Exploring the Importance of Innovations with Patent Citations

By

Uryia Dolev

Thesis submitted in partial fulfilment of the requirements for the Degree of Master of Commerce in Economics

at

Department of Economics
College of Business and Economics

The University of Canterbury

2006

Acknowledgements

First and foremost, I would like to sincerely thank my supervisor, Prof. Les Oxley, for his ongoing support, guidance and encouragement throughout this year. Especially, for allowing me to study a much more narrowly focussed topic than was initially intended.

I would also like to express my heartfelt gratitude and thanks to those individuals who were instrumental in helping me to complete this project:
Dr. Marco Reale for introducing me to the world of non-parametric statistics, Dr. Philip Meguire for numerous discussions, some of which have been of signal and some of noise, but definitely interesting and stimulating, Dr. Søren Højsgaard for assistance with getting the MIM program to work with such a large dataset, Prof. Manuel Trajtenberg for information on his estimation procedure, Dr. (*forthcoming*) Paul Walker for making his private library a public good, and the entire Department of Economics at the University of Canterbury for many great educating years.

Parts of this thesis were presented at the New Zealand Econometrics Study Group meeting and at the New Zealand Treasury during 2006. This thesis was made possible due to the financial assistance of the Royal Society of New Zealand, Marsden Grant, 04-UOC-020-SOC, *Winners and Losers in the Knowledge Society,* and the New Zealand Treasury.

I am responsible for any remaining errors.

Dedication

To Dr. Yael Frank, a wonderful grandma and a great source of wisdom. *Toda Al Hakol*

Abstract

This thesis begins by outlining the theoretical and empirical foundations of the economics of innovations. It then proceeds by analysing four econometric issues in the measurement of technological knowledge embedded in patented innovations and modelling the statistical relationship of the value of patented innovations originating in the G-5 countries overtime. This thesis contributes to the economics of innovation literature in four areas: (1) a comprehensive review of the proxies available to elicit the value embodied in patented inventions (2) a direct comparison of regression estimates based on citations count dependent variable versus citations-weighted dependent variable (3) an introduction and application of Regression Tree and Graphical Modelling methodologies to model patented inventions (4) estimation of the fluctuations and associations in the values of patented innovation in the G-5 countries using patent citations.

There is a well known saying that 'Consistency is the Hobgoblin of small minds'. If this is true, then economics has been well endowed with minds that are not small."

Joseph E. Stiglitz, Nobel Laureate in Economic 2001

Table of Contents

List of Figures

List of Tables

1 Introduction

Economists have increasingly recognised the importance of technological knowledge assets in the economic growth process and the social well-being of all. The modern economic inquiry into technological knowledge stems from a number of theoretical developments led by Romer (1986, 1990), Lucas (1988), Aghion and Howitt (1988) and Grossman and Helpman (1991). Ideas, inventions, research and scientific discoveries are at the heart of modern growth theory. The difficulty comes in capturing these dynamic processes empirically, in a systematic and consistent manner. However, "in this desert of data, patent statistics loom up as a mirage of wonderful plentitude and objectivity" (Griliches, 1990 p. 1661).

Patents fascinate economists as they represent an excellent source of information regarding innovative activity, technological developments and intellectual property. Most applied work however overlooks the fact that the value of patents is highly skewed to the right. Very few patents have a significant technological and economic impact on the society, while very many patents have a limited and insignificant impact. Recently, upon investigation, researchers have found that patent citations provide a strong indicator of the ex-post technological and economical value patents represent. Patent citations are the 'prior-to-new-art-link' and appear on the patent grant document.

Patent citations are a complicated system that requires a careful analysis. The objective of this thesis is to improve the estimates of the value of innovations using only information contained in patent citations. This thesis provides an original contribution to the economics of innovation literature in four areas. First, it provides a comprehensive and a rich review of the elicitation of the ex-post value of patented innovations literature. Secondly, it contrasts regression estimates based on citation count dependent variable against citation-weighted dependent variable. Thirdly, it takes advantage of techniques rare to the field, such as Regression Tree and Graphical Modelling, to model patented inventions. Fourthly, it estimates the fluctuations and associations of the averaged citation received by patented inventions originating in the G-5 countries: the US, the UK, Japan, Germany and France.

The remainder of the thesis is organised as follows:

Chapter 2 provides a rich and detailed synthesis of the main theoretical approaches to the economics of innovation. It describes the reasons technology, innovative activity and scientific discoveries are found at the heart of the modern economic inquiry into the determinants of economic growth. The discussion particularly concentrates on models based on the neoclassical and endogenous growth schools of thought.

Chapter 3 starts with a description of the meaning of innovation. It is followed by a discussion on the measurement of innovations and the reasons patent statistics are found at the centre stage in estimating innovative activity. The following sections are devoted to the empirical literature on patent data and patent values. The chapter concludes with a detailed description of the various proxies available to elicit the value embodied in patented innovations, and encourages the use of patent citations.

Chapter 4 provides a comprehensive discussion about the conceptual, theoretical and empirical background behind the use of patent citations for measuring the ex-post technological and economic value embodied patented inventions. This review sets the scene and provides a context for the analysis of the quality of patents innovations in Chapters 6, 7 and 8.

Chapter 5 provides a description of the citation data that is used in Chapters 6, 7 and 8. The data is taken from the National Bureau of Economics Research U.S. patent citations program. It comprises three million U.S. patents granted between January 1963 and December 1999.

Chapter 6 is the starting point for the analysis of the value of patented innovations. To ensure full comprehension of the methodology and use of the patent citation data, the chapter replicates an original econometric estimation of the value of patented innovations approximated by patent citations, suggested by Trajtenberg (2001). The results set the scene for the analysis in Chapter 7.

Chapter 7 builds on Trajtenberg's analysis and provides rigorous econometric estimation of the quality of patented innovations in the context of Count Data models and atheoretical Regression Tree.

Chapter 8 provides an approach that uses patent citation data in the context of a cointegration test and Graphical Modelling to identify whether long run geographic trends exists in the value of patented innovations. The analysis is based on the theoretical framework developed in Chapter 2, 3 and 4. The final chapter draws together the main findings of this thesis.

2 Why Do Economists Care About Innovations?

This chapter provides the main theoretical foundations of the economics of innovations. It discusses the reasons technology, inventions and scientific discoveries are found at the heart of the modern economic inquiry into the determinants of economic growth and the theoretical underpinning of this thesis.

