen, tide & wave, solar and wind). This group constitutes the supply side and is referred as product technologies. The second group is linked to the demand side. The purpose of these technologies is to enlarge the energy efficiency. They have mainly an industrial application and are referred as process technologies (heat exchange, heat pump, Stirling engines and waste heat recovery).

We study the impact of these main determinants in 4 countries (Germany, France, Japan and the US) over a long period (from 1978 to 2003). We use as proxy for technology innovations, patent applications (Johnston et al. 2008 [93]). Demand pull determinant, that is market demand for new technology, is approximated by energies prices. Technology push determinant, which is technology opportunity, is identified by the stocks of national and international knowledge which inventor could benefit from.

Data on demand side are easily available, but data on supply side are not readily available (R&D expenditures are not available at a technological level for private sector). The main contribution of this paper is to use patent citations to estimate the technological opportunities available to inventors. The patent data are taken from the EPO Worldwide Patent Statistical Database (called Patstat). This database has been made recently available for all country and presents the advantage to provide a large volume of information relating to inventors, technological fields, citations..., covering a long time period on an international level. Our work consists in two steps. In a first step we develop the knowledge input-based indicator which represents the knowledge available to inventor. In a second step, we estimate the contribution of the two main determinants on the innovation process.

The knowledge indicator (the stock of knowledge) is built on a modified form of the extensively used “quasi-structural citations function”, Hall et al. (2001) [71]. We follow Popp (2002) [146] in using the productivity parameters estimated by the citation function to build our knowledge indicator. The knowledge stock thus created is a stock of patents weighted by their subjective productivity (depending on the inventor’s characteristics who receives the knowledge). Given the global nature of environmental and energy issues, a particularly important role can be played by the diffusion of knowledge at the international level. For this reason, we extend previous methodology according to previous chapters in order to take the role of technological externalities into account by differentiating the productivity parameters according to the geographic origin of knowledge. The indicator quantifies the knowledge used by an inventor within a specific space and technology dimension depending on the origin of this

knowledge. Stocks of knowledge are constructed for each technology and each country and distinguish between domestic and foreign knowledge.

Results show that high energy price induces innovation. But the mechanism is not automatic and leaves space for public policies aimed at promoting innovation since we also find a strong respond to technological opportunity. Thus our results suggest that the increase in energy price will enhance energy saving innovations as long as R&D investment had been previously made. Taking polled estimates of the G5 countries, we find that prices have a strong and significant effect on innovation with elasticity close to unity. Knowledge opportunity plays also a significant role with an elasticity of 0.39 for domestic knowledge and 0.45 for foreign knowledge.

Paper is organized as follows. Section 7.2 surveys the economic literature about the sources of technological change. Section 7.3 describes the models used to test the hypothesis of induced innovation. Section 7.4 presents the data and the construction of the knowledge stocks. Section 7.5 presents estimations and results. Finally, the last section concludes.

## 7.2 Literature Review

### 7.2.1 Demand pull theory

Intuitively, economic theory suggests that if the relative price of energy increases, energy intensity of the economy will fall as a result of a series of behavioral changes: agents would drive slower, they would turn down their thermostats, they would replace their goods by more efficient models, consuming less energy... In order to answer to this request of lower energy consumption, firms will propose new goods allowing a reduced consumption of energy. This mechanism is called “induced invention” and was for the first time described by Sir John Hicks in 1932:

“a change in the relative prices of the factors of production is itself a spur to innovation and to invention of a particular kind directed for economizing the use of a factor which has become relatively expensive” (Hicks, 1932, p.124) [76].

Hicks assumption received little attention until the arrival of a strong debate following the Salter’s explicit criticism of Hicks’s assumption :

“at competitive equilibrium marginal each factor is being paid its marginal value product; therefore all factors are equally expensive to firm”...” The entrepreneur is interested

in reducing cost in total, not particular cost or capital costs. When labor cost rise and advance that reduces total cost is welcome, and whether this is achieved by saving labor or saving capital is irrelevant” (Salter, 1960, p.43-44) [154]

This criticism was gradually refuted by a series of paper during the 1960s, establishing solids theoretical and empirical foundations for the induced innovation hypothesis. But the literature on the induced theory became important only after the publications of Ahmad (1966) [5] and Kamien & Schwarz (1968) [95] which gave the first bases of microeconomic formulation of Hicks’ theory.

Ahmad (1966) [5] introduces the concept of Innovation Possibility Curve (IPC). At each period, there exists a set of potential production processes determined by the basic state of knowledge available. Each process in the set is characterized by an isoquant with rather narrow possibilities for substitution, and requires a specific R&D allocation allowing this process to be applied in production. The IPC represents the efficient border of the envelope including all the isoquants which the entrepreneur might develop with a given amount of R&D expenditure. Figure A (see Appendix) presents an example of IPC. At the time  $t$ , the IPC is represented by  $IPC_t$ . It is the envelope of several isoquants, each one corresponding to different a technology. It represents the isoquant minimizing the production cost taking into consideration relative price, illustrated by line  $P_t P_t$ . In the following period, technological advance induced a displacement of  $IPC_t$  to  $IPC_{t+1}$ . If the relative price remains unchanged, new technology is represented by isoquant  $I_{t+1}$ . If the price of the factors changes, for example a rising in labor price, a different isoquant is selected:  $I'_{t+1}$ . New prices lead to technological changes using relatively less labor.

Later, Kamien & Schwarz (1968) [95] develop a more general model where the rate and direction of the technological change are endogenously determined. They develop a formulation on the conditions under which a firm, maximizing its profit with a given amount of R&D, will arbitrate between different types of technological change (i.e. neutral and no neutral technological changes). The authors show that the best alternative depends on initial technology, the relative price of factors and the relative cost of acquisition of new technologies. The main result is that the adoption of a neutral technological change can remain optimal in presence of relative changes in factors prices, if factors’ markets are in a perfect competition.

In the precise case, of studies relating to efficient energies, Atkinson & Halvorsen (1984) [12] found that new fuel saving invention for car motors responds more than proportionally to changes in expected fuel prices. This result was confirmed by Wilcox (1984) [190]. Through a creation of a quality index

of engines between 1949 and late 1960's, he shows that prices of oil and legislative constraints related to pollution induce an increase of almost 20% of engine efficiency. Ohta & Griliches (1986) [136] found that gasoline price changes over the period 1970–1981 could themselves explain much of the observed change in related automobile characteristics. Goldberg (1998) [54] examines the impact of the standard CAFE<sup>1</sup> on car sales, prices and consumption of fuel between 1984 and 1990. The author combines an oligopolistic supply model with products differentiation and demand for vehicles and observes a significant impact. Newell et al. (1999) [128] study the impact of energy prices on new models of electric household equipment available to sale between 1958 and 1993, taking into account the communication towards the consumers. The induced invention is then characterized by a change of the frontier of transformation surface of saving energy consumption goods. These authors find that the variation of the prices of energy induces the commercialization of new models and eliminates the old ones. Finally, Popp (2002) [146] tests price induce theory on energy efficiency innovations. He shows that prices play a crucial role on patenting activities with few lags, especially for energy aimed at the development of new sources. As highlighted by Popp (2001) [145], main studies on induced innovation in energy field focus only on the role of demand-side mechanism as determinant of innovation. For a detailed review of demand-Side energy efficiency policies reader could refer to Gillingham, Newell and Palmer (2006) [53].

### 7.2.2 Technology push theory

Main criticism directed toward induced innovation literature was done by Nordhaus in 1973 [129] and reproach not to consider the origin of knowledge. If we connect the induced theory to the border of innovation possibilities, this theory indicates to us how the IPC varies compared to the expenditure in R&D but not how this border was initially created. By ignoring the determinants of the IPC, the induced innovation hypothesis does not truly endogenize the path of technological change.

Technology push literature gives some solutions by focusing on the role of existing knowledge as technological opportunities for developing new innovations. Technological advances make new invention possible. Investor, even if energy prices are high, will not invest in innovating activity if probability of success is low due to a lack of background knowledge. Quoting Mowery & Rosenberg:

---

<sup>1</sup>The Corporate Average Fuel Economy (CAFE) regulations in the United States, first enacted by Congress in 1975, exist to regulate and improve the average fuel economy of cars and light trucks (trucks, vans and sport utility vehicles) sold in the US in the wake of the 1973 Arab Oil Embargo.



“Rather than viewing either the existence of a market demand or the existence of a technological opportunity as each representing a sufficient condition for innovation to occur, one should consider them each as necessary, but not sufficient, for innovation to result; both must exist simultaneously» (Mowery & Rosenberg 1979) [125].

Early literature, based on studies using questionnaires or interviews with technology managers, argued that demand side was the primary factor driving innovation (Schmookler (1962) [162], Langrish et al. (1972) [103], Myers & Marquis (1969) [127]). Mowery & Rosenberg (1979) [125] review these studies and find that questionnaires present often a bias of interpretation toward market demand. Since this study, literature focuses on the role of both demand and supply side in technological change determinant (Rosenberg (1982) [152], Dosi (1982) [40], Mowery & Rosenberg (1989) [126], Utterback (1996) [180] and Rycroft & Kash (1999) [153]). This literature also shows that these two sides are intricately intertwined and this differentiation between both of them is often difficult, not to say impossible:

“...the old debate about the relative relevance of “technology push” versus “market pull” in delivering new products and processes has become an anachronism. In many cases one cannot say with confidence that either breakthroughs in research “cause” commercial success or that the generation of successful products or processes was a predictable “effect” of having the capability to read user demands or other market signals accurately” (Rycroft and Kash 1999).

Traditional measures of knowledge were based on R&D indicators, but unfortunately at a technological level, R&D data are not available, and it is also the case for the energy technologies (IEA supply some, but only for public expenditures). Interesting solution to this data gap has been brought by Popp (2001, 2002) who applies the model of Jaffe & Trajtenberg (1996) in using patent citations as proxy for knowledge flow. Popp uses national patent citations to build its US national knowledge stock and shows that the stock of knowledge plays a more important role than the energy price, the latter preserving however a strong significant impact.

### 7.3 Modeling

Innovating activity is affected by both demand and supply factors. The demand-side determinants of innovation are affected by changes that make inventive activity more (or less) profitable at a given level

## Chapter 7. Demand-Pull and Technological-Push mechanisms in Energy-Efficient Technologies

---

of scientific knowledge (changes can affect aggregate demand or population for instance). Supply-side determinants of innovation are affected by changes in technological opportunity, including scientific and technological advancements, which make possible additional innovation more profitable or less costly at a fixed level of demand. Since a structural model is difficult to define, we use a log-log model that follows usual econometric studies on induced innovation. We regress first patent application on energy price, knowledge stocks and policies incentives. The first patent application is our indicator of innovation, the price of energy is for the market demand for innovation, policies are the public demand for innovation and knowledge stock represents the technological opportunities available to the inventors. This specification allows results to be interpreted directly in term of elasticities. Model (7.1) is defined as follows:

$$\text{Log} \left( \frac{\text{patent}_{t,c,i}}{TPatent_{t,c}} \right) = cst + \gamma_1 \text{Log} \cdot P_{t-1,c} + \gamma_2 \text{Log} \cdot Know_{t,c,c,i} + \gamma_3 \text{Log} \cdot Know_{t,c,-c,i} + \gamma_4 \text{Log} \cdot Policies_{t,c} \quad (7.1)$$

With:

- *Patent* is the number of first patent application made for technology *i*, in country *c* at year *t*.
- *TPatent* is the total number of first patent application in country *c* at year *t*.
- *P* is the energy price index in country *c* at year *t*.
- *Know* is the knowledge stock for technology *i*, in country *c* at year *t*.
- *Policies* is a composite variable of different policies for renewable energy, in country *c* at year *t*.

In order to collect the structural effects of patents application variation, we use as dependent variable the fraction of first patents application, i.e. the number of patents by technology and by country, divided by the total number of first patents application by country (for all technologies class). Changes affecting all the patents will influence, at the same time, the numerator and the denominator, making constant the dependent variable. Popp (2001) [145] finds that energy innovations respond quickly to change in energy price and considers current and lagged price in his analysis. To avoid colinearity between price values we keep only the lagged price that seems more relevant. Recent theoretical and empirical literature provides robust evidences that knowledge spillovers exist (see chapter 2 & 3). We expect it is also the case in energy efficiency technology and we take into account knowledge coming

from abroad in the construction of our supply-side determinant. Thus, knowledge stock is divided into two components: a domestic and a foreigner one. First component represents the national knowledge stock build upon patent published by national inventor and weighted by their productivities for national inventor. Second component represents international knowledge stock build upon patent published by foreign inventors and weighted by their productivities for national inventor. International knowledge includes knowledge coming from the four countries considered minus the country studied *c*. Finally, Johnstone et al. (2007) [93] show that public policies are an important driver for renewable energies (that we have called product technologies). We control public incentives for innovation by the variable Policies. This variable is a composite variable of different public policies aimed at the development of renewable energy.

## 7.4 Data

### 7.4.1 Energy patent data set

**Patents** We use as a proxy for innovating activity the number of patents application by technology<sup>2</sup>. We retain only the first application, thus the application date refers to the priority date that is the first date of filing of a patent application, anywhere in the world, to protect an invention. It is the earliest and therefore closest to the invention date (OECD 2008 [132]). Data on patent are taken from the EPO Patstat database. A data set of 11 energy technology groups was constructed for the 4 countries (Germany, France, Japan and United States). To identify the country of origin of the innovation, we class patent by inventor country.

**Energy technologies selected** Energy technology groups are determined by International Patent Classification (IPC)<sup>3</sup>. The classification used is based on the work of Popp (2001) [145], Johnstone et al. (2007) and on the definitions of environmental technologies given by the Minister of the Economy, Industry and Employment (France). Since Popp classification is being expressed in term of American classification, we use the “Energy Information Administration” concordance table (UPTSO Table) to transform it into IPC8 classification<sup>4</sup>. Some qualitative modifications were made in order to correct

---

<sup>2</sup>Since the seminal work of Griliches (1990), we know that patents are a good indicator for innovating activities. Patent provides a uniquely detailed source of information on innovation process (including spatial, temporal, technology dimension) but it also present some weaknesses. All these issues are detailed in chapter 4.

<sup>3</sup>As from now, IPC would always mean International Patent Classification.