2.1 *Introduction*

The interest of economists in the economics of innovations stems from an answer to a question that has preoccupied economists since the days of Adam Smith: What drives the long-run growth of nations? The answer is "The engine of economic growth is invention" (Jones 2002, p.195).[1]

Economists have always considered innovative activity and scientific discoveries to be the key assets that foster and accelerate economic progress (Machlup, 1962; Freeman 1974). The early analysis goes back to the work of the classical economists. Adam Smith, 250 years ago, said, "man educated at the expense of much labour and time…may be compared to one of those expensive machines," emphasising the importance of knowledge and ideas embodied in individuals. Alfred Marshall was more explicit and described knowledge as "our most powerful engine of production (Marshall 1980). Karl Marx's analysis of capitalist economy in the nineteenth century showed that the process of technological change is the driver of "capitalist development" (Rosenberg 1971). The process of technological development was also

[1] In introducing Growth Theory and the Economics of Technological Change and Innovation, I closely follow texts by Aghion and Howitt (1998), Helpman (2004), Jones (2005), Marsh (2004) and Rosenberg (1971).

evident in the work of Malthus (1980) and Ricardo (1921), although it played a "kind of [an] afterthought" in relation to the traditional economic assets of land and capital formation (Rosenberg 1971).

The classical economists emphasised inventions, but did not model the economics behind them. Their analysis was static in nature, which limited their ability to model the dynamic process of inventive activity. The first to bring technical innovation into the centre of the economic analysis was Joseph Schumpeter in the early 1940s. Schumpeter suggested that firms compete through innovation, whereby the development of new technologies generates a "creative destruction". These innovations, carried out by profit seeking entrepreneurs, rival and may destroy the existing structure of industries, leading to the dynamic transformation of economic systems. In Schumpeter's vision, these dynamic waves were the causes and forces behind the short run and long run economic fluctuations.

Schumpeter's theory was limited in scope, as it had "little to say [about the] economic factors shaping inventive activity, which seemed to have a life of its own" (Rosenberg 1971, p.9). The foundations and inspiration for the most significant progress towards an analytical examination of invention growth in modern economics is contributed to the two famous articles by Robert Solow in the mid 50s. Solow's growth framework emphasised the accumulation of physical capital, such as factories and machines, and human capital, such as education and skills, as the forces that move economies towards an equilibrium path of growth. Economic incentives were assumed to influence the accumulation of these resources, which made this framework very convenient for later analysis and gained it superiority among the leading economists at the time such as

Moses Abramovitz, Kenneth Arrow, Simon Kuznets and others. Much of their work is now known as the neoclassical growth theory.

2.2 *Neo-Classical Growth Theory*

The neo-classical growth theory provided answers to a number of key questions. Firstly, rich economies are those that have high capital investment rates and low net population growth. Their capital per worker is high and leads to a more productive and skilful labour force.

Secondly, a country's growth rate is, *inter alia*, a function of capital intensity. As a country becomes richer, its capital to labour ratio intensifies and the return to capital diminishes over time. Consequently, its growth rate will decline. This is a key implication of the neoclassical growth. The theory hypothesises that there is an inverse relationship between capital intensity and growth rate, a relationship commonly referred to as the convergence hypothesis. This means that there would be a point in a country's growth rate where further increase in capital per worker will stimulate no further growth, a property known as the steady state.

Thirdly, the reason that some countries enjoy sustained long-run growth is technology. A direct consequence of the convergence property is that a country's growth rates will eventually decay due to the diminishing returns to capital. In the long run, the only factor that could offset this tendency is technological progress, where technological progress is the ability to produce more or better output from the same amount of input. The change of technology available is therefore the transformation in the processes by which economies produce outputs, a process driven by ideas, scientific discoveries and their diffusion throughout.

The neoclassical model provided a simple framework for the estimation of a country's technological progress. The methodology, commonly referred to as 'growth accounting' as proposed by Solow (1957), decomposes the change in the scale of production into the factors of production components (Capital, Labour). The growth of output that is not explained by the growth of factors of production is called the total factors productivity (TFP). Under this structure, the TFP measures the technological progress that occurred in an economy during a period, and the ratio of TFP to GDP is the growth of wealth explained by technological progress. The overwhelming evidence (see Helmpan 2004 for survey) shows that among the industrialist countries, the TFP ratio, over the second half of last century, ranges from 20 to 50 percent depending on the country and the quality adjustments of inputs.

The assumptions and predictions of neoclassical theory raised a number of concerns. Although the convergence of the growth rate to the rate of technology progress is the theory's greatest prediction, it cannot explain how technology enters into the economic system as technology is considered to be exogenous. The TFP measure provided an estimation of the rate of growth of technological progress but could not explain what causes it to grow (Helpman 2004).[2] Meanwhile, the growth rates of many countries has been rising rapidly in a non-converging manner, rather than falling as predicted by the theory, challenging its convergence hypothesis (Jones 2005). Furthermore, the theory assumed that markets operated in a perfectly competitive environment, which would have prevented firms from covering the costs of their innovative activities, due to the public-good nature of knowledge (Marsh 2004). This implied that research and development, the inputs of technology progress, would not have been profitable and would have not been undertaken by firms.

[2]Interested readers can refer to Chapter 5 in Lipsey et al. (2005) for a discussion on why changes in TFP may not track changes in technology.

2.3 *New Growth Theory*

Prior to the late eighties, research on technological development, its flow into the economic system and consequential growth, were subordinate to business cycle research, which took research centre stage. Technology was seen as a 'black box', which was "outside the specialised competence of most economists" (Freeman 1994). The revolution came with the formalisation of the effects of the accumulation of technological ideas on the economic system by Paul Romer and Robert Lucas in the late eighties. Their theories endogeneised the deliberate process economic agents undergo, in which the invention and diffusion of new processes is explained within the model. In that sense, these models extended the Solow framework by explaining the development of inventions. According to Romer (1986), technological ideas are nonrivalrous, implying that use of the idea by one firm does not preclude other firms from benefiting from the idea simultaneously. The production of new ideas increases the aggregate stock of ideas available in the economy, and that stock is a function of, *inter alia,* the number of researchers. The Japanese automobile inventory system of just-in-time service cannot stop American and European car manufactures from utilizing it. Once the idea is produced, it can be reproduced with no extra cost (Jones, 2005). Consequently, firms' productivity increases and increasing returns to scale would characterise production due to the high fixed costs. The presence of increasing returns to scale of production implies that inventors must expect to price above marginal cost in order to cover the high fixed of cost of producing the idea, which necessitated a move towards an economy that competes imperfectly, otherwise firms would never cover their costs and would not engage in research and development (Jones, 2005).