<sup>4</sup>Now and thereafter, IPC holds for International Patent Classification

## Chapter 7. Demand-Pull and Technological-Push mechanisms in Energy-Efficient Technologies

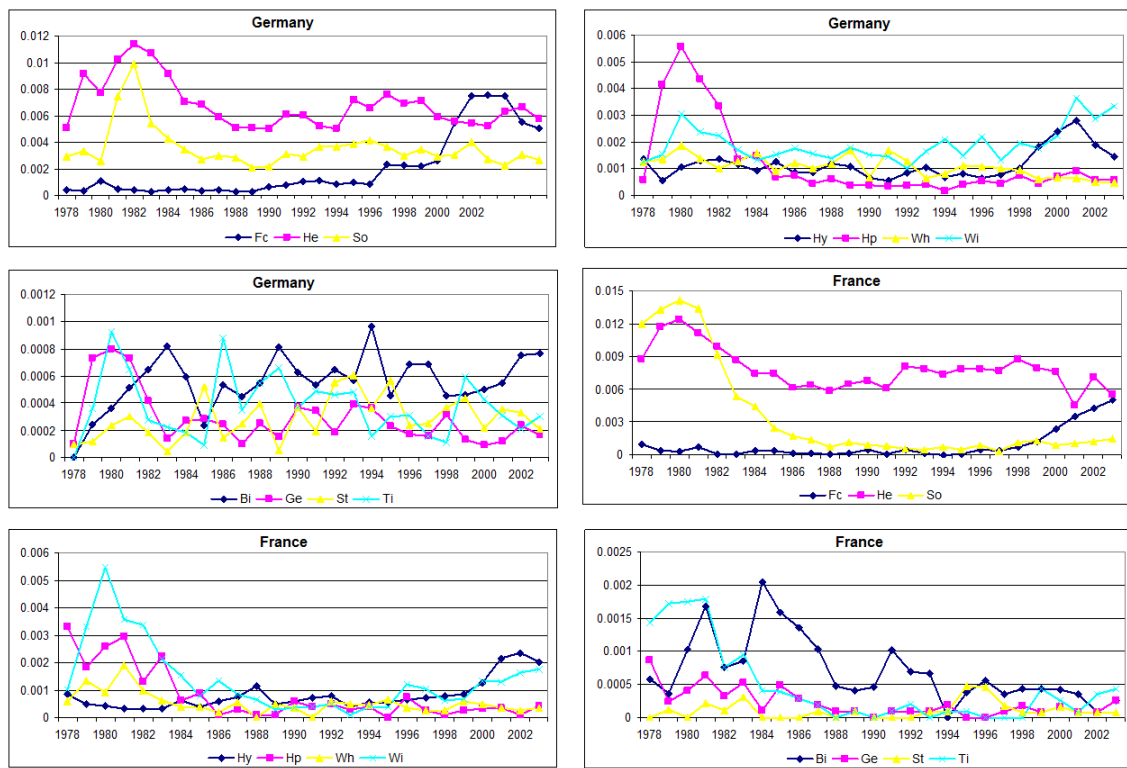


Figure 7.1: First patent application in energy efficiency technologies

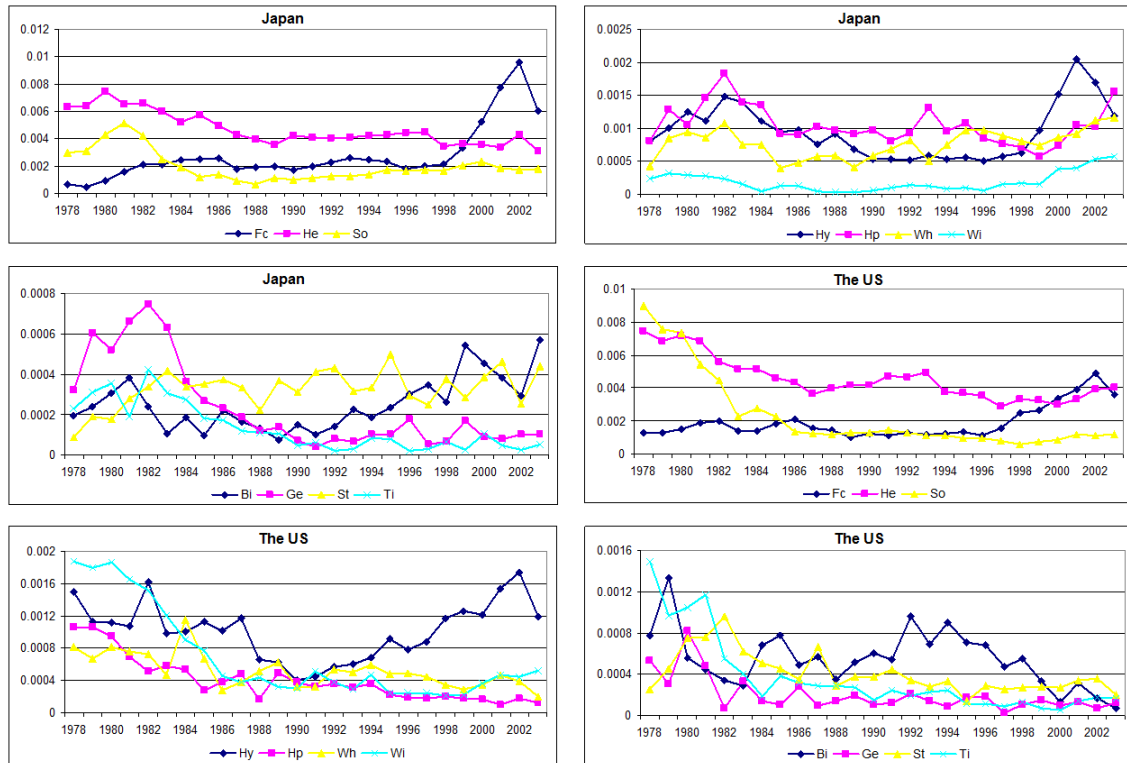


Figure 7.2: First patent application in energy efficiency technologies (continued)

some inappropriate approximation made by UPTSO table. Finally, the selected technology could be separated into two groups. The first group embodies innovations designed to develop new energy sources and is referred as product technologies.

It includes Biomass (Bi), Fuel Cell (Fc), Geothermal energy (Ge), Solar energy (So), Tide & Wave (Ti) and Wind (Wi). The second group embodies energy efficiency innovations that focus mainly on industrial energy consumption and is referred as process technologies. It includes Heat exchange (He), Heat Pump (Hp), Stirling engines (St) and Waste Heat recovery (Wh). IPC patent classification used to select energy innovations is given in Appendix. Figure 7.1 & 7.2 plots the relative annual count of successful patent applications for all technologies from 1978 to 2003. Figures illustrate trend in the data. For most of the technology groups, there is a jump in patent application after the energy crisis in the late of 1970's and early 1980's, suggesting that energy prices do play an important role in inducing energy efficient technological change. For some technology groups such as fuel cell or hydrogen, the increase seems to happen only in recent years. The final sample selected is composed of 60 069 patents applied by innovators in 4 countries between 1978 and 2003. Table 7.1 presents the descriptive statistics

Technology	Country				Total
	Germany	France	Japan	United States	
<b>Bi</b>	393	183	492	668	1 736
<b>Fc</b>	1 893	257	5 532	2 647	10 329
<b>Ge</b>	183	55	450	216	904
<b>He</b>	4 550	2 051	9 688	5 313	21 602
<b>Hp</b>	743	195	2 152	421	3 511
<b>Hy</b>	870	224	1 871	1 303	4 268
<b>So</b>	2 456	845	3 889	2 345	9 535
<b>St</b>	206	30	695	455	1 386
<b>Ti</b>	254	111	260	356	981
<b>Wh</b>	678	138	1 566	614	2 996
<b>Wi</b>	1 409	359	343	710	2 821
<b>Total</b>	13 635	4 448	26 938	15 048	60 069

Table 7.1: Patents in Energy Efficient Technologies by Innovating country, 1978-2003

and information on the distribution of the patenting activity over the technology and country analyzed. Countries considered accounts for the same amount of inventions except France which represents one third of the innovation activity of other countries. Japan presents the highest innovation pattern, but if we control it per patent breadth, values are in range with Germany and the US. Concerning the pattern of innovation, Figure 7.3 shows the distribution of innovations by technology within these countries. More than one third of energy efficient innovations are directed toward Heat Exchange, but a significant number of patents are also applied for other technologies, with Solar and Fuel Cell being the second and the third technologies. Technology distribution is relatively stable among the different countries, maybe at the exception of wind that seems to be an European purpose.

#### 7.4.2 Energy prices

Energy price serves as the proxy for changes in the demand for innovation in energy efficiency technologies. It represents the demand based on our assumption of induced innovation. Ideally, it would be preferable to take different energy prices into account according to the different technologies selected. For instance process technologies are mainly used in industrial application, so the right proxy would be a shadow industrial energy price. However, since the data is not available for all the technologies and countries selected and because we need homogeneous and long set, we use the OECD energy consumer price index<sup>5</sup>. Energy Prices are in constant 2000 dollars, deflated by the OECD consumer price index.

---

<sup>5</sup>Consumer Price Indices (CPIs) measure the average changes in the prices of consumer goods and services purchased by households. Energy Consumer Price index refers to items "electricity, gas and other fuels" as defined under the Classification of Individual consumption According to Purpose (COICOP 04.5) and "fuel and lubricants for personal

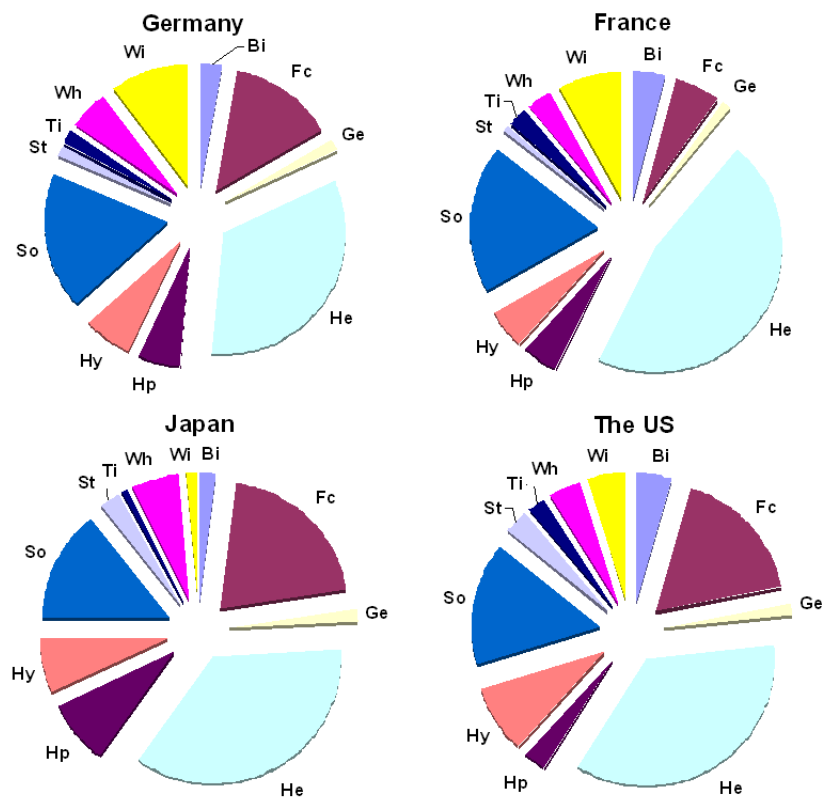


Figure 7.3: Distribution of technologies between countries

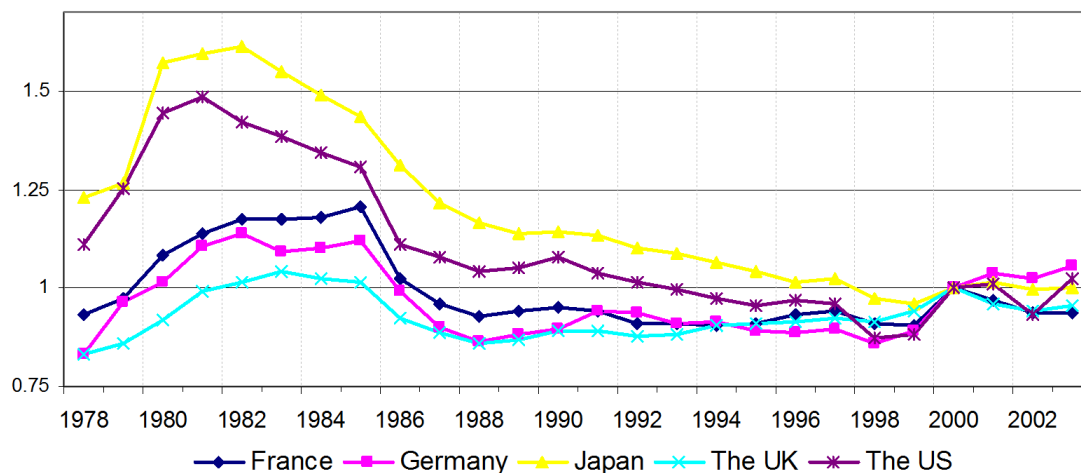


Figure 7.4: Energy prices

Energy price indexes are normalized to 100 in year 2000. They are reported in Figure 7.4.

### 7.4.3 Policies

Policies variable is a principal component variable of different kinds of public policies aimed at developing renewable energy sources. This variable is the same than the one used by Johnstone et al. (2007). It is initially compiled by the International Energy Agency (IEA) 2004 [79]. This variable takes into account public drivers for innovation. The different policies included are: support for R&D, investment incentives (third-party financing, investment guarantees), taxes (exemptions, rebates), price-based policies (tariffs, guaranteed prices), obligations and tradable permits. We first construct a variable per policy which takes the value of 0 prior to introduction of the policy and 1 thereafter. Afterward, we build a composite policy variable using principal component analysis and we keep only the first principal component. As stressed by Johnstone and al. (2007), this variable does not help to draw conclusions on the efficiency of the different types of policies, but is a good indicator of the intensity of environmental regulation. Table 7.2 presents the implementation of policies by type in the G5 countries.

---

transport equipment" (COICOP 07.2.2).



	France	Germany	Japan	The US
Research & Development	1985	1974	1974	1974
Investment Incentives	1980	1985	1994	1980
Taxes	1980	1999		1978
Tariffs	1996	1991		1978
Voluntary Programmes		1996	2000	1993
Obligations	2000	2000	2001	
Tradable Permits			2001	

Source: adapted from IEA (2004)

Table 7.2: Date of introduction of different public policies for renewable technologies

#### 7.4.4 Knowledge Stocks

Since data on R&D expenditures are not available for the technological level selected, we use data on patent citation to build our knowledge stocks as it has been done in the previous chapter. But due to constraint data we are not able to construct one stock by cited country, thus we only implement one stock for domestic and one for foreign knowledge.