Romer's model has two important implications. First, increasing returns to production imply that growth rates need not be declining, but can instead be increasing. This stands in sharp contrast to the notion of convergence as implied in the neoclassical theory (see, for example, Baumol (1986) and Mankiw et al. (1992) for empirical examinations of the implications of the two viewpoints). Secondly, the nonrivalry and low excludability attributes of ideas, originally suggested by Arrow in the early sixties, imply a positive externality in knowledge production and spillover – a "standing on shoulders" effect (Jones, 2005).

These early models by Lucas (1988) and Romer (1990) are now viewed as the foundations of the 'new' growth theory. Since then, an overwhelming number of models have been developed in this spirit. In 1990, Romer also provided the second milestone in endogenising technology. Instead of pursuing an aggregate ideas accumulation approach as he did in 1986, Romer suggested a disaggregated model of the business sector that provides an explicit analysis of the competitive behaviour of firms (Helpman, 2004). In this model, firms innovate by engaging in R&D and are driven by market incentives. Free-riding on ideas is protected with patents, which restore the incentive to innovate. However, some of the ideas still spillover towards a common knowledge pool, which reduces the production costs for all firms, but at the same time induces development of competitive products, which cuts down the profits to all firms. The novelty in Romer's model is a knowledge spillover mechanism that maintains a constant innovation incentive, which provides a nondiminishing rate of growth.

To illustrate this explicit link between innovation and economic growth, consider the following model suggested by Romer (1990) and generalised by Jones (1995,2005) which explains why and how long-run growth is sustained.

2.3.1 Romer (1990)

Consider an economy with a standard Cobb-Douglas production function:

$$Y = K^{1-\alpha}(AL_y)^{\alpha}$$

where

Y is the output produced in the economy

A is the available stock of technology

K is Capital

L is labour. The population in this economy engages either in output production (L_y) or in innovative activities (L_A).

α is a parameter between 0 and 1.

For a given level of technology, the production function in this economy exhibits constant returns to scale. However, when technological knowledge becomes part of the production process, increasing returns to scale characterise the production function.

> Once Steve Jobs and Steve Wozniak invented the plans for assembling personal computers, those plans…did not need to be invented again. To double the production of personal computers, Jobs and Wozniak needed only to double the number of intergraded circuits, semiconductors, etc., and find a larger garage. (Jones 2005, p. 98)

The capital in this economy accumulates according an exogenously determined saving rate, s_K and depreciation rate d: $\dot{K} = s_K Y - dK$.

The link between invention and growth becomes explicit once we model the flow of innovation into the model. Romer suggests that $A(t)$ represents the existing technology, which is the summation of all existing technological innovations at time t. The production of innovation, $\dot{A}$, is equal to the number of people engaged in innovative activities, such as Research and Development, multiplied by the rate at which they successfully develop new inventions, $\bar{\delta}$: $\dot{A} = \bar{\delta} L_A$. The innovation success rate depends on the available stock of technology: $\bar{\delta} = \delta A^{\phi}$, where δ and ϕ are constants. The parameter ϕ indicates the degree of externality inherent in technological knowledge. If it is greater than zero, it implies that an increase in the stock of technological innovations increases the invention success rate whereas if it is less than one it implies the opposite.[3] The invention equation then becomes $\dot{A} = \delta L_A A^{\phi}$ which implies that the increase in TFP is proportional to the labour units engaged in innovative activities and the existing stock of technology. By dividing the equation by the technology stock, the innovation growth rate is denoted $\frac{\dot{A}}{A} = \delta \frac{L_A}{A^{1-\phi}}$. The steady state of inventive creation is found by logging and differentiating the equation with respect to time: $0 = \frac{\dot{L}_A}{L_A} - (1-\phi)\frac{\dot{A}}{A}$ which reduces to $\frac{\dot{L}_A / L_A}{(1-\phi)}$ and "pins down…all the growth rate in this model" (Jones 1995, p. 767). Thus, the long run growth in output per worker in this model is tied to the growth of inventions and to the nature of the innovation externality.

Alternative ways of endogenising technology have been suggested in subsequent work. Later models can be crudely divided into either R&D or Human Capital based models

[3] For a detailed discussion on the sign of ϕ, see Jones (1995).

(Klenow, 1998). In R&D based models such as these of Grossman and Helpman (1991) and Aghion and Howitt (1992), R&D efforts generate new ideas, which once embodied in intermediate goods raise the level of productivity and generate growth. In these models, distinctively from Romer (1990), the productivity increases along a quality ladder, where the product's quality affects its substitutability among older products. The higher the quality of a product, the bigger the "creative destruction" imposed by the new product is. Such models are referred to as Schumpeterian growth models as they fulfil the original prediction of Schumpeter in the late 30s. The second class of models emphasises the accumulation of Human Capital embodied in workers as the productivity factor that stimulates long run growth. Such models include these of Jones and Manuelli (1990) and Rebelo (1991).

To illustrate the link between innovative activities, Human Capital and economic growth, consider the following model suggested by Jones and Manuelli (2005), developed in the spirit of recent work by Boldrin and Levine (2002).

2.3.2 Jones and Manuelli (2005)

c_t - is the final consumption at time t.

A - is a productivity factor.

δ - is the depreciation rate.

β - is a discount factor.

λ - is long run growth

α - is a parameter between 0 and 1.

L - is labour supply. There are two types of labour supply, inventors, L_1, and workers L_2, where $L = L_1 + L_2$ is the total supply of labour within each household. A continuum of households exists. Each supplies labour into invention creation and into production of output. Each of the individuals in the household has his own level of human capital.

H_t - is the level of Human Capital per inventor at time t. If households are symmetric, then H_t represents the economy absolute Human Capital skills frontier at time t.

h_t - is average level of Human Capital per worker is at time t

Inventors can devote their efforts into either the research and development of inventions L_{1t}^H or the training and educating of the workers L_{1t}^h. Thus, $L_1 = L_{1t}^H + L_{1t}^h$

Workers can devote their efforts into either the training and learning from inventors, L_{2t}^h or into to the production of current consumption goods L_2^c. Thus, $L_2 = L_{2t}^h + L_{2t}^c$

The increase in the inventors' Human Capital stock is a function of the existing H_t and the production of new inventions $H_{t+1} = H_t + A_H L_{1t}^H H_t$. The development of new inventions is determined $A_H L_{1t}^H H_t$ by the inventors' inventive efforts, the current stock of inventive Human Capital and an inventive productivity factor. Note, depreciation does not enter this production function. Thus, inventors' Human Capital cannot go backwards.

The increase in the average Human Capital per worker is determined by (after depreciation) per worker Human Capital and the education and training workers undergo at time t $h_{t+1} = (1-\delta_t)h_t + A_h (L_{1t}^h H_t)^{\alpha} (L_{2t}^h h_t)^{1-\alpha}$. The education and training component $A_h (L_{1t}^h H_t)^{\alpha} (L_{2t}^h h_t)^{1-\alpha}$ is a Cobb-Douglas production function. An increase in the skills and knowledge embodied in workers is a function of two factors. The first factor $L_{1t}^h H_t$ is the effort inventors (trainers) put into for training and educating the workers and their Human Capital. The second factor $L_{2t}^{h} h_t$ is effort workers (trainees) put into training and learning from the inventors and their average Human Capital skills.

The more time inventors spend on training and educating the workers, the less they engage in inventing activities. The more time workers spend on training and learning, the less time they have to for production. This effect of Human Capital on inventive output and consequential growth becomes obvious it is by introduced into the production of current consumption goods $c_t = A_c L_{2t}^{c} h_t$. The amount of output produced is determined by the workers production efforts, their average Human Capital skills and a production productivity factor. Therefore, although education and training increases the productivity of workers, they constrain the growth of the inventive Human Capital frontier. In the extreme case where the inventors devote all their efforts towards training (L_{1t}^{H} =0), the inventive Human Capital frontier would remain static and thus, the average Human Capital per worker would be bounded. Subsequently, the production of current consumption goods would also be bounded. Therefore, economic growth is only possible if new inventions are produced. The long run economic growth in a steady state becomes $\beta[1 + A_H L_1^H]$. Thus, if we compare countries with a similar discount factor and productivity factor, the countries that devote more labour into research and inventive activities would have a more skilful labour market and enjoy higher economic growth rates.

2.4 *Alternative Theoretical Approaches*

2.4.1 General Purposes Technologies

The growth theories discussed above view inventiveness as an incremental process (Helpman 2004). However, in 1995, Breshahan and Trajtenberg suggested that certain innovations are radical in nature, which could lead the transformation of industries and economies over time. Steam engines, the railroad, electricity and computers are examples of innovations that gradually penetrated into the economy. These innovations are called General-Purpose Technologies (GPTs). Once the economy adjusts and implements the implications of the new technologies, an accelerated productivity growth rate will spread, leading to an economic discontinuity (Helpman 1998). A growing number of growth students now theorise the economic implications of such innovations. Nevertheless, the study of GPT is still very young and the concept is interpreted in a variety of ways in the literature (See Helpman (1998) for details). Yet, the theory has opened a window to the study of radical breakthroughs in science as the powerful engine of growth in modern economies.

2.4.2 Evolutionary and Systems of Innovations

As growth theory evolved over time with inventive activity gaining the central stage of the analytical analysis, supplementary theories started to evolve. An evolutionary approach that uses biological analogies for the dynamic process of technological change started to evolve in 1982 in the work of Nelson and Winter. This approach advocates the notion of bounded rationality and asymmetric and costly information to explain the inventive decisions undertaken by individuals and firms and institutions. The production of ideas, the centre of this framework, then follows a stochastic process

(Mokyr, 1999). This idea, to an extent, was further developed by a number of scholars who suggested the notion of a National System of Innovations (NIS), where the inventiveness competency of a country is hypothesized to lie within the interactions of individuals, firms and governments either at the domestic or international level (Lundvall 1992). It is an interaction between "institutions in the public and private sectors, whose activities … initiate, import, modify and diffuse new technologies" (Freeman 1987).

Although these alternative theories may provide a more pragmatic description of reality than is provided in the mainstream literature, they are very general and their departure from the neoclassical paradigm of maximization and equilibrium leads to a "propensity to produce sheer nonsense" (Paul Krugman speaking in front of the *European Association for Evolutionary Political Economy* in November 1996). The aggregated approach under NIS creates measurement difficulties, and the implication of the indeterminacy of knowledge creation implied by evolutionary economics does not simplify the estimation (Marsh 2004). Nonetheless, the implications of both theories, integrated within mainstream literature, could provide a more holistic description of the innovative process.

2.5 *Summary*

The implications of growth theory that inventions are the engine behind economic growth have led a research trajectory into the determinants of innovation. Chapter 3 reviews the literature on the measurement of innovations.

3 How Do Economists Measure Innovations?

Chapter 2 provided the theoretical foundations underlining the interest in innovations. This chapter discusses the measurement of innovations and motivates the use of patent data as an indicator of inventive activity. The chapter then describes the empirical literature on patent data, the heterogeneity problem inherent in patent values and the methods available to elicit the value of each patented innovations. It shows that patent citations, prior-to-new-art link, appear to be the most objective and systemic proxy for patent value.

3.1 *What is Innovation?*

Before discussing the measurement of innovations, it will be useful to provide an introduction to the meaning of innovations in this thesis. Economists use innovation to mean "the economic application of a new idea" (Black 2002), that is, new ideas on ways "inputs to the production process [could be] transformed into output" (Jones 2002, p.79). Innovations are predominantly thought to represent technological knowledge, which is knowledge "transmitted [inter alia] by mathematical theorems or computer programs that can be reproduced through known procedures" (Howitt, 1998, p. 99). Economists therefore regard innovations as actions that yield "new products … and new devices to be used in economic production… [and exclude] social inventions, new methods of inducing human beings to compete and cooperate in social progress…[and] creative work of an esthetic character, in which economic use is not the major aim or test" (Kuznets, 1962, p.19)

3.2 *How to Measure Innovations?*

The measurement of innovation is constrained by a shortage of adequate and precise data. Approximations of inventiveness are therefore used in the literature. Although many proxies exist, such as the study of scientific publications (see Kleinknecht 1996), the subsequent performance of inventive firms (see Hansen 1992) and quality indexes of improved products, R&D activity and patent counts are the most commonly used indicators of inventive activity.