**Citations and knowledge flows**<sup>6</sup> When a patent is granted, it contains several citations to earlier patents that are related to the current invention. The granting of the patent is a legal statement that the knowledge embodied in the patent represents a new and useful contribution over and above the previous state of knowledge, as represented by the citations. Thus, in principle, a citation of patent A by patent B means that A represents a piece of previously existing knowledge upon which B is built. Using the interpretation of patent citation as measuring flows of knowledge, patents that receive many citations from subsequent patents must have provided greater technological opportunity. The importance of technological opportunity does not depend exclusively on the cited patent but also on the characteristic of the citing patent.

**The quasi structural citation function** Studies on patent citations were encouraged by the new finding that citations appear to be correlated with the value of innovations (Trajtenberg 1990) [178]. Economists undertook work aimed primarily at demonstrating the potential usefulness of citations for a variety of purposes: as an indicator of spillovers (Jaffe, Trajtenberg and Henderson 1993 [89], Caballero & Jaffe 1993 [25]), and as an ingredient in the construction of measures for other features of innovations,

---

<sup>6</sup>Recent works have confirmed the role of patent citation as a proxy in the measurement of knowledge flow. Detailed discussions on the link between patent citation and knowledge flow are proposed by Jaffe & al. (2000), Breschi and Lissone (2004), Duguet & Mac Garvie (2005) and Gay & Le Bas (2005). In chapter 4 and chapter 5 we have extensively detailed this issue.

## Chapter 7. Demand-Pull and Technological-Push mechanisms in Energy-Efficient Technologies

---

such as “originality” and “generality” (Trajtenberg, Jaffe and Henderson 1993). Jaffe & Trajtenberg (1996) [83] develop Caballero & Jaffe (1993) [25] model of citation to take into account particularities of different cohorts<sup>7</sup>. They develop what it nowadays called the “quasi-structural” approach. This approach identifies separately the contribution of citations lag distribution (obsolescence and diffusion of knowledge), fertility (for our purpose : productivity of knowledge embodied) and proportion to cite in the citation process. The number of citation depends on the size of each cohort<sup>8</sup>, so it is necessary to control the size effect by looking at the probability of citation. The probability of citation is given by:

$$P_{cit_{ctd,ctg,i,c,d}} = \frac{Cit_{ctd,ctg,i,c,d}}{N_{ctd,d,i}N_{ctg,c,i}} \quad (7.2)$$

Where:

- $Cit_{ctd,ctg,i,c,d}$  is the total number of citations received by a patent cohort published in year  $ctd$  (cited year), in country  $d$  and in technology field  $i$  by a subsequent patent cohort applied in year  $ctg$  (citing year), in country  $c$  and in the same technology field  $i$ .
- $N_{ctd,d,i}$  is the number of patent published in year  $ctd$ , in country  $d$  and in technology field  $i$ .
- $N_{ctg,c,i}$  is the number of patent application in year  $ctg$ , in country  $c$  and in technology field  $i$ .

Country  $d$  is either the domestic country,  $d = c$ , or all the foreign countries,  $d = -c$ . We construct these data for each of the eleven technology fields  $i$ . All citations concern only patents belonging to the same class of innovation  $i$ , we do not allow inter-technological spillovers. For our purpose, we adapt the formulation of Caballero & Jaffe (1993), Jaffe & Trajtenberg (1996), Jaffe & Trajtenberg (1998) and Popp (2002) in order to take into account of international knowledge externalities. We assume that the citation frequency is determined as a multiplicative function of :

- Cited year : the usefulness of the knowledge represented in the patent being cited :  $\alpha_{ctd,d,i}$
- Citing year : the frequency by which patents applied in citing year cite earlier patents :  $\alpha_{ctg}$
- Obsolescence : the rate at which the knowledge represented by cited patents becomes obsolete :

$$\beta_1(\beta_{1,d}, \beta_{1,i})$$

---

<sup>7</sup>For different applications of this method see : Jaffe & Trajtenberg (1996), Hall et al. (2002), Bacchiocchi & Montobbio (2004), Branstetter & Ogura (2005), Adams, Clemmons & Stephan (2006) and Marco (2007).

<sup>8</sup>Since it not possible to study patent productivity for each patent published, we work with patent cohort i.e. all patents applied (citing patent) or published (cited patent) during the related year.

- Diffusion : the rate at which knowledge diffuses :  $\beta_2$
- A set of control variable that can be interpreted as the fixed effect in traditional linear model :  $\alpha_d$  and  $\alpha_i$

Note that the first parameter is the value of interest for this paper. It tells us the likelihood that patents from year  $ctd$  will be cited by subsequent patents. Higher values indicate that patent cohort in question is more likely to be cited by subsequent patents. This implies that the knowledge embodied in those patents is particularly useful. The other parameters control other facets of the patenting process that might affect the likelihood of citation.

**Estimations of productivity** Productivity parameters are estimated through the quasi structural approach mentioned above. The probability of citation is expressed as follows:

$$Pcit_{ctd,ctg,i,d} = \alpha_{ctd,d,i} \cdot \alpha_{ctg} \cdot \alpha_d \cdot \alpha_i \cdot e^{(-\beta_1\beta_{1,d}\beta_{1,i}(ctg-ctd))} \cdot \left(1 - \exp(-\beta_2(ctg-ctd))\right) + \epsilon_{ctd,ctg,i,d} \quad (7.3)$$

It's a combination of two exponential processes and shifts parameters<sup>9</sup>.  $\alpha_{ctd,c,i}$  is the productivity of domestic patents and  $\alpha_{ctd,-c,i}$  is the productivity of foreign patents. Because this function is non-linear, it is possible to identify distinct  $\alpha_{ctd,c,i}$  or  $\alpha_{ctd,-c,i}$  from  $\alpha_{ctg,i}$  effects, at least in principle. In practice, we found that the estimation is difficult with a full set of unconstrained cited year and citing year effects. For this reason we group citing year in four-year interval and we allow cited effect to vary every year for domestic  $c$  and foreign patents  $-c$ . Multiplicative parameters of obsolescence, i.e.  $\beta_{1,d}$  and  $\beta_{1,i}$ , are function of geographic localization and technological field, because we expect different patterns of knowledge diffusion according to these characteristics. For instance, we expect that domestic knowledge diffuse faster to domestic inventors than foreign knowledge. Finally, multiplying parameters  $\alpha_d$  and  $\alpha_i$  controls for differences in the propensity of the different localizations or technology groups to be cited. Equation (7.3) is estimating with non linear regressor, using all patents published from 1978 to 2003 for the cited year and all patent applied from 1978 to 2006 for the citing year. Estimations are made separately for each country  $c$ . Finally, since this is grouped data, observations are weighted by  $\sqrt{N_{ctd,d,i}N_{ctg,c,i}}$  to avoid problems with heteroskedasticity (Greene 1993).

---

<sup>9</sup>Recall that captures the depreciation or obsolescence of knowledge and captures its diffusion. The maximum value of the citation frequency is approximately determined by and the modal lag is approximately equal to .

## Chapter 7. Demand-Pull and Technological-Push mechanisms in Energy-Efficient Technologies

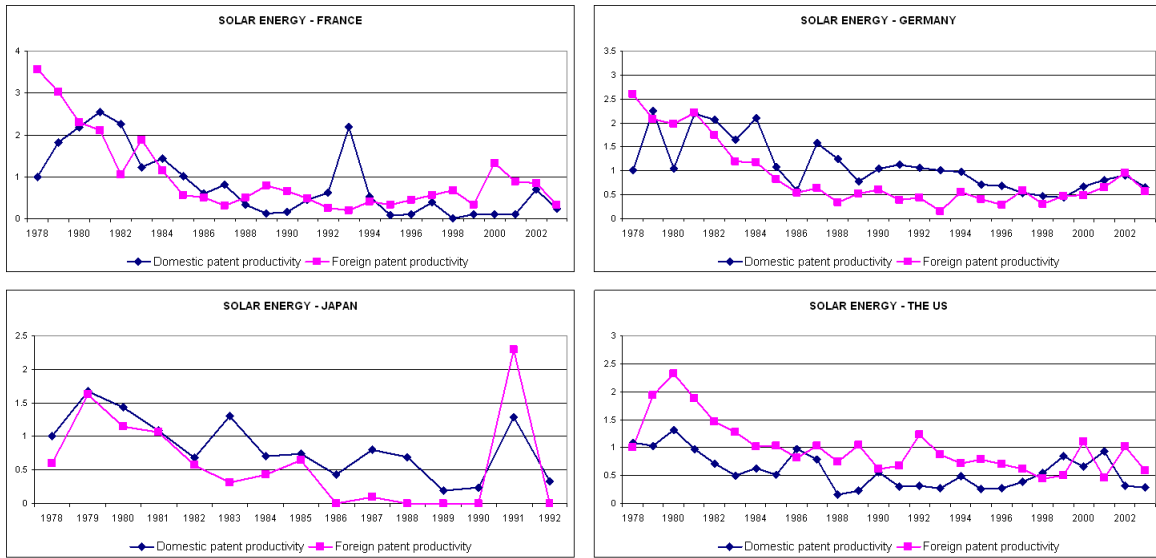


Figure 7.5: Productivity estimates for Solar energy technology

**Results** To overcome the identification problem we normalize  $\alpha_{ctd=1978,d=c,i} = 1$  for each  $i$ ,  $\alpha_{ctg=1978} = 1$ ,  $\alpha_{d=c} = 1$ ,  $\alpha_{i=he}$ ,  $\beta_{1,d=c} = 1$  and  $\beta_{1,i=he}$ . The preceding parameters can be interpreted as the proportional difference in citation intensity for a given year, technological field or country relative to the base group constraint to 1. Thus, estimates higher than one, for  $\alpha_{ctd,d,i}$ , mean that patents published in year  $ctd$  in country  $d$  are more useful to future inventors than domestic patents published in 1978. Similarly, estimates higher than one, for  $\alpha_{ctg}$ , mean that patents applied in this year  $ctg$  tend to make more citations than patents applied between 1978 and 1979. These estimations help to identify the knowledge that national inventors benefit from. Foreign knowledge may have different implications not only because of the distribution of countries included in the foreign flows but also because of the characteristics of the countries receiving these flows. Here, foreign knowledge operates in a same logic than in traditional spillovers studies (see chapter 2 and 3). In empirical literature, impact of foreign knowledge depends on the proximity between nations (in term of geographic distance, technology proximity, bilateral exchange...). Foreign knowledge is thus weighted by these proximity parameters. In the present study, we estimate directly the weighted parameters (that is our productivity parameters and localization multiplicative parameters) and the lag of diffusion of foreign knowledge.

Because of the large number of results ( $11 \times 4$  set of results) we present here only one result by category of energy innovation<sup>10</sup>. Figure 7.5 displays the results for the solar productivity parameters

<sup>10</sup>Other results are available upon request.

## Chapter 7. Demand-Pull and Technological-Push mechanisms in Energy-Efficient Technologies

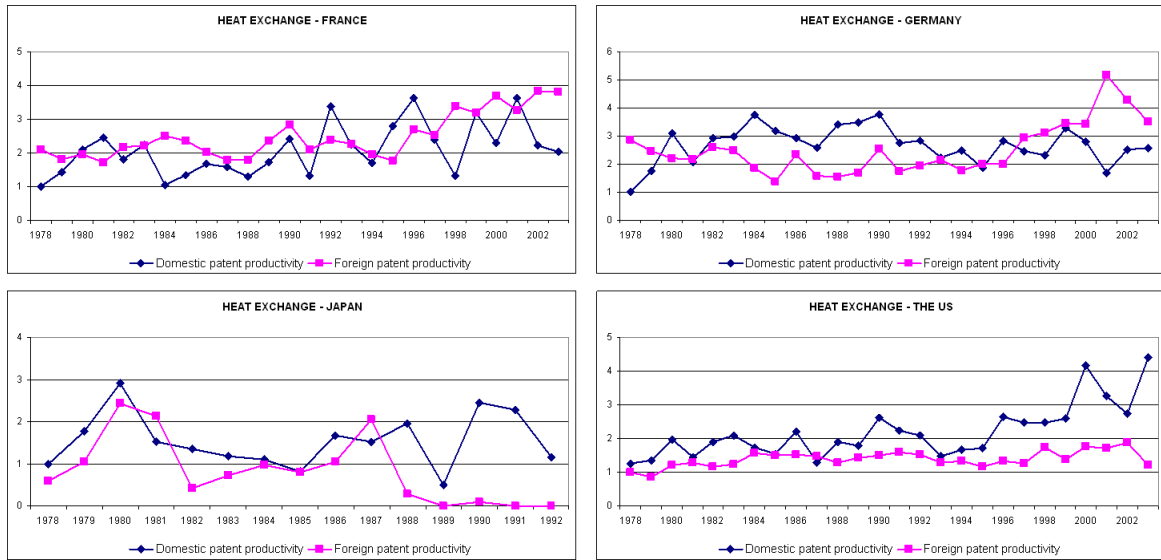


Figure 7.6: Productivity estimates for Heat Exchange technology

and figure 7.6 for the heat exchange productivity parameters. The other estimates are provided in Appendix (Table 7.6). Figure 7.5 & 7.6 show patent cohort productivities by published year and by country of interest. For example, considering the US solar productivity, we assume that patents published in solar energy field around the world embody more useful knowledge until 1986 than patents published in the US in 1978 for US inventors. Some parameters are no different from zero mainly for Japan and France meaning that patents published in this year present not useful knowledge or corresponding to year where there are no patents published in the related technology group. Concerning results provided in Appendix, it's interesting to note that the diffusion parameters are always higher for domestic knowledge. In order to highlight the implication of diffusion and obsolescence parameters, examples of knowledge flows are given in Figure 7.7, presenting knowledge flows created in 1982 that benefit an US and a German inventors.

We see that patterns of diffusion differ according to origin and destination of knowledge. For German inventors, foreign knowledge seems to be as important as domestic knowledge but its diffusion takes more time. Domestic knowledge median lag is 2.64 years whereas it is 7.33 years for foreign knowledge. At the opposite US inventors benefit with a closer lag from domestic and foreign knowledge (with a respective lag of 3.39 and 5.5 years) but domestic knowledge flows seem to be more relevant than foreign ones.