3.2.1 Research and Development (R&D)

R&D is the "process of knowledge creation, with the knowledge applicative as a production technique, either directly or indirectly. Enhanced knowledge improves the productivity of existing inputs and these productivity gains – taking the form of cost reductions – are the returns to R&D" (Smith 1991, p.2). Therefore, R&D data are believed to represent the inputs to the production of innovations. The advantage of R&D data is that they typically have a dollar sign attached. Therefore, economists frequently use R&D data as an approximation for the share and intensity of resources devoted to developments of inventions. Alternatively, the number of scientists involved in R&D activities is sometimes used.

R&D data are slowly starting to emerge. However, R&D data is often subject to classification problems due to strategic decisions by firms, institutions or countries in classifying their R&D activities (tax advantages for example), which imposes a great constraint on the reliability and validity of R&D based proxies.

3.2.2 Patent Statistics

A patent is a "document, issued by an authorised government agency, granting the right to exclude anyone else from the production or use of a specific new device, apparatus or process for a number of years. The …purpose of the patent system is to encourage invention and technical progress both by providing temporary monopoly for the inventor and by forcing the early disclosure of the information necessary for the production of this item or the operation of the new process" (Griliches 1990, p. 1662-1663).

Whereas R&D measures the input side of innovative production and is subject to classification constraints, patents represent the inventive output and are "based on…an objective and slowly changing standard" (Griliches, 1990 pp. 1661) and are:

1. A voluntary economic system
2. Contain highly detailed information about each invention granted
3. Provide over 250-300 years of data

For these reasons, a substantial body of empirical literature in economics relies on patents data as an indicator of inventive activity. The following section reviews this literature.

3.3 *Patents Statistics in Economics*

Patents have been extensively used in empirical economics since the mid 60s. Basberg (1987), Pavitt (1988) and Griliches (1990) provide surveys of the extensive use of patent statistics in economics.

The early use of patent statistics in economics goes back to the 1966 book by Jacob Schmookler, *Invention and Economic Growth*. In his book, Schmookler "demonstrated that patent statistics…provide a unique source of systematic information about the inventive process" (Jaffe and Trajtenberg, 2002. p. 6). Using patents as a surrogate for an innovation, Schmookler showed that "not only that one could explain the *diffusion* of existing inventions in economic terms … but that one could even explain the pattern of inventive activity itself" (Rosenberg, 1974, p. 90).

However, it was Zvi Griliches in the late 70s who laid the foundations for a systematic and concise use of patent statistics as a defined economic indicator. Griliches "transformed the study of productivity growth from the study of a residual to a study of the measurable factors that caused increases in the output available from given configurations of inputs, and in so doing changed both official statistical procedures, and our understanding of how productivity improvements occur" (Heckman 2006, p.4). During the 1980s, Griliches led an NBER research program into the sources of productivity growth. Griliches and colleagues developed detailed panel data that allowed a through investigation of the relationship between patents, R&D expenditure and productivity at the firm level (Griliches 1984). [4] Their work formed the basis of

[4] Fredric Scherer (1982) was carrying out another large-scale patent related project. Scherer created a detailed patent dataset, with the patent being sorted according to technology type and industry of origin.

future empirical studies on the determinants of growth such as these of Bount et al. (1984), Hall, Griliches and Hausman (1986), Cockburn and Griliches (1988) and many others. Jaffe and Trajtenberg (2002) summarise some of the key findings:

- Strong relationship between R&D and patents at the firm level
- Strong correlation between R&D and patents over time
- R&D expenditure is a strong predictor of firm's performance.

The use of patents statistics as an innovation indicator stimulated new research trajectories into a host of economic questions, such as the factors that influence the decision to innovate (Duguet and Kabla, 2000); the existence of radical innovation (Hall and Trajtenberg, 2005); the effects of Government innovation policies (Henderson et al. 1998; Mowery et al. 2001); the private returns to innovations (Hall 1998); the social returns to innovations (Trajtenberg 1990); the spillovers of ideas (Jaffe 1983; 1986); fluctuations in inventive activities across countries (McAleer et al. 2006); and the diffusion of innovations across time (Sokoloff ,1988; Sokoloff and Khan 1989; Magee 1999).

Although patents are a very objective and concise indicator for innovative activity, they are an imperfect measure of innovation. This is the subject of the next sections.

3.3.1 Limitations of Patent Statistics

A number of problems arise in the use of patents data.

3.3.1.1 The Identification Problem

The first is an identification problem. This is because not all innovations are patented or patentable. Therefore, patents represent only a subset of all the existing inventions. There are a number of reasons an innovation might not be patented.

1. Inventors may strategically decide not to register their innovation through the patent system. Although a patent provides a temporary monopoly on an invention, it forces the inventor to disclose all "the information necessary for the production of this item or the operation of the new process" (Griliches 1990, p. 1663). Therefore, it might be in the best interests of the inventors to use a secrecy approach to protect their innovation. The Coca-Cola formula is an example of such a situation.

2. Innovations might not be patented because they are not a device. Inventors can only patent their invention if they:

> invent or discover any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof,…. The word 'process' is defined by law as a process, act or method, and primarily includes industrial or technical processes. The term 'machine' used in the statute needs no explanation. The term 'manufacture' refers to articles that are made, and includes all manufactured articles. The term 'composition of matter' relates to chemical compositions and may include mixtures of ingredients as well as new chemical compounds. These classes of subject matter taken together include practically everything that

> is made by man and the processes for making the products … A patent cannot be obtained upon a mere idea or suggestion

(US Patent and Trademark Office, 2006).

3. Innovations might not be patentable because they are an idea, although they may provide a 'better' way of doing things. For example, Maxwell's equations of the behaviour of electric fields cannot be patented (Trajtenberg et al. 1997).

4. Inventions may not be patentable if they are regarded as trivial. For example, a marginal improvement of a 'mousetrap' would not be patentable (Trajtenberg et al. 1997) as the invention

> must be sufficiently different from what has been used or described before that it may be said to be nonobvious to a person having ordinary skill in the area of technology related to the invention. For example, the substitution of one colour for another, or changes in size, are ordinarily not patentable.

(US Patent and Trademark Office, 2006).

This identification problem means that patents may not capture the "purely scientific advances devoid of immediate applicability, as well as run-of-the-mill technological improvement that are too trite to pass for discrete, codifiable innovations" (Trajtenberg 2001 p. 336). However, this problem is "widely believed … not too [be] severe" (Trajtenberg et al. 1997, p. 54-55) as the non-patented inventions reflect the outliers in the innovation curve and can be countered by adjusting the measures of this deficiency (Scherer, 1965).