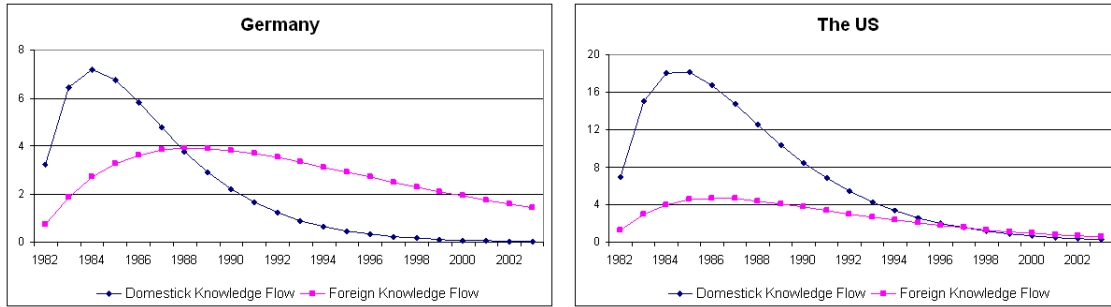


Figure 7.7: 1982 Solar knowledge flow for German and US inventors

**Construction of knowledge stock** We use productivity parameters estimated above to build our knowledge stocks. We assume that knowledge stock available for an inventor located in country  $c$  at date  $T$  could be expressed as follows:

$$Know_{T,c,i} = \sum_{t=0}^T \alpha_{t,c,d,i} N_{t,d,i} \cdot e^{-\beta_1 \beta_{1,d} \beta_{1,i} (T-t)} \left(1 - e^{-\beta_2 (T-t)}\right)$$

Where  $t = ct d$  and  $d$  is the origin of knowledge. Initial stock was built with perpetual inventory method (PIM), see chapter 4. Unlike the previous chapter, the foreign stock is not a sum of all the stocks of foreign countries because we have treated foreign knowledge as coming from only one country. Figure 5 & 6 plot knowledge stocks for solar energy and heat exchange technology. In all cases, figures highlight the importance of foreign knowledge even for the US. Popp (2002) find decreasing return to energy research over time both through his downward trend in his productivity estimates and in his stocks that falling also over time. In Appendix, in Figure 7.10, we present US knowledge stocks for the nine other technologies. We also find decreasing value of the domestic stocks and this especially during the nineties. In all cases the diminishing domestic returns is accompanied by increasing foreign returns that seem to compensate the domestic decrease at the exception of the heat pump technology and maybe of the solar and wind technologies.

## 7.5 Econometrical results

Having constructed knowledge stock for each country and each technology, we move to the estimation of the induced innovation relationship. The equation to estimate derived from model (7.1) is:

## Chapter 7. Demand-Pull and Technological-Push mechanisms in Energy-Efficient Technologies

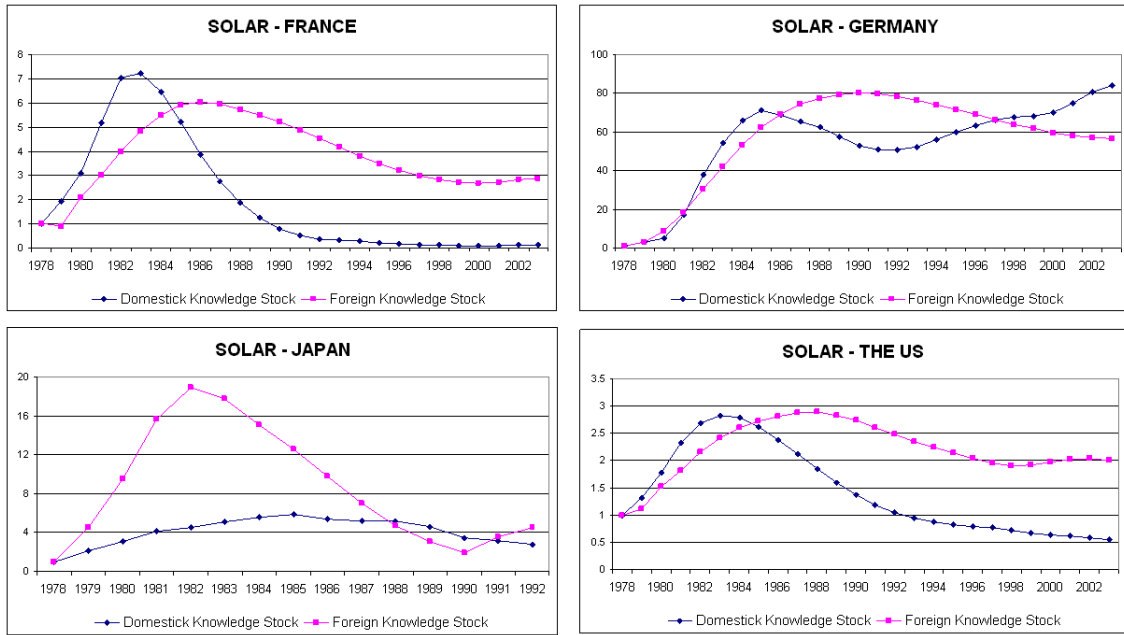


Figure 7.8: Solar Knowledge stocks

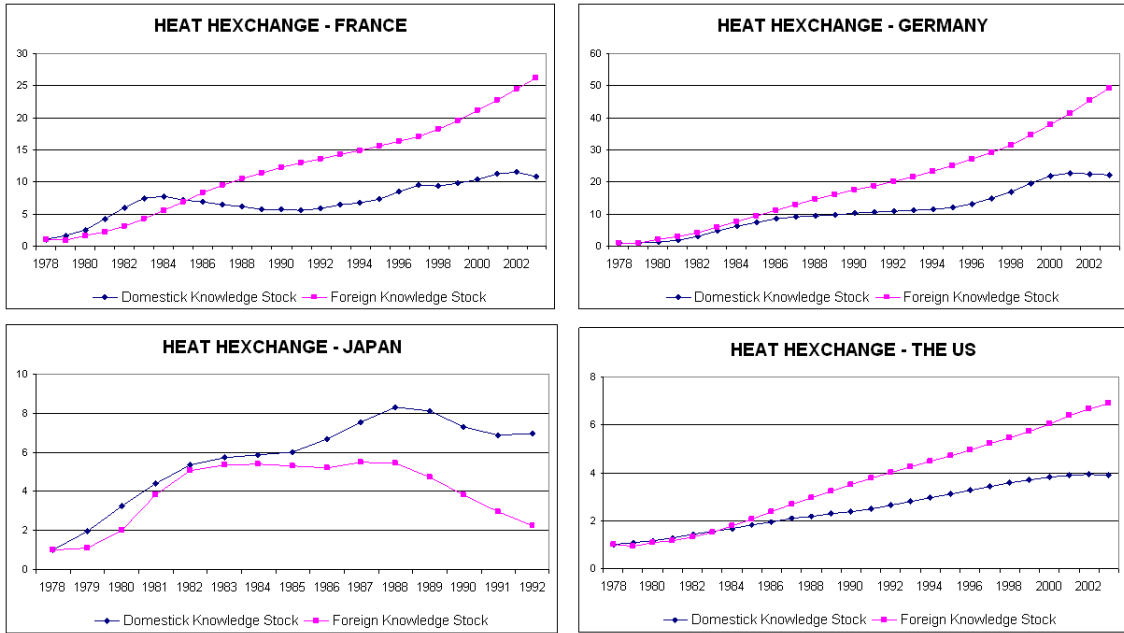


Figure 7.9: Heat Exchange Knowledge Stocks

## Chapter 7. Demand-Pull and Technological-Push mechanisms in Energy-Efficient Technologies

---

$$\begin{aligned} \text{Log} \left( \frac{\text{patent}_{t,c,i}}{TPatent_{t,c}} \right) = cst + \gamma_1 \text{Log} \cdot P_{t-1,c} + \gamma_2 \text{Log} \cdot \text{Know}_{t,c,c,i} + \gamma_3 \text{Log} \cdot \text{Know}_{t,c,-c,i} \\ + \gamma_4 \text{Log} \cdot \text{Policies}_{t,c} + \phi_c + \phi_i + \phi_t + \varepsilon_{t,c,i} \end{aligned} \quad (7.4)$$

where  $i = 1, \dots, 11$  indexes technologies,  $c = 1, \dots, 4$  indexes countries and  $t = 1978, \dots, 2003$  indexes time. Fixed effects are added for county  $\phi_c$ , technology  $\phi_i$  and year  $\phi_t$ . Due to auto correlation we use the Cochrane–Orcutt transformation. We pooled the countries and the technologies groups to obtain single estimates for each parameter. Regression will proceed in two steps. The first step will implement regression with no distinction between kind of innovations, thus product and process technologies will be put together. In a second time, regressions will be done depending on innovation categories. First step results of equation (3) are presented in Table 7.3.

All technologies		
Independant variables	Estimates	S.E.
Lagged Energy Price	0.974	0.308 **
Domestic Know Stock	0.390	0.049 **
Foreign Know Stock	0.451	0.041 **
Policies	0.092	0.068
Nbs of obs	982	
R-squared	0.75	

Notes: \* and \*\* refer to 5%, and 1% level of statistical significance  
Country, technology and time dummy are included

Table 7.3: All technologies estimation results

Main interest result is the effect of energy price on innovating activity. As shown in Table 7.3, energy price plays an important role in inducing new energy innovations. Elasticity of lagged energy price on innovation is positive and significant with a value close to unity at 0.974. A change of 10% in energy prices induces a change of 9.74% in patenting activity. Regression shows that not only prices play an important role in determining the level of innovating activity, but the knowledge available to inventor is also an important factor. Both domestic and foreign knowledge are positives and significant with a respective value of 0.390 and 0.451. Foreign knowledge has a higher effect on innovative activities than domestic knowledge: this is a common characteristic in knowledge spillovers literature. If we sum the two knowledge estimates, value is below price elasticity at 0.741. Lastly, Policies variable is positive but not significant.

Table 7.4 shows results for product technologies only, i.e. biomass, fuel cell, geothermal, hydrogen,



Product technologies		
Independant variables	Estimates	S.E.
Lagged Energy Price	1.348	0.421 **
Domestic Know Stock	0.338	0.062 **
Foreign Know Stock	0.416	0.058 **
Policies	0.142	0.075
Nbs of obs	623	
R-squared	0.70	

Notes: \* and \*\* refer to 5%, and 1% level of statistical significance  
Country, technology and time dummy are included

Table 7.4: Product technologies estimation results

solar, tide & wave and wind. Price elasticity appears with a higher value than the unity, meaning that an increase in energy price induces an increase more than proportional of innovation activities. New energy innovations in product technologies are one-third more dependents of energy price than all technologies. Knowledge stocks are still positives and significant but with a lower effect than in previous regression. It seems that product innovations are more directed by energy prices than by technological opportunities overall and relatively to all technologies results. But the main difference appears with Policies variable that becomes significant. Public policy is a significant determinant of patenting in new product technologies.

Table 7.5 presents results for process technologies only, i.e. heat exchange, heat pump, Stirling engines and waste heat recovery. Energy price elasticity is still positive and significant with a value close to unity which is very similar to the first regression. Knowledge stocks play again a positive and significant role on new energy innovations. The gap between the two knowledge elasticities is higher than previously, meaning that innovators benefit even more from foreign knowledge.

Process technologies		
Independant variables	Estimates	S.E.
Lagged Energy Price	0.975	0.496 *
Domestic Know Stock	0.312	0.098 **
Foreign Know Stock	0.533	0.068 **
Nbs of obs	373	
R-squared	0.81	

Notes: \* and \*\* refer to 5%, and 1% level of statistical significance  
Country, technology and time dummy are included

Table 7.5: Process technologies estimation results

Results of estimations show that both energy price and technology opportunity play a crucial role in determining the level of energy saving innovation. Higher energy prices provide incentive for increased

patenting activities if there is no decrease in the stock of knowledge. Knowledge elasticity appears to be lower than price elasticity with a cumulative value around 0.8. Patenting reacts to knowledge opportunity changes but with lower effect; innovating activity seems to be most directed by energy prices than by knowledge opportunities. Results stress the importance of international knowledge spillovers in energy innovations, particularly in energy process innovations where the elasticities is around 0.5. Finally, results suggest that public policies play a significant role in inducing innovations in renewable energies.

### 7.6 Conclusion

This paper validates induced innovation theory for energy saving technologies for four main countries (France, Germany, Japan and the US) over the period 1978-2003. The estimations carried out show a strong influence of energy prices variations on innovating activities, measured by the successful patent applications. A rise in the prices of energy will bring relatively quickly to a rise in patent application, representing an increase of inventions activities in the field of energy efficiency technologies. A change of 10% in energy price induces by and large a change of 10% in patenting activities toward energy efficient technologies. Our results also confirm the importance of the supply side in the determination of innovation process and highlight the role of international knowledge spillovers in energy efficiency technologies. Domestic and foreign knowledge, both constructed with patent citation data, present an elasticity of 0.39 and 0.451 that is not so far than energy price elasticity. Estimations suggest that induced innovation plays a key role, but is only possible if technological opportunities are presents. Our results also confirm the importance of public policies as driver for renewable technologies innovations. Finally, the diminishing returns to energy research suggested by Popp (2002), is not obvious in our results if we take into account the role of international knowledge externalities.