3.3.1.2 The problem of high variance in patent values

The second problem that arises in the use of patent data is the high variance in patent values. It is well recognised that the ex-post value of innovation embodied in each patent varies significantly across patents. Therefore, any aggregation patent counts leads to a highly biased estimation of the underlying innovation activity. This is the topic of the next section.

3.4 *Heterogeneity in Patent Values*

The value of patents is highly skewed to the right. Very few patents have a significant technological and economic impact on the society, while very many patents have a limited and insignificant impact. The patent granting office does not classify or scale the granted patents according to some hypothesised ex-post value measure. The office simply determines whether the patent application meets the non-triviality, novelty and usefulness patent criteria. This implies that any aggregation of patent records would result in a severe bias estimation of the real innovation activity. As Simon Kuznets stressed almost fifty years ago, any systematic measurement of innovations must be sensitive “with respect to the magnitude of technical problem overcome, technical potential, and economic contribution” (Kuznets, 1962. p. 30) of each invention.

The variance in patent values is frequently overlooked in applied work as most researchers simply use patent counts to measure the underlying invention ‘success’, thereby attaching a value of one to all patents. The underlying hypothesis is that the quality of any sampled patent is simply “a random variable with some probability distribution” (Scherer, 1962 p. 1098). Given that the sample size studied is sufficiently large, the average value of the sampled patents will approach the average of the patent population, and the variance of the estimator will be reduced.

The reliance on the law of large numbers to minimise the bias of the estimator does not appear to be valid in the case of patents (Bertran 2004). The findings of numerous studies have suggested that the variance of patents values would remain high despite the size of the patent sample used. The first comprehensive evidence comes from a series of surveys and interviews of US patent holders conducted by the Patent Foundation Study in the late 50s by Barkev Sanders, Joseph Rossman and James Harris. The researchers studied the utilisation of a random sample consisting of two percent of the total patents granted in 1938, 1948 and 1952. One of their most striking findings was that the economic benefits of the patents to their assignees was highly skewed across patents (Sander et al. (1958); for a brief discussion of the results see Schmookler (1966) p. 47-55 and Griliches (1990) p. 1679). Patents that has been utilised at the time the research was carried out had a mean economic value of \$557,000, whereas the median was \$25,000. Griliches (1990) recomputed Sander et al.'s (1958) results for this group of patents and estimated the variance coefficient under log normality to be 2.5 and the standard deviation to be \$1.5 million.

Scherer (1965) used these data to carry a graphical test, confirming that

> …the existence of a Pareto-type distribution of profits with a α coefficient of less than 0.5. Asymptotically such a distribution possesses neither a finite mean nor a finite variance, and so one cannot be sure that the mean economic value of any particular sample of patents converges (under the weak law of large numbers) towards the true population mean value if large enough samples are drawn. …patent statistics are likely to measure run-of-the-mill industrial inventive output much more accurately than they reflect the occasional strategic

> inventions which open up new markets and new technologies. The latter must probably remain the domain of economic historians (Scherer, 1965 p. 1098).

Counting patents leads to two measurement problems (Lanjouw et al. 1998):

1. When systematic differences exist in the value of inventions across different groups of patents, the analysis would lead to biased inferences.
2. Since the relationship between patents counts and values is ambiguous, then even a comparison of patent groups with similar average value is difficult to interpret.

The implications of highly dispersed patent values on measuring inventive activity have led to research into the adjustments of patent counts. This is the topic of the following section.

3.4.1 How to Estimate Patent Values?

The issue of dispersed patent values has led to a new line of research. Its main objective is to identify the value of patents and ways to control for their variability. This section reviews the main approaches in the literature.

1. **Estimation of the value of patents through direct communication with patent holders**

This approach typically relies on surveys, where the patent holders are asked for the monetary value their patent has generated in subsequent years. Thus, the figures elicited are the private economic value of patents to assignees.

This line of work, starting with Sanders et al.'s (1958) and Scherer's (1965) work with US patents, was popularised by Dietmar Harhoff and colleagues in the late 90s and extended to European patents (see Harhoff et al. 1999; Scherer and Harhoff 2000; Harhoff et al. 2003b). Giuri and Mariani (2005) and Reitzig (2003) are a few later examples.

The use of surveys and interviews can be quite useful, but this approach is exposed to a number of important limitations. Firstly, surveys are limited in scope and can be quite expensive, limiting the possibility of using them in large-scale productivity and growth research. Secondly, surveys are prone to bias since patent holders are reluctant to provide their true returns on innovative investments.

2. Observations on the propensity of patent holders to renew patents

Patent holders must pay a periodic fee in order to keep their patent in force. Failure to do so results in the termination of the patent. This renewal cost increases over time in order to keep only useful patents and weed out less valuable ones.

Observations over time of the propensity to renew patents at different patent ages, and the renewal cost schedule, can then provide detailed information on the value of patents: "The lower the incremental fee at which payment is discounted, the smaller is the patent right's estimated value" (Harhoff et al. 2003a, p. 280).

The underlying view of this approach is that the decision whether to renew a patent is based on economic criteria. Thus, patents are renewed if the discounted stream of profits that could be earned in the subsequent period exceeds the cost of renewing the patent (See Lanjouw, Pakes and Putnam, 1998 and Pakes and Simpson, 2001 for a review of empirical work along this line).

Early work on the renewal mechanism was carried out by Dernburg and Gharrity (1961) in the early 60s. However, it was Nordhaus' (1969) thesis in the late 60s that introduced patent renewal data to the discipline. The first to popularise this approach were Pakes and Schankerman in 1984 using European patents data. Pakes and Schankerman's (1984) patent renewal model was deterministic in nature, allowing the stream of returns generated by renewed patents to decay deterministically over time. Pakes (1986) relaxed this assumption and introduced uncertainty in the model. According to Pakes this uncertainty is because inventors often patent their inventions at an early stage of the innovation process in order to obtain immediate protection. This implies that the decision to patent often occurs prior to receiving market feedback about the commercial potential of the invention. However, Pakes' findings showed that the uncertainty gradually fades away and almost perfectly clears when the patents reach the age of five. Given this result, Pakes and Schankerman (1986)[5] re-estimated the deterministic model, but this time for patents older than five. Their results showed that half of the patents reach the age of ten and half do not. Only ten percent of all patents survive the entire renewal period. This implies that the majority of patents are not valuable enough, and the expected discounted profit generated does not exceed their maintenance costs. The mean value of a patent in the UK and France was $7,000, whereas in Germany, where the application is more rigorous, the mean value was $17,000. Their findings confirmed the highly skewed distribution of patents values as half of the values estimated belonged to about five percent of the entire patent population analysed.