### 7.7 Appendix

#### 7.7.1 Definitions of energy technologies

Biomass

C10L 05/4. Solid fuels (produced by solidifying fluid fuels - essentially based on materials of non-mineral origin

C10L 01/14. Liquid carbonaceous fuels - organic compounds

F02B 43/08. Engines operating on gaseous fuels from solid fuel - e.g. wood  
B01J 41 /16. Anion exchange; Use of material cellulose and wood  
C10B 53/02. Destructive distillation, specially adapted for particular solid raw materials or solid raw materials in special form of cellulose-containing material

#### Fuel-Cell

H01M 08/\*\*. Fuel cells; Manufacture thereof  
H01M 04/86. Inert electrodes with catalytic activity, e.g. for fuel cells

#### Geothermal

F03G 04/\*\*. Devices for producing mechanical power from geothermal energy  
F24J 03/\*\*. Other production or use of heat, not derived from combustion - using natural or geothermal heat  
H02N 10/00. Electric motors using thermal effects

#### Heat-exchange

F28. Heat exchange in general  
H01P 7/10. Controlling of coolant flow by throttling amount of air flowing through liquid-to-air heat- exchangers  
H01L 23/46. Arrangements for cooling, heating, ventilating or temperature compensation - involving the transfer of heat by flowing fluids

#### Heat-pumps

F24D 03/18. Heat central with heat water using heat pump  
F24D 05/12. Hot-air central heating systems using heat pumps  
F24D 11/02. Other central heating systems using heat pumps  
F24D 15/04. Other domestic- or space-heating systems using heat pumps  
F24D 17/02. Domestic hot-water supply systems using heat pumps  
F24H 04/00. Fluid heaters using heat pumps  
F25B 15/00. Sorption machines, plant, or systems, operating continuously, e.g. absorption type  
F25B 30/00. Heat pumps

#### Hydrogen

C01B 03/\*\*. Hydrogen; Gaseous mixtures containing hydrogen; Separation of hydrogen from mixtures containing it

#### Solar

H02N 6 / 00. Generators in which light radiation is directly converted into electrical energy (solar cells or assemblies thereof H01L 25/00, H01L 31/00)

## Chapter 7. Demand-Pull and Technological-Push mechanisms in Energy-Efficient Technologies

---

- F03G 6 /\*\*. Devices for producing mechanical power from solar energy
- F24J 02/\*\*. Use of solar heat, e.g. solar heat collectors
- F25B 27/00. Machines, plant, or systems, using particular sources of energy - sun
- H01L 25/00. Assemblies consisting of a plurality of individual semiconductor or other solid state device
- H01L 31/042. Semiconductor devices sensitive to infra-red radiation - including a panel or array of photoelectric cells, e.g. solar cells
- E04D 13/18. Roof covering aspects of energy collecting devices, e.g. including solar panels
- B60L 08/00. Electric propulsion with power supply from force of nature, e.g. sun, wind
- B60K 16/00. Arrangements in connection with power supply from force of nature, e.g. sun, wind
- F26B 03/28. Drying solid materials or objects by processes involving the application of heat by radiation – e.g. sun

### Stirling-Engines

- F01B 29/10. Machines or engines with pertinent characteristics other than those provided for in main groups reciprocating-piston machines or engines not otherwise provided for Engines
- F02G 01/043. Hot gas positive-displacement engine plants of closed-cycle type; the engine being operated by expansion and contraction of a mass of working gas which is heated and cooled in one of a plurality of constantly communicating expansible chambers, e.g. Stirling cycle type engines
- F25B 09/14. Compression machines, plant, or systems, in which the refrigerant is air or other gas of low boiling point characterized by the cycle used, e.g. Stirling cycle

### Tide-&-Wave

- F03B 07/00. Water wheels
- F03B 13/12. to F03B 13/26. Adaptations of machines or engines for special use- characterized by using wave or tide energy
- F03G 07/05 Mechanical-power-producing mechanisms - ocean thermal energy conversion
- F03G 07/04 Mechanical-power-producing mechanisms - using pressure differentials or thermal difference

### Waste-Heat-Recovery

- F01K 23/06. Plants characterized by more than one engine delivering power external to the plant the engine cycles being thermally coupled combustion heat from one cycle heating the fluid in another cycle
- F01K 23/10 ...with exhaust fluid of one cycle heating the fluid in another cycle
- F01K 23/14. ...including at least one combustion engine
- F01K 27/02. Plants modified to use their waste heat, other than that of exhaust, e.g. engine -friction heat Methods of steam generation by exploitation of the heat content of hot heat carriers

F22B 01/16. ...the heat carrier being hot liquid or hot vapour, e.g. waste liquid, waste vapour

F22B 01/18. ...the heat carrier being a hot gas, e.g. waste gas

F23G 05/46. Incineration of waste - recuperation of heat

F25B 27/027. Machines, plant, or systems, using waste heat, e.g. from internal-combustion engines

F02G 5/00-04. Hot gas or combustion Profiting from waste heat of exhaust gases

F01K 25/14. Plants or engines characterized by use of industrial or other waste gases

## Wind

F03D 1/\*\*. Wind motors with rotation axis substantially in wind direction (controlling F03D 7/00)

F03D 3/\*\*. Wind motors with rotation axis substantially at right angle to wind direction (controlling F03D 7/00)

F03D 5/\*\*. Other wind motors (controlling F03D 7/00)

F03D 7/\*\*. Controlling wind motors F03D 9/ Adaptations of wind motors for special use; Combinations of wind motors with apparatus driven thereby (aspects predominantly concerning driven apparatus, see the relevant classes for such apparatus)

F03D 11/\*\*. Details, component parts, or accessories not provided for in, or of interest apart from, the other groups of this subclass

B63H 13/\*\*. Effecting propulsion by wind motors driving waterengaging propulsive elements

B60L 8/\*\*. Electric propulsion with power supply from force of nature, e.g. sun, wind

7.7.2 Tables

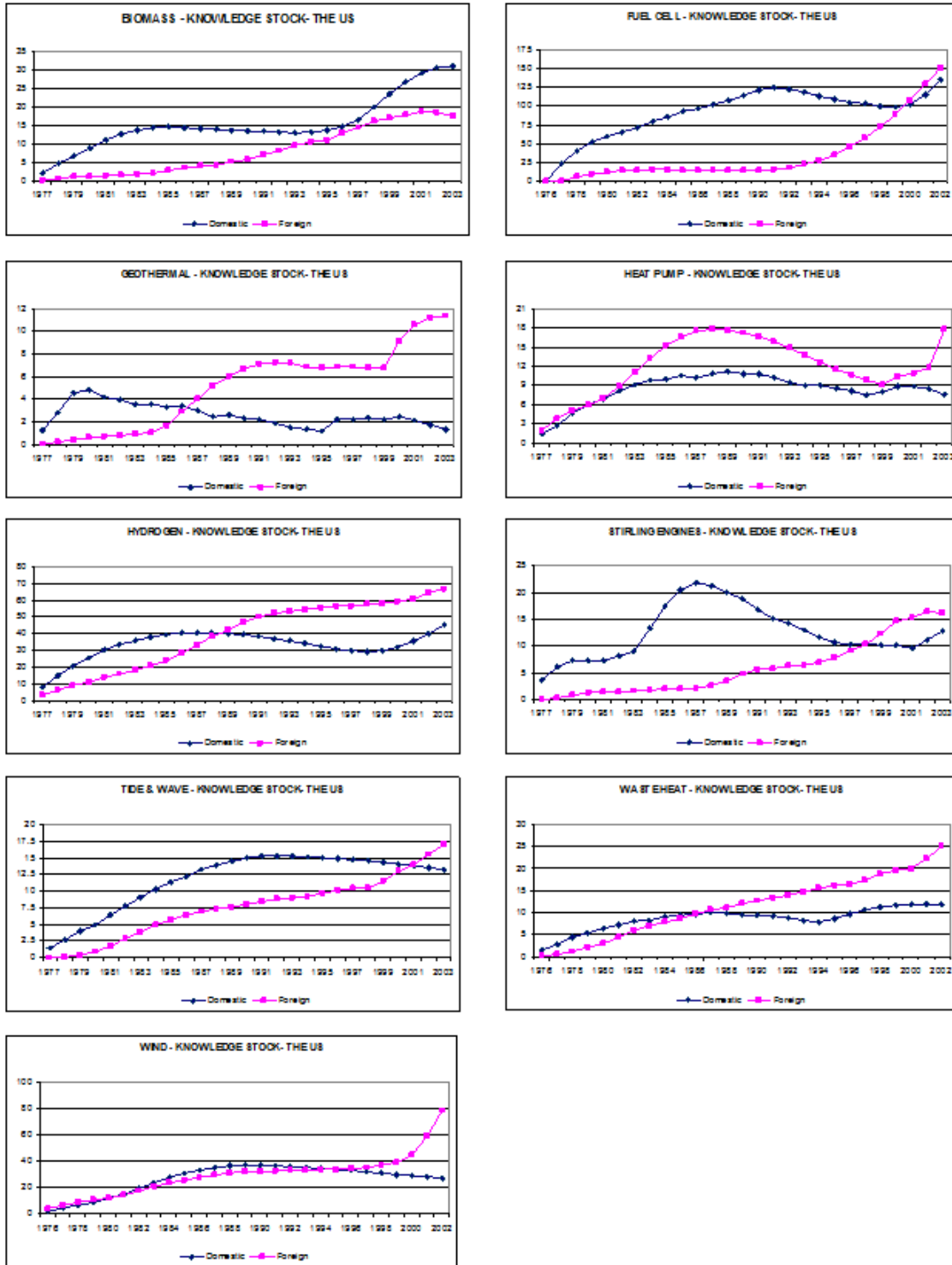


Figure 7.10: US knowledge stocks in energy efficiency technologies

Chapter 7. Demand-Pull and Technological-Push mechanisms in Energy-Efficient Technologies

	France		Germany		Japan		The US	
Adjusted R-squared	0.278		0.387		0.506		0.462	
Std. error of regression	0.047		0.031		0.049		0.107	
	Estimates	S.E	Estimates	S.E	Estimates	S.E	Estimates	S.E
<b>Citing year effect</b>								
1978-1979	1	n.a	1	n.a	1	n.a	1	n.a
1980-1983	1.056	0.083	0.479	0.085	0.853	0.048	0.993	0.043
1984-1987	1.220	0.077	0.529	0.059	1.848	0.045	0.786	0.030
1988-1991	1.072	0.076	0.786	0.083	0.843	0.045	0.702	0.027
1992-1995	0.808	0.073	0.940	0.088	0.284	0.057	1.036	0.025
1996-1999	1.157	0.064	1.277	0.074	.	n.a	1.405	0.025
2000-2003	1.103	0.064	1.837	0.109	.	n.a	1.299	0.025
2004-2006	0.944	0.091	1.343	0.089	.	n.a	0.631	0.029
<b>Cited inventor country code</b>								
Domestic	1	n.a	1	n.a	1	n.a	1	n.a
Foreign	0.125	0.026	0.076	0.010	0.028	0.008	0.147	0.014
<b>Technological field</b>								
Biomass	2.268	0.612	1.601	0.316	4.318	2.413	1.048	0.153
Fuel Cell	4.049	1.359	8.622	0.521	8.883	2.266	4.699	0.162
Geothermal	0.977	0.406	0.951	0.144	0.283	0.105	1.306	0.166
Heat Pump	0.861	0.290	1.682	0.386	2.045	0.753	2.427	0.171
Heat Exchange	1	n.a	1	n.a	1	n.a	1	n.a
Hydrogen	1.509	0.369	2.374	0.307	2.704	0.628	1.669	0.146
Solar	2.080	0.428	2.964	0.250	1.643	0.316	2.497	0.237
Stirling Engines	3.229	1.601	0.892	0.258	3.407	1.197	3.120	0.229
Tide & Wave	0.371	0.127	0.570	0.164	1.881	0.869	0.402	0.042
Waste Heat	2.309	1.030	1.677	0.252	1.530	0.316	1.244	0.200
Wind	1.402	0.328	3.324	0.229	2.174	0.904	1.022	0.072
<b>Beta 1</b>	0.501	0.017	0.345	0.011	0.584	0.036	0.170	0.004
<b>Beta 2</b>	0.054	0.003	0.027	0.002	0.061	0.006	0.076	0.003
<b>Rate of obsolescence :</b>								
<i>by cited inventor country code</i>								
Domestic	1	n.a	1	n.a	1	n.a	1	n.a
Foreign	0.292	0.030	0.361	0.033	0.681	0.115	0.613	0.050
<i>by technological group</i>								
Biomass	1.493	0.210	1.281	0.156	1.520	0.407	1.199	0.131
Fuel Cell	1.876	0.321	1.852	0.067	1.268	0.150	1.249	0.032
Geothermal	0.857	0.255	0.713	0.072	0.401	0.156	1.275	0.120
Heat Pump	1	n.a	1	n.a	1	n.a	1	n.a
Heat Exchange	0.678	0.144	1.506	0.210	0.965	0.157	1.299	0.067
Hydrogen	0.983	0.129	1.272	0.101	1.156	0.141	1.153	0.076
Solar	1.290	0.137	1.094	0.058	1.190	0.114	1.735	0.112
Stirling Engines	1.072	0.233	1.106	0.200	1.198	0.209	1.677	0.085
Tide & Wave	0.559	0.173	0.777	0.147	0.887	0.224	0.219	0.043
Waste Heat	1.860	0.431	1.104	0.104	0.817	0.097	1.505	0.171
Wind	1.102	0.141	1.021	0.044	0.851	0.178	0.571	0.040

Table 7.6: Patent citation function estimation results

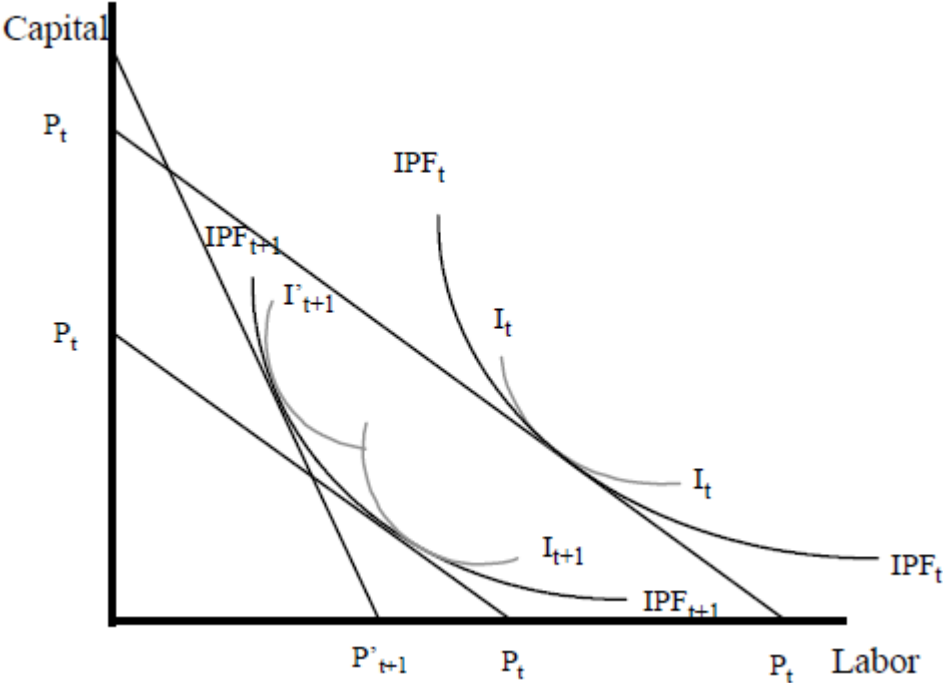


Figure 7.11: Ahamd's induced innovation model





## Chapter 8

# Conclusions

Knowledge spillovers determine the long run economic growth due to the constant social return that they can induce. The econometric literature on this subject was particularly abundant during the last three decades. Unfortunately “Knowledge flows...are invisible; they leave no paper trail by which they may be measured and tracked” Krugman (1991). According to this, empirical studies have focused for a long time on the effect of one external stock of knowledge on the innovative activity outcome of an agent. Economists can only measure knowledge spillovers indirectly without the advent of new databases on patent. A patent is a set of exclusive rights granted by a state (or regional office) to an inventor or his assignee for a limited period of time in exchange for a public disclosure of an invention. One aspect of this disclosure is the citation and can be interpreted as a track for knowledge diffusion and be directly used to measure different aspects of the knowledge based economy.

The main objectives of the present dissertation were threefold. Firstly, we wanted to analyze how knowledge spillovers flow and what are the main determinants that underlie their empirical impact. Secondly, we determined to what extent we can use patent and patent citation data, taken from the PATSTAT database in order to have a direct look into the knowledge diffusion patterns. Finally, we proposed a new input-based indicator built on these data that can be very helpful when traditional indicators are missing.

## 8.1 Main conclusions

- Concerning traditional studies, it exists, in the literature, numerous channels associated to knowledge flows, all of them lying on different assumptions. The reported estimates are significantly different according to these different channels and are statistically significant even after being controlled by the possible presence of a publication bias effect. Nonetheless, the presence of the publication effect has been econometrically proved and makes knowledge empirical effects larger than they are because researches with statistically positive and significant results are most likely to be reported in comparison to those with null or no significant results. Finally, the spillovers effects are increasing over time, since the estimates reported in the literature are higher if the studies refer explicitly to data covering a period after 1990.
- The massive use of patent citations from the early 1990s has opened a “new window” on the knowledge economy, by increasing the number of possible measure of technological change with new perspectives. Specifically patent citations provide a direct measure of knowledge diffusion renewing the knowledge spillovers literature.
- Concerning the empirical analysis of patent citation between the DPMA, INPI, IPO, EPO JPO, USPTO and the WIPO, we find evidence of a domestic bias effect of citations toward patents belonging to the same office. This effect can be partly straightened by prior equivalence correction.
- When we explore the similarities of family search report between national offices and the EPO, we find very strong similarities with INPI and JPO, these offices present a respective matching rate of 87.2 % and 58.9 %. Concerning the DPMA and IPO the respective matching rates are quite lower with a respective value of about 44 % and 40 %.
- The use of patent citations both drawn up by examiners of the DPMA, EPO, INPI, IPO, JPO and USPTO are a good measure for knowledge diffusion. This assessment is supported by econometric tests on the explanatory power of our input-based indicator, built with international patent citations data taken from these offices, on domestic productivity growth.
- International patent and patent citation data can provide a measure of knowledge stock that could be very helpful when data on R&D are not available. This indicator is the most useful because it allows a direct search for knowledge spillovers.

- Concerning international knowledge diffusion patterns, the conclusions are twofold. First, domestic knowledge is predominant, it diffuses almost twice as fast as the foreign knowledge, although the probability of domestic citations compared to those from abroad decreases with time. Second, when foreign knowledge is controlled by its technological productivity, the geographic location effect seems to disappear.
- Regarding the application of our indicator on a case study on the determinants of innovation in energy efficiency technologies, we find that innovations are more directed by demand pull mechanisms than by technological push ones. Nevertheless, both domestic and foreign knowledge have a positive and significant impact on energy efficiency innovations with a predominance effect of international knowledge spillovers.

## 8.2 Limits

The set of indicators employed here, based on patent and patent citation data is by no means exhaustive. A number of criticisms could be raised against the empirical use of these data for measuring the creation of knowledge and the knowledge diffusion patterns.

- The first set of drawbacks are intrinsic to patents and are well identified in the literature, namely: not all inventions meet the patentability criteria set by the different patent offices; inventor has to make the strategic decision to patent, so not all patentable inventions are patented and, finally, there are several differences among technological sectors, country and time. Nevertheless, we have followed the famous Schmookler dictum: "We have a choice of using patent data statistics cautiously and learning what we can from them, or not using them and learning nothing about what they alone can teach us" Schmookler (1966, p.56).
- The second drawback comes from the PATSTAT database and concerns the percentage of inventor country-code available. Although inventor country code coverage is quite high for DPMA, EPO, INPI, WIPO and USPTO that is not the case for JPO and IPO. This is not a big deal for JPO because almost all prior applications are made by Japan applicants, but it becomes problematic for data taken from the IPO because 20% of first patent applications are not made by English applicants. This situation restricts significantly the sample for the United-Kingdom and raises a barrier to the implementation of several case studies as the one done in Chapter 7 where the UK data were not included.

### 8.3 Future researches

There is clearly room for further works and several empirical extensions would be worthwhile:

- In this dissertation, we have only considered patent citations to be only intra-sectoral or intra-technological, investigating only the international spillovers. An obvious extension of this study will be to extend our works in order to take into account inter-sectoral spillovers both at a national and international level. Several studies conducted on the USPTO data have investigated these issues but their aggregation level is quite high, they generally consider six technologies (Chemical, Computer and Comm, Drug and Medical, Elec. & Electronics, Mechanical and Other). Here, to increase our understanding of knowledge diffusion across fields it would be natural to extend our analysis toward inter-technological citations according to our database and our methodologies developed. The Schmoch concordance table allows the identification of 44 technologies, in the present work; due to time constraints, we have grouped these 44 technologies into 13 ones. Nevertheless it would be easy to consider these 44 technological fields.
- Another fundamental concern of knowledge spillovers is how knowledge flows from public toward private sector. In this thesis, we have not make any distinctions between public and private developed technology. An extending way would be to investigate the dissemination of knowledge between University (and Public Laboratory) and private sector. Once the origin of knowledge is identified, it is possible to investigate all possible combinations, i.e. between Firm and University or University and University or Firm and Firm.
- Concerning the geographical pattern of knowledge diffusion, we have only considered a small set of five countries. Moreover, a major limitation of the study of knowledge geographical pattern with the Patstat database is the data quality about the address of inventors. One solution could be provided by the OECD with the REGPAT database. The REGPAT database link patent data to regions according to the addresses of the applicants and inventors. The level of desaggregation is important since more than 5000 regions are covered (i.e. NUTS 3<sup>1</sup>).
- Concerning the citation function, transformation or addition of new determinants can be easily achieved, allowing thus to analyze new aspects of knowledge diffusion. For instance, we have considered the obsolescence and the diffusion parameter to be constant over the period, it should

---

<sup>1</sup>Nomenclature Teritorial Units for Statistics.

## Chapter 8. Conclusions

---

be interesting to allow these parameters to vary over time. Such findings could assess the impacts of different public policies designed to increase the rate of knowledge diffusion.

- It is commonly assumed that the triadic patents present higher technological or/and monetary value. It might be interesting to study how this pattern is reflected in the patent citation data, i.e to what extent the number of citation received by a patent family depends on the size of this family (and if these patents are triadic or not).
- Finally, we have applied our indicator on energy technologies, but there are other specific cases for which our indicator can be helpful. For instance, methodology of Chapter 7 could be used for the investigation of the General Purpose Technology (for example nanotechnologies). GDP are at the source of important externalities, but their impact on the economy depends on the complementarity between GPT inventions and the development of applications in the general dynamics of technology that could be assessed by our patent citation function.

Modelling and measuring scientific production is not an easy task, given the fact that science inputs and outputs, are difficult to quantify in terms of both quantity and quality. Nevertheless, this dissertation has tried to shed some light on these issues, and particularly it has attended to provide several new perspectives on the measurement of knowledge spillovers and on how to model the cumulative process of knowledge creation.



# Bibliography

- [1] James D Adams. Fundamental Stocks of Knowledge and Productivity Growth. *Journal of Political Economy*, 98(4):673–702, August 1990.
- [2] James D. Adams and Adam B. Jaffe. Bounding the Effects of R&D: An Investigation Using Matched Establishment-Firm Data. *RAND Journal of Economics*, 27(4):700–721, Winter 1996.
- [3] Philippe Aghion and Peter Howitt. A Model of Growth through Creative Destruction. *Econometrica*, 60(2):323–51, March 1992.
- [4] Ajay Agrawal, Devesh Kapur, and John McHale. How do spatial and social proximity influence knowledge flows? Evidence from patent data. *Journal of Urban Economics*, 64(2):258–269, September 2008.
- [5] S. Ahmad. On the theory of induced innovation. *Economic Journal*, 76:344 – 357, 1966.
- [6] Francesco Aiello and Paola Cardamone. Similarity and Geographical Issues in Evaluating the Impact of R&D Spillovers at Firm Level: Evidence from Italy. *Icfai University Journal of Applied Economics*, 0(4):52–72, July 2007.
- [7] Francesco Aiello and Paola Cardamone. R&D spillovers and firms performance in Italy. *Empirical Economics*, 34(1):143–166, February 2008.
- [8] Juan Alcacer and Michelle Gittelman. How do I Know what you Know? Patent Examiners and the Generation of Patent Citations. SSRN working paper., 2004.
- [9] Roland Berger Market Analysis. Study on the costs of patenting. Technical report, Final report prepared for the EPO, 2005.



- [10] Dolores Anon Higon. The impact of R&D spillovers on UK manufacturing TFP: A dynamic panel approach. *Research Policy*, 36(7):964–979, September 2007.
- [11] Kenneth J. Arrow. The Economic Implications of Learning by Doing. *Review of Economic Studies*, No. 29 (June):155–173, 1962.
- [12] Scott E. Atkinson and Robert Halvorsen. A New Hedonic Technique for Estimating Attribute Demand: An Application to the Demand for Automobile Fuel Efficiency. *The Review of Economics and Statistics*, 66(3):417–26, August 1984.
- [13] E. Bacchiocchi and F. Montobbio. Knowledge diffusion from university and public research. A comparison between US, Japan and Europe using patent citations. *The Journal of Technology Transfer*, 34(2):169–181, April 2009.
- [14] S. Bergeron and C. Le bas. Redistribution technologique et diffusion intersectorielle de l'innovation. Technical report, LESA, Université Lyon II, 1997.
- [15] Jeffrey I Bernstein. The Structure of Canadian Inter-industry R&D Spillovers, and the Rates of Return to R&D. *Journal of Industrial Economics*, 37(3):315–28, March 1989.
- [16] Jeffrey I Bernstein and M Ishaq Nadiri. Research and Development and Intra-industry Spillovers: An Empirical Application of Dynamic Duality. *Review of Economic Studies*, 56(2):249–67, April 1989.
- [17] T.H.A. Bijmolt and R. Pieters. Meta-analysis in marketing when studies contain multiple measurements. *Marketing Letters*, 12:157–169, 2001.
- [18] Jürgen Bitzer and Monika Kerekes. Does foreign direct investment transfer technology across borders? New evidence. *Economics Letters*, 100(3):355–358, September 2008.
- [19] Nick Bloom, Mark Schankerman, and John Van Reenen. Identifying technology spillovers and product market rivalry. *Proceedings*, 2005.
- [20] Laura Bottazzi and Giovanni Peri. Innovation and spillovers in regions: Evidence from European patent data. *European Economic Review*, 47(4):687–710, August 2003.
- [21] Lee G. Branstetter. Are knowledge spillovers international or intranational in scope?: Microeconomic evidence from the U.S. and Japan. *Journal of International Economics*, 53(1):53–79, February 2001.

## BIBLIOGRAPHY

---

- [22] Thomas Brenner. Local Knowledge Resources and Knowledge Flows. *Industry & Innovation*, 14(2):121–128, 2007.
- [23] S. Breschi and F. Lissoni. Knowledge networks from patent data: methodological issues and research targets. In *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems*. H. Moed, W. Glänzel & U. Schmoch, 2004.
- [24] M. Brown and H. A. Conrad. *The Theory and Empirical Analysis of Production, vol. 3 of Studies in Income and Wealth*, chapter The Influence of Research and Education on CES Production Relations, pages 275–340. New York: Columbia University Press, 1967.
- [25] R. Caballero and A. Jaffe. How High are the Giants’ Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth. *National Bureau of Economic Research Macroeconomics Annual, MIT Press in O. Blanchard and S. Fischer, eds*, 8, 1993.
- [26] John S Chipman. External Economies of Scale and Competitive Equilibrium. *The Quarterly Journal of Economics*, 84(3):347–85, August 1970.
- [27] Michele Cincera. Firms productivity growth and R&D spillovers: An analysis of alternative technological proximity measures. *Economics of Innovation and New Technology*, 14(8):657–682, November 2005.
- [28] Michele Cincera and Bruno Van Pottelsberghe de la Potterie. International R&D Spillovers: a Survey. *Brussels Economic Review/Cahiers Economiques de Bruxelles*, 0(169):3–31, 2001.
- [29] Maria Cipollina and Luca Salvatici. Reciprocal trade agreements in gravity models: a meta-analysis. Economics & Statistics Discussion Papers esdp07035, University of Molise, Dept. SEGeS, April 2007.
- [30] David T. Coe and Elhanan Helpman. International R&D spillovers. *European Economic Review*, 39(5):859–887, May 1995.
- [31] Wesley M Cohen and Daniel A Levinthal. Innovation and Learning: The Two Faces of R&D. *Economic Journal*, 99(397):569–596, September 1989.

- [32] Australian Industry Commission. Research and Development. Technical report, Australian Government Publishing Service, Report no 44, 1995.
- [33] Timothy G. Conley and Ethan Ligon. Economic Distance and Cross-Country Spillovers. *Journal of Economic Growth*, 7(2):157–87, June 2002.
- [34] Robin Cowan and Dominique Foray. The Economics of Codification and the Diffusion of Knowledge. *Industrial and Corporate Change*, 6(3):595–622, September 1997.
- [35] Katrin Cremers. Settlement during patent litigation trials. An empirical analysis for Germany. *The Journal of Technology Transfer*, 34(2):182–195, April 2009.
- [36] Jesus Crespo-Cuaresma, Neil Foster, and Johann Scharler. Barriers to technology adoption, international R&D spillovers and growth. *Economics Bulletin*, 15(3):1–7, 2008.
- [37] Paola Criscuolo. The 'home advantage' effect and patent families. A comparison of OECD triadic patents, the USPTO and the EPO. *Scientometrics*, 66(1):23–41, 2006.
- [38] Paola Criscuolo, Rajneesh Narula, and Bart Verspagen. Role of home and host country innovation systems in r&d internationalisation: a patent citation analysis. *Economics of Innovation and New Technology*, 14(5):417–433, July 2005.
- [39] A. Dechezlepretre, M. Glachant, H. Ivan, J. Nick, and Y. Ménière. Invention and Transfer of Climate Mitigation Technologies on a Global Scale: A Study Drawing on Patent Data. Technical report, Report for AFD (Agence Française de Développement), 2008.
- [40] Giovanni Dosi. Technological paradigms and technological trajectories : A suggested interpretation of the determinants and directions of technical change. *Research Policy*, 11(3):147–162, June 1982.
- [41] Nigel Driffield, James Love, and Karl Taylor. Productivity and Labour Demand Effects of Inward and Outward FDI on UK Industry. Working Papers 20080001, The University of Sheffield, Department of Economics, January 2008.
- [42] Thomas Döring and Jan Schnellenbach. What do we know about geographical knowledge spillovers and regional growth?: A survey of the literature. *Regional Studies*, 40(3):375–395, May 2006.