[5] The data contained information on the renewals of patents between 1950 and 1979.

The issue concerning the uncertainty of the inventors' learning process led to a division of the literature. The first group (see, among others, Schankerman 1998, Sullivan 1994) assume that all the relevant information is available at the time the renewal is made, whereas the second group (see, among others, Pakes and Simpson 2001, Lanjouw 1993) allows the patent value to follow a stochastic process.

There are a number of limitations to the patent renewal approach. Griliches (1990) suggested that identification problems might arise in the renewal models. Since an 'open-ended' class of patents exists[6] that pay the full renewal fees throughout time and a stable renewal cost schedule, the estimations are very sensitive to the assumptions underlying functional form for patents rights. Furthermore, Scherer[7] cautioned that since the technologies rapidly change, early patent dropout might not be indicative of low value. Many inventions are of high value when first introduced but become obsolete shortly after. Levin[8] indicated that exogenous factors might influence the decision to renew patents, such as institutional factors in the pharmaceutical industries. For example, because of the long regulatory delays between drugs development and their introduction to the market, the high patent renewals might be biased for pharmaceutical and drugs patents (Pakes and Simpson, 1989).

3. Independent proxies that correlate with ex-post value of patents

This approach is typically an econometric analysis of value-dominated variables that are hypothesised to exhibit strong correlations with the value of patents. Most of the variables used in the literature are taken from the grant document which is issued when

[6] The statutory limit is usually between 15 to 20 years.
[7] In the General Discussion in Pakes and Simpson (1989).
[8] In the General Discussion section in Pakes and Simpson (1989).

the invention is patented. This document contains all the general characteristics related to the invention. The underlying view of this work is that some of the information (See Appendix) is representative of the importance of the invention. Thus, by constructing the correlated factual details into a cross-sectional set of variables, an econometric analysis could be carried out to provide a framework for the study of patent values.

The value-dominated variable used in the literature are:

i. Patent Family Size

The value-dominated variable 'family size' refers to the number of countries in which a patent grant has been sought. This proxy, first proposed in Putnam's (1996) PhD thesis, suggested that the collection of international patent grants is an indicator of patents value. Putnam used the patent family size as an extension to Pakes and Schankerman's (1984) original patent renewal decision model to allow the application and renewal of patents in more than one country. A number of papers have used family size as a proxy for patent value (see Harnoff et al. 2003b; Lanjouw et al. 1998; Guellec and Potterie 2000; Lanjouw and Schankerman 2004 among others). Harnoff et al. (2003b) used the Derwents World Patent Index (WPI) to estimate the family size of a sample of German patents. Their Ordered Probit regression estimations showed that family size contains particularly valuable information about patent values. Lanjouw and Schankerman (2004) developed an index of patent quality using family size for patents applied by US firms. Their results showed a strong positive association between the equality index and the firm's valuation. The authors showed this finding is robust and holds even when not controlling for year effects. Guellec and Potterie (2000) use the family size dummies as an explanatory variable in an econometric analysis of the likelihood that a European patent would be successful in the grant. Their results support Putnam's original hypothesis.

ii. Patent Claims

The value-dominated variable is patent claims, the number of 'components' embodied in patented invention, which appear at the front of the patent document.

The view underlying this approach is that each individual patent represents a bundle of inventive components. Thus, the number of components could be indicative of the value of each patent. Tong and Frame (1994) were the first to use patent claims data to model the technological performance of patents. The authors used the following example to explain how patent counts are the true measure of the value of a patented invention:

> Let us say that Martha invents the first stool and applies for a patent to protect her invention. In the claims section, she might write the following claim: 'I claim a device that can be used for sitting. This device is composed of a seat that is elevated off the ground by means of legs.'
>
> Let us assume that George spots Martha's invention and is stimulated to think of an improvement to it. He determines that the stool would be more comfortable if it has a back support to it. He thus invents a chair. Note, however, that in his claims section he can only claim the back of the chair, since the sitting component is already covered in Martha's stool patent. This is fitting, since George's true invention contribution is not the whole chair, but simply the seatback (Frame and Tong, 1994 p. 134).

To validate the use of patent claims, the authors compared the correlation of patent claims vis-à-vis patent counts with technological and economic strength indicators such as R&D expenditure, the number of scientists and engineers, gross national product, value of exports and counts of scientific and technological papers published. The sample contained 7531 patents granted in the US, but originating in the US, Japan, the UK, Germany and France. Their analysis showed that claims are much better than patent counts and have stronger correlations with the technology related indicators. Based on this relationship, Lanjouw and Schankerman (2004) used patent claims to formulate a factor model to analyse research productivity in the US. The factor model was estimated with more than 100,000 patents in seven different technological fields, applied between 1975 and 1993. Their results supported the Frame and Tong findings, as the number of claims was the most important determinant of research quality in six out of the seven technological fields.

iii. Patent Subclasses

The value-dominated variable is the number of subclasses the patent grant assigns the patented invention. This approach was suggested by Joshua Lerner (1994). The author pointed out that patent claim analysis, although having the potential to be a valid proxy for patent value, requires rigorous analysis of each patent and is not practical for large-scale economic research. Instead, the author showed that the number of International Classification of Patents (IPC) subclasses is more useful as they are determined through a rigorous bureaucratic procedure in the patent grant office (see p. 320 for details). The strength of the IPC system, which originated in the 1964 'European Convention on the International Classification of Patents for Invention', is its high standards, frequent revision, and strong industry and profession focus.

However, this definition of patent value is closer to technological diversity than to patent strength. It remains unclear how the broadness of the patent scope represents the technological or economic value of the invention.

iv. Patent Application Process

This approach examines the refusal, withdrawal or success of patent application across various dimensions (ownership, domestic and international co-operation, the number of applicants, and technology category) as a signal for patent value (Guellec and Potterie 2000; Guellec and Potterie 2002). The view underlying this approach is that a patented invention corresponds to a higher technological and economic value than an unsuccessfully patented invention. Although this approach provides some interesting insights about the value of patents, it does not appear to provide a systematic and consistent way to assess and analyse patent values.

v. Patent Citations

The value dominated variable is the number of citations a patented invention subsequently receives from future patents. Citations imply the use of the ideas embodied in existing patents to develop new and/or better patents. The view underlying this approach is that patent citations represent the impact each invention has had on creating new knowledge. Thus, this knowledge impact indicates the technological value of the ideas embodied in each invention. This technological value could then be used to measure the economic value each invention contributed to its inventors and assignees (private economic value of invention) and to society (social economic value of an invention).