## BIBLIOGRAPHY

---

- [43] Louis-Marc Ducharme and Pierre Mohnen. Externalités et taux de rendement sociaux de la R&D. *Économies et Sociétés, série Dynamique technologique et organisation*, 8:189–213, 1996.
- [44] Emmanuel Duguet and Megan MacGarvie. How well do patent citations measure flows of technology? Evidence from French innovation surveys. *Economics of Innovation and New Technology*, 14(5):375–393, July 2005.
- [45] Matthias Egger, George Davey Smith, Martin Schneider, and Christoph Minder. Bias in meta-analysis detected by a simple, graphical test. *BMJ, British medical journal*, 315(7109):629–634, 1997.
- [46] A. Steven Englander, Robert Evenson, and Masaharu Hanazaki. R&D, innovation and the total factor productivity slowdown. *Revue économique de l'OCDE*, 11, 1988.
- [47] R. E. Evenson. *The contribution of agricultural research and extension to agricultural production*. PhD thesis, University of Chicago, 1968.
- [48] R. E. Evenson, S. Kortum, and E. J. Putnam. Estimating Patents by Industry Using Yale Canada Patent Concordance. Technical report, Yale University, Mimeo, 1989.
- [49] Bernard Felix. Patent procedures and statistics: an overview. Technical report, Eurostat, 2006.
- [50] Michael Funke and Annkatrin Niebuhr. Regional Geographic Research and Development Spillovers and Economic Growth: Evidence from West Germany. *Regional Studies*, 39(1):143–153, January 2005.
- [51] A. Gambardella, P. Giuri, and M. Mariani. Study on evaluating the knowledge economy. What are patents actually worth? The value of patents for today economy and society. Technical report, Commission européenne, project ETD/2004/IM/E3/77, 2005. Detailed Technical Studies.
- [52] C. Gay and C. Le Bas. Uses without too many abuses of patent citations or the simple economics of patent citations as a measure of value and flows of knowledge. *Economics of Innovation and New Technology*, 14(5):333–338, July 2005.
- [53] Kenneth Gillingham, Richard G. Newell, and Karen L Palmer. Energy Efficiency Policies: A Retrospective Examination. *Annual Review of Environment and Resources*, 31, 2006.

- [54] Pinelopi Koujianou Goldberg. The Effects of the Corporate Average Fuel Efficiency Standards in the US. *Journal of Industrial Economics*, 46(1):1–33, March 1998.
- [55] Akira Goto and Kazuyuki Motohashi. Construction of a Japanese Patent Database and a first look at Japanese patenting activities. *Research Policy*, 36(9):1431–1442, November 2007.
- [56] Akira Goto and Kazuyuki Suzuki. R&D Capital, Rate of Return on R&D Investment and Spillover of R&D in Japanese Manufacturing Industries. *The Review of Economics and Statistics*, 71(4):555–64, November 1989.
- [57] S. Greif and G. Potkowik. Patente und Wirtschaftszweige, Zusammenführung der Internationalen Patentklassifikation und der Systematik der Wirtschaftszweige. Technical report, Heymann, Köln, 1990.
- [58] Rachel Griffith, Stephen Redding, and John Van Reenen. R&D and Absorptive Capacity: Theory and Empirical Evidence. *Scandinavian Journal of Economics*, 105(1):99–118, 03 2003.
- [59] Rachel Griffith, Stephen Redding, and John Van Reenen. Mapping the Two Faces of R&D: Productivity Growth in a Panel of OECD Industries. *The Review of Economics and Statistics*, 86(4):883–895, December 2004.
- [60] Zvi Griliches. Research Costs and Social Returns: Hybrid Corn and Related Innovations. *Journal of Political Economy*, 66:419, 1958.
- [61] Zvi Griliches. Research Expenditures, Education, and the Aggregate Agricultural Production Function. *American Economic Review*, 54(6):961–974, 1964.
- [62] Zvi Griliches. *Price indexes and quality change: Studies in new methods of measurement*. Harvard University Press, 1971.
- [63] Zvi Griliches. *Science and technology in economic growth*, chapter Research expenditures and growth accounting, pages 59–95. London, Macmillan, 1973.
- [64] Zvi Griliches. Issues in Assessing the Contribution of Research and Development to Productivity Growth. *Bell Journal of Economics*, 10(1):92–116, Spring 1979.
- [65] Zvi Griliches. Patent Statistics as Economic Indicators: A Survey. *Journal of Economic Literature*, 28(4):1661–1707, December 1990.

## BIBLIOGRAPHY

---

- [66] Zvi Griliches. The Search for R&D Spillovers. *Scandinavian Journal of Economics*, 94(0):S29–47, Supplemen 1992.
- [67] Zvi Griliches and Frank Lichtenberg. Interindustry Technology Flows and Productivity Growth: A Re-examination. *The Review of Economics and Statistics*, 66(2):324–29, May 1984.
- [68] Gene M Grossman and Elhanan Helpman. Quality Ladders in the Theory of Growth. *Review of Economic Studies*, 58(1):43–61, January 1991.
- [69] Dominique Guellec and Bruno Van Pottelsberghe de la Potterie. 'From R&D to Productivity Growth: Do the Institutional Settings and the Source of Funds of R&D Matter? *Oxford Bulletin of Economics and Statistics*, 66(3):353–378, 07 2004.
- [70] Bronwyn H. Hall. *The Private and Social Returns to Research and Development: What Have We Learned?*, chapter Technology, R&D, and the Economy, pages 140–183. The Brookings Institution and the American Enterprise Institute, 1996.
- [71] Bronwyn H. Hall, Adam B. Jaffe, and Manuel Trajtenberg. The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. (8498), October 2001.
- [72] P. Hanel. Inter-industry flows of technology: an analysis of the Canadian patent matrix and input-output matrix for 1978-1989. *Technovation*, 14(8):529–548, 1994.
- [73] Dietmar Harhoff, Francis Narin, F. M. Scherer, and Katrin Vopel. Citation Frequency And The Value Of Patented Inventions. *The Review of Economics and Statistics*, 81(3):511–515, August 1999.
- [74] Dietmar Harhoff and Markus Reitzig. Determinants of opposition against EPO patent grants; the case of biotechnology and pharmaceuticals. *International Journal of Industrial Organization*, 22(4):443–480, April 2004.
- [75] Rebecca Henderson, Adam Jaffe, and Manuel Trajtenberg. Patent citations and the geography of knowledge spillovers: A reassessment: Comment. *American Economic Review*, 95(1):461–464, March 2005.
- [76] John R Hicks. *The theory of wage*. Macmillan, London, 1932.

- [77] Albert G. Z. Hu and Adam B. Jaffe. Patent citations and international knowledge flow: the cases of Korea and Taiwan. *International Journal of Industrial Organization*, 21(6):849–880, June 2003.
- [78] R. E. Huffman, W. E. & Evenson. *Science for agriculture*. Iowa state university press, 1991.
- [79] IEA. Renewable Energy - Market and Policy Trends in IEA Countries. Technical report, International Energy Agency, Paris, France, 2004.
- [80] Adam B Jaffe. Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value. *American Economic Review*, 76(5):984–1001, December 1986.
- [81] Adam B Jaffe. Demand and Supply Influences in R&D Intensity and Productivity Growth. *The Review of Economics and Statistics*, 70(3):431–37, August 1988.
- [82] Adam B. Jaffe. The importance of "spillovers" in the policy mission of the advanced technology program. *The Journal of Technology Transfer*, (2):11–19, 1998.
- [83] Adam B. Jaffe and Manuel Trajtenberg. Flows of Knowledge from Universities and Federal Laboratories. *Proceeding of the National Academy of Sciences*, 93:12671–12677, 1996.
- [84] Adam B. Jaffe and Manuel Trajtenberg. Flows of Knowledge from Universities and Federal Labs: Modeling the Flow of Patent Citations Over Time and Across Institutional and Geographic Boundari. NBER Working Papers 5712, National Bureau of Economic Research, Inc, August 1996.
- [85] Adam B. Jaffe and Manuel Trajtenberg. International Knowledge Flows: Evidence from Patent Citations. NBER Working Papers 6507, National Bureau of Economic Research, Inc, April 1998.
- [86] Adam B. Jaffe and Manuel Trajtenberg. International Knowledge Flows: Evidence from Patent Citations. *Economics of Innovation and New Technology*,, 8:105–136, 1999.
- [87] Adam B. Jaffe and Manuel Trajtenberg. *Patents, Citations, and Innovations: A Window on the Knowledge Economy*, volume 1 of *MIT Press Books*. The MIT Press, January 2005.
- [88] Adam B. Jaffe, Manuel Trajtenberg, and Michael S. Fogarty. Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors. *American Economic Review*, 90(2):215–218, May 2000.

## BIBLIOGRAPHY

---

- [89] Adam B Jaffe, Manuel Trajtenberg, and Rebecca Henderson. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, 108(3):577–98, August 1993.
- [90] Tim Jeppesen, John A. List, and Henk Folmer. Environmental Regulations and New Plant Location Decisions: Evidence from a Meta-Analysis. *Journal of Regional Science*, 42(1):19–49, 2002.
- [91] Daniel K N Johnson and David Popp. Forced Out of the Closet: The Impact of the American Inventors Protection Act on the Timing of Patent Disclosure. *RAND Journal of Economics*, 34(1):96–112, Spring 2003.
- [92] Daniel K.N. Johnson. The OECD Technology Concordance (OTC): Patents by Industry of Manufacture and Sector of Use. OECD Science, Technology and Industry Working Papers 2002/5, OECD Directorate for Science, Technology and Industry, March 2002.
- [93] Nick Johnstone, Ivan Hascic, and David Popp. Renewable Energy Policies And Technological Innovation: Evidence Based On Patent Counts. NBER Working Papers 13760, National Bureau of Economic Research, Inc, January 2008.
- [94] D.W. Jorgenson, F.M. Gollop, and B.M. Fraumeni. *Productivity and US Economic Growth*. Cambridge MA, 1987.
- [95] Morton I. Kamien and Nancy L. Schwartz. *Market Structure and Innovation*. Number 9780521293853 in Cambridge Books. Cambridge University Press, August 1968.
- [96] Wolfgang Keller. Technology Flows between Industries: Identification and Productivity Effects. *Economic Systems Research*, 9(2):213–20, June 1997.
- [97] Wolfgang Keller. Are international R&D spillovers trade-related?: Analyzing spillovers among randomly matched trade partners. *European Economic Review*, 42(8):1469–1481, September 1998.
- [98] Wolfgang Keller. Geographic Localization of International Technology Diffusion. *American Economic Review*, 92(1):120–142, March 2002.



- [99] Wolfgang Keller and Stephen R Yeaple. Multinational Enterprises, International Trade and Productivity Growth: Firm-Level Evidence from the US. CEPR Discussion Papers 3805, C.E.P.R. Discussion Papers, February 2003.
- [100] A. Kleinknecht, K. van Montfort, and E. Brouwer. The non rival choice between innovation indicators. *Economics of Innovation and New Technology*, 11(2):109–121, 2002.
- [101] Paul Krugman. *Geography and Trade*, volume 1 of *MIT Press Books*. The MIT Press, January 1992.
- [102] Simon Smith Kuznets. *Economic growth and structure : selected essays*. New York : Norton, 1956.
- [103] J. Langrish, W. Evans, M. Gibbons, and F. Jevon. *Wealth from knowledge: studies of innovation in industry*. 1972.
- [104] Jean O. Lanjouw and Mark Schankerman. Characteristics of Patent Litigation: A Window on Competition. *The RAND Journal of Economics*, 32(1):129–151, 2001.
- [105] Gwanghoon Lee. Direct versus indirect international R&D spillovers. *Information Economics and Policy*, 17(3):334–348, July 2005.
- [106] Gwanghoon Lee. The effectiveness of international knowledge spillover channels. *European Economic Review*, 50(8):2075–2088, November 2006.
- [107] Frank R. Lichtenberg and B. van Pottelsberghe. International R&D spillovers: A comment. *European Economic Review*, 42(8):1483–1491, September 1998.
- [108] Zhiqiang Liu. Foreign direct investment and technology spillovers: Theory and evidence. *Journal of Development Economics*, 85(1-2):176–193, February 2008.
- [109] Bart Los and Bart Verspagen. R&D spillovers and productivity: Evidence from U.S. manufacturing microdata. *Empirical Economics*, 25(1):127–148, 2000.
- [110] James H. Love. Technology sourcing versus technology exploitation: an analysis of US foreign direct investment flows. *Applied Economics*, 35(15):1667–1678, October 2003.

## BIBLIOGRAPHY

---

- [111] Carmen López-Pueyo, Sara Barcenilla-Visús, and Jaime Sanaú. International R&D spillovers and manufacturing productivity: A panel data analysis. *Structural Change and Economic Dynamics*, 19(2):152–172, June 2008.
- [112] Robert E. B. Lucas. On the Mechanics of Economic Development. *Journal of Monetary Economics*, 22:3–42, 1988.
- [113] Olivier Lumenga-Neso, Marcelo Olarreaga, and Maurice Schiff. On ‘indirect’ trade-related R&D spillovers. *European Economic Review*, 49(7):1785–1798, October 2005.
- [114] Megan MacGarvie. The Determinants of International Knowledge Diffusion as Measured by Patent Citations. *Economics Letters*, 87:121–126, 2005.
- [115] Megan MacGarvie. Do Firms Learn from International Trade? *The Review of Economics and Statistics*, 88(1):46–60, May 2006.
- [116] Franco Malerba, Maria Luisa Mancusi, and Fabio Montobbio. Innovation, international R&D Spillovers and the sectoral heterogeneity of knowledge flows. CESPRI Working Papers 204, CESPRI, Centre for Research on Innovation and Internationalisation, Università’ Bocconi, Milano, Italy, October 2007.
- [117] Maria Luisa Mancusi. International spillovers and absorptive capacity: A cross-country cross-sector analysis based on patents and citations. *Journal of International Economics*, 76(2):155–165, December 2008.
- [118] Edwin Mansfield. How Rapidly Does New Industrial Technology Leak Out? *Journal of Industrial Economics*, 34(2):217–23, December 1985.
- [119] Edwin Mansfield and al. Social and Private Rates of Return from Industrial Innovations. *The Quarterly Journal of Economics*, 91(2):221–40, May 1977.
- [120] Alfred Marshall. *Principles of economics*. Macmillan, 1920.
- [121] Per Botolf Maurseth and Bart Verspagen. Knowledge Spillovers in Europe: A Patent Citations Analysis. *Scandinavian Journal of Economics*, 104(4):531–45, December 2002.
- [122] J. Michel and B. Bettels. Patent Citation Analysis; A closer look at the basic input data from patent research reports. *Scientometrics*, 51:181–201, 2001.