Patent citations are the boundaries of the new invention. Citations delimit the scope of the new innovation as they indicate all the relevant existing knowledge-base that leads to the development of the new idea. Therefore, patent citations, unlike scientific citations, are the result of the legal requirement to validate the creation of new knowledge needed (Trajtenberg, 1990). Therefore, research shows that the legal process that provides the list of citations is generated by the incentives of the people involved (Campbell and Nieves 1979). The computerised system at patent grant offices now makes the citations retrievable. These properties make patent citations a superior proxy to the other alternatives discussed above (Bertran 2004) and have made patent citations the most preferred proxy to determine the value of patents. For these reasons, the empirical research followed in this thesis uses patent citations as a proxy for patent values.

4 What Are Patent Citations?

This chapter provides the conceptual, theoretical and empirical background for the use of patent citations as the value-dominated variable for the measure of patented inventions.

4.1 *Introduction*

Patent citations appear on the patent grant document (See Appendix). The citations indicate the 'prior-art' that the current patented invention is building on:

> If Patent B cites patent A, it implies that Patent A represents a piece of previously existing knowledge upon which patent B builds, and over which patent B cannot claim. (Hall et al. 2001 p. 14)

The legal requirements behind patent citations, contrary to scientific citations, give them a dimension of objectiveness. Campbell and Nives (1979, Appendix II) explain:

> "Patent Citations have a distinct legal and technical meaning and are produced by a distinctive process. These [citations] come from two sources: (1) the patent attorney and his or her client and (2) the patent examiner. For both these sources, the motivation to cite … another patent is embedded in patent law.
>
> First, the inventor's attorney must include by law citations to references in the specification of the patent application and in amendments that deal with related prior art. The inventor is obliged by law to bring to the attention of the patent examiner any relevant prior art of which he or she is aware. Failure to do so is considered fraud on the patent office, which places the patent (if issued) in risk of invalidation, if it can be shown that the patent was anticipated.

Second, the patent examiner must include citations in the file of the patent and in the printed patent that were used to further limit the claims and specification and the prior art, these references serve to narrow the scope of the of the patent. Thus, the citations contained in a patent's file represent the legal and technical judgement of the patent examiner (acting as an expert interested referee) and the patent attorney and inventor (as expert interested parties) with respect to the scope of the discovered hand. Not all citations included in the patent's file are printed in the patent. Only those citations specify the most relevant prior art of the patent in question.

4.2 *The Validity of Patent Citations as a Proxy*

The first studies to use patent citations were primarily bibliometric focused and concerned with the technological merits hypothesised in the citations patent receive over time (Ellis et al. 1978; Campbell and Nieves 1979; Carpenter et al. 1981; Narin et al. 1987; Lieberman 1987; Albert et al. 1991).[9] Ellis et al. (1978) used US patent citations as a source of information to map the history of specific technological fields. Carpenter et al. (1981) tested whether the average number of patent citations received is higher for patents whose underlying product received the IR100 award.[10] Their results showed that the group of 'important patents' received 2.5 times as many citations as the randomly selected control patents. Albert et al (1991) asked 20 researchers and research managers in Kodak, working in the area of silver halide technology, to rate 77 silver halide Kodak patents according to the technological impact each patent has had, where the number of citations each of these patents received ranged between zero and ten or

[9] With the exception of Lieberman (1987) who examined the relationship between patent citation and price change of a sample of 24 chemical products.
[10] The IR100 award is given by the *Journal of Industrial Research and Development* for the 100 most significant products developed (see Carpenter et al. (1981) for further discussion of the award).

more. Their results showed that the highly cited patents were of much greater technological significance than the infrequently cited or not cited patents.

The first systematic use of patent citations in economic research goes back to Trajtenberg's 1983 PhD thesis. Trajtenberg (Trajtenberg 1990a; 1990b) pointed out that the value of an innovation could be equated with the social benefits that it generates. Trajtenberg's underlying hypothesis was that patent citations could be used to indicate the ex-post social value of the ideas embodied in patented inventions. In an attempt to validate his hypothesis, Trajtenberg studied Computed Tomography (CT) technology. CT is a major medical innovation and is considered "the gold standard in the diagnosis of a large number of different disease entities" (Wikipedia, 2006). With the use of a discrete choice model, the social value of the CT innovation was measured as the incremental changes in the consumer and producer surplus of CT scanners marketed in the US. These estimations were then analyzed with all the 456 US CT granted patents from 1971 to 1986. The results showed that patents weighted by citations were highly correlated with this measure of social welfare.

With Trajtenberg's findings setting a benchmark, later studies further explored this relationship (Harnoff et al. 1999; Lanjouw and Schankerman 1999; Jaffe et al. 2000; Gay et al. 2005; Maurseth 2005). Harnoff et al.'s (1999) study showed a strong relationship between patent citation frequency and the private value of patented inventions as estimated through two surveys (one in Germany and one in the US) of the companies that hold these patents. Hall et al. (2005) indicated a strong association between patent citation and private economic value of innovations. Maurseth (2005) linked patent citation and renewal data, and showed that citations were positively correlated with the survival time of patents. Lanjouw and Schankerman (1999) analyzed the variance of a range of patent value indicators and showed that forward citations (citation received) were the least idiosyncratic. Jaffe et al. (2000) took a qualitative approach to investigate the nature of the patent citation mechanism as a signal of communications through surveying patent inventors. The perceived technological and economic values of the patents were found to be correlated with citation frequency. The authors could not ascertain which of the two perceived values had stronger associations with the citation frequency.

4.3 *Patent Citation in Applied Economics*

A growing literature has emerged as a response to the promising findings that patent citations could contain rich economic information. Consequently, a number of broad research areas have been explored using citations as the primary research tool:

4.3.1 The Value of Intangible Assets

Over the last two decades, there has been significant work done to estimate the value of the tangible and intangible assets of publicly traded firms. This body of literature[11] uses a 'hedonic' Tobin's Q model, $V_{it} = (A_{it} + \lambda K_{it})$ where the value (V) of a firm i at time t is a function of physical assets (A), knowledge assets (K) and the shadow value of intangible assets versus tangible assets (λ).

A number of studies have applied patent citations to advance the estimations by adjusting (K) according to the quality of the intangible assets, rather than simply the stock of invented inventions. Hall et al. (2005) matches citations