- [123] P. Mohnen. R&D Externalities and Productivity Growth. *STI Review*, 18:39–66, 1996.
- [124] Pierre Mohnen and Normand Lepine. R&D, R&D spillovers and payments for technology: Canadian evidence. *Structural Change and Economic Dynamics*, 2(1):213–228, June 1991.
- [125] David Mowery and Nathan Rosenberg. The influence of market demand upon innovation: a critical review of some recent empirical studies. *Research Policy*, 8(2):102–153, April 1979.
- [126] David C. Mowery and Nathan Rosenberg. *Paths of Innovation: Technological Change in 20Th-Century America*. Cambridge University Press, 1989.
- [127] S. Myers and D Marquis. Successful industrial innovations. Technical report, National Science Foundation, Washington, D, 1969.
- [128] Richard G. Newell, Adam B. Jaffe, and Robert N. Stavins. The Induced Innovation Hypothesis And Energy-Saving Technological Change. *The Quarterly Journal of Economics*, 114(3):941–975, August 1999.
- [129] William D Nordhaus. Some Skeptical Thoughts on the Theory of Induced Innovation. *The Quarterly Journal of Economics*, 87(2):208–19, May 1973.
- [130] Giovanna Oddo. Patstat. *IPR helpdesk Bulletin*, 30, Nov-Dec, 2006.
- [131] OECD. *The Knowledge Based Economy*. OECD Publishing, 1996.
- [132] OECD. Compendium of Patents Statistics. Technical report, OECD, Paris, 2008.
- [133] OECD. *OECD Patent Statistics Manual*. OECD Publishing, 2009.
- [134] OECD. *Research and Development Expenditure in Industry 2009 ANBERD*. OECD Publishing, 2009.
- [135] European Patent Office. World Wide Database Coverage Report. Technical report, EPO, 2008.
- [136] Makoto Ohta and Zvi Griliches. Automobile Prices and Quality: Did the Gasoline Price Increases Change Consumer Tastes in the U.S.? *Journal of Business & Economic Statistics*, 4(2):187–98, April 1986.
- [137] Michael J. Orlando. Measuring Spillovers from Industrial R&D: On the Importance of Geographic and Technological Proximity. *RAND Journal of Economics*, 35(4):777–786, Winter 2004.

## BIBLIOGRAPHY

---

- [138] Raffaele Paci and Stefano Usai. Knowledge flows across European regions. Working Paper CRENoS 200704, Centre for North South Economic Research, University of Cagliari and Sassari, Sardinia, 2007.
- [139] Ariel Pakes and Zvi Griliches. Estimating Distributed Lags in Short Panels with an Application to the Specification of Depreciation Patterns and Capital Stock Constructs. *Review of Economic Studies*, 51(2):243–62, April 1984.
- [140] G. Park and Y. Park. On the measurement of patent stock as knowledge indicators. *Technological Forecasting and Social Change*, 73:783–812, 2006.
- [141] Giovanni Peri. Knowledge Flows and Knowledge Externalities. CESifo Working Paper Series CESifo Working Paper No., CESifo GmbH, 2002.
- [142] Giovanni Peri. Determinants of Knowledge Flows and Their Effect on Innovation. *The Review of Economics and Statistics*, 87(2):308–322, 06 2005.
- [143] W.L. Peterson. Return to poultry research in the united states. *Journal of farm economics*, 49:656–669, 1967.
- [144] H Pillu and G. Koleda. Induced Innovation and International Technological Opportunity in the Field of Energy : Evidence from World Patent Citations. *ENERDAY 3rd Conference on Energy Economics and Technology*, 2008.
- [145] David Popp. Induced Innovation and Energy Prices. NBER Working Papers 8284, National Bureau of Economic Research, Inc, May 2001.
- [146] David Popp. Induced Innovation and Energy Prices. *American Economic Review*, 160-180:92, 2002.
- [147] F Raines. The Impact of Applied Research and Development on Productivity. *Working Paper, Washington University, St. Louis, MO.*, 6814, 1968.
- [148] James Rollinson and Davide Lingua. DATA CATALOG for the EPO Worldwide Patent Statistical Database. Technical report, European Patent Office, 2008.
- [149] Paul M Romer. Increasing Returns and Long-run Growth. *Journal of Political Economy*, 94(5):1002–37, October 1986.

- [150] Paul M Romer. Endogenous Technological Change. *Journal of Political Economy*, 98(5):S71–102, October 1990.
- [151] Andrew K. Rose and T. D. Stanley. A Meta-Analysis of the Effect of Common Currencies on International Trade. *Journal of Economic Surveys*, 19(3):347–365, 07 2005.
- [152] Nathan Rosenberg. *Inside the black box*. Cambridge University Press, 1982.
- [153] R. W. Rycroft and D. E. Kash. *The Complexity Challenge. Technological Innovation for the 21st Century*. London, Pinter, 1999.
- [154] W.E.G Salter. *Productivity and technical change*. Cambridge University Press, 1960.
- [155] Mark Schankerman and Ariel Pakes. Estimates of the Value of Patent Rights in European Countries during the Post-1950 Period. *Economic Journal*, 96(384):1052–76, December 1986.
- [156] F.M. Scherer. Inter-Industry Technology Flows and Productivity Growth. *The Review of Economics and Statistics*, 64(4):627–34, November 1982.
- [157] F.M. Scherer. *R&D, Patents and Productivity*, chapter Using Linked Patent and R&D Data to Measure Interindustry Technology Flows. University of Chicago Press, 1984.
- [158] F.M. Scherer. Technology flows matrix estimation revisited. Working Papers 02-18, Federal Reserve Bank of Philadelphia, 2002.
- [159] Thomas Scherngell, Manfred M. Fischer, and Martin Reismann. Total factor productivity effects of interregional knowledge spillovers in manufacturing industries across Europe. *Romanian Journal of Regional Science*, 1(1):1–16, December 2007.
- [160] Andrew Schmitz and David Seckler. Mechanized Agriculture and Social Welfare: The Case of the Tomato Harvester. *American Journal of Agricultural Economics*, 52(4):569–578, 1970.
- [161] U. Schmoch, F. Laville, P. Patel, and R. Frietsch. Linking Technology Areas to Industrial Sectors. *Final Report to the European Commission, DG Research*, 2003.
- [162] J. Schmookler. Economic sources of inventive activity. *The Journal of Economic History*, 2(1)19, 1962.
- [163] J Schmookler. *Invention and Economic Growth*. Harvard University Press, Cambridge, 1966.

## BIBLIOGRAPHY

---

- [164] T W Schultz. *The economic organization of agriculture*. McGraw-Hill, 1954.
- [165] Adam Smith. *An Inquiry into the Nature and Causes of the Wealth of Nations*. NY (1937), 1776.
- [166] Robert M. Solow. A Contribution to the Theory of Growth. In *Growth economics. Selected Readings*. Sen, A. pp. 161-192. Penguin, Harmondsworth, 1956.
- [167] Robert M. Solow. Technical Change and the Aggregate Production Function. *The Review of Economics and Statistics*, 39(3):312–320, 1957.
- [168] T. D. Stanley. Wheat from Chaff: Meta-analysis as Quantitative Literature Review. *Journal of Economic Perspectives*, 15(3):131–150, Summer 2001.
- [169] T. D. Stanley. Beyond Publication Bias. *Journal of Economic Surveys*, 19(3):309–345, 07 2005.
- [170] T. D. Stanley and Stephen B. Jarrell. Meta-Regression Analysis: A Quantitative Method of Literature Surveys. *Journal of Economic Surveys*, 3(2):161–70, 1989.
- [171] T.D. Stanley, Chris Doucouliagos, and Stephen B. Jarrell. Meta-regression analysis as the socio-economics of economics research. *The Journal of Socio-Economics*, 37(1):276–292, February 2008.
- [172] Alessandro Sterlacchini. R&D, Innovations, and Total Factor Productivity Growth in British Manufacturing. *Applied Economics*, 21(11):1549–62, November 1989.
- [173] Alex J. Sutton, Keith R. Abrams, David R. Jones, Trevor A. Sheldon, and Fujian Song. Methods for Meta-analysis in Medical Research. *Institute of Medical Biometry and Medical Informatics*, 22:3112 – 3114, 2000.
- [174] N. Terleckyj. Effects of R&D on the Productivity Growth of Industries: An Exploratory Study. Technical report, National Planning Association, Washington, DC., 1974.
- [175] Peter Thompson. Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-added Citations. *The Review of Economics and Statistics*, 88(2):383–388, 06 2006.
- [176] Peter Thompson and Melanie Fox-Kean. Patent Citations and the Geography of Knowledge Spillovers: A Reassessment. *American Economic Review*, 95(1):450–460, March 2005.

- [177] Marcel P. Timmer, Mary O'Mahony, and Bart van Ark. EU KLEMS Growth and Productivity Accounts: An Overview. *International Productivity Monitor*, 14:71–85, Spring 2007.
- [178] Manuel Trajtenberg. A Penny for Your Quotes: Patent Citations and the Value of Innovations. *RAND Journal of Economics*, 21(1):172–187, Spring 1990.
- [179] Bulent Unel. R&D spillovers through trade in a panel of OECD industries. *Journal of International Trade & Economic Development*, 17(1):105–133, 2008.
- [180] James M. Utterback. *Mastering the dynamics of innovation*. Harvard Business School Press, Boston, MA., 1996.
- [181] Hans Van Meijl. *Endogenous Technological Change: The Case of Information Technology*. PhD thesis, MERIT, Maastricht University, Maastricht, 1995.
- [182] Hans Van Meijl. Measuring Intersectoral Spillovers: French Evidence. *Economic Systems Research*, 9(1):25–46, March 1997.
- [183] Bruno van Pottelsberghe de la Potterie. Issues in Assessing the Effect of Interindustry R&D Spillovers. *Economic Systems Research*, 9(4):331–56, December 1997.
- [184] Bart Verspagen. Estimating international technology spillovers using technology flow matrices. *Review of World Economics (Weltwirtschaftliches Archiv)*, 133(2):226–248, 06 1997.
- [185] Bart Verspagen. Measuring Intersectoral Technology Spillovers: Estimates from the European and US Patent Office Databases. *Economic Systems Research*, 9(1):47–65, March 1997.
- [186] Bart Verspagen, Ton van Moergastel, and Maureen Slabbers. MERIT concordance table: IPC - ISIC (rev.2). Technical report, MERIT Research Memoranda 1994-004, 1994.
- [187] David Warsh. *Knowledge and the Wealth of Nations: A Story of Economic Discovery*. New York : W, 2006.
- [188] Colin Webb, H el ene Dernis, Dietmar Harhoff, and Karin Hoisl. Analysing European and International Patent Citations: A Set of EPO Patent Database Building Blocks. OECD Science, Technology and Industry Working Papers 2005/9, OECD, Directorate for Science, Technology and Industry, October 2005.

## BIBLIOGRAPHY

---

- [189] Robert Wieser. Research And Development Productivity And Spillovers: Empirical Evidence At The Firm Level. *Journal of Economic Surveys*, 19(4):587–621, 09 2005.
- [190] James A Wilcox. Automobile Fuel Efficiency: Measurement and Explanation. *Economic Inquiry*, 22(3):375–85, July 1984.
- [191] Bin Xu and Jianmao Wang. Capital Goods Trade and R&D Spillovers in the OECD. *Canadian Journal of Economics*, 32(5):1258–1274, November 1999.







## **Résumé :**

Dans cette thèse, nous analysons les différents aspects des externalités de connaissance et la façon dont les citations de brevet peuvent être utilisées comme un indicateur de ces flux. La première partie de cette thèse examine la littérature traditionnelle sur les externalités de connaissance, et cela d'un point de vue à la fois qualitatif et quantitatif (la forme quantitative est réalisée grâce à une méta-analyse). Nous insistons sur les conséquences résultant de l'utilisation de différents canaux de mesure de ces externalités, précisément nous nous attardons sur les hypothèses sous-jacentes et sur leurs implications en termes d'estimations empiriques. Ce point est important car ces canaux sont la principale source d'hétérogénéité des résultats empiriques. Dans la seconde partie, nous explorons des données de brevets et de citations de brevet encore peu étudiées (ces données sont extraites de la base de données Patstat pour les offices de brevets du G5, de l'OEB et de l'OMPI). Cette analyse est à nouveau réalisée à la fois en termes qualitatifs et quantitatifs. La troisième partie, dans un premier temps, examine de façon empirique les caractéristiques des flux de connaissance entre et au sein des inventeurs des pays du G5 et cela pour 13 secteurs industriels. Dans un deuxième temps, cette partie propose et valide la création d'un indicateur de stocks de connaissance qui prend en compte les externalités de connaissance internationales. Cet indicateur se révèle particulièrement utile puisque les indicateurs traditionnels ne sont pas toujours disponibles (comme les indicateurs basés sur les dépenses de R&D). Enfin, l'indicateur précédemment créé sera appliqué à une étude de cas consacrée à l'analyse des déterminants de l'innovation pour les technologies énergétiques efficaces.

## **Mot clés :**

Brevets; Citation de brevets; Base de donnée sur les brevets; Externalité de connaissance; Diffusion de connaissance; Innovation induite; Technologie énergétiques

## **Abstract:**

In this dissertation, we analyze the different aspects of knowledge spillovers with special emphasis on the use of patent citations as an indicator of knowledge flows. The first part of the thesis reviews the traditional knowledge spillovers literature both with a qualitative and a quantitative view (the quantitative form is done through a meta-analysis). A particular focus is placed on the consequences of using the different channels reflecting different weighted functions, i.e. about the underlying assumptions and the implications in terms of estimates. This point is important since these channels are the main source of heterogeneity in the analysis of knowledge spillovers. In the second part, we shed some light on patent and patent citation data that have not been much studied (taken from the Patstat database for the G5 patent offices plus the EPO and the WIPO). Once again this analysis is conducted both in qualitative and quantitative terms. The third part, in a first time, investigates empirically the diffusion pattern of knowledge between and within the G5 inventors for 13 manufacturing sectors. In a second time, it proposes and validates the creation of an input-based indicator of knowledge that takes into account international knowledge spillovers. This indicator can be very helpful since traditional indicators are not always available (such as R&D indicators). Finally, the previously created indicator will be applied in a case study dedicated to analyze the determinants of innovation for energy efficient technologies.

## **Keywords:**

Patents; Patent citations; Patent database; Knowledge spillovers; Knowledge diffusion; Induce innovation; Energy technologies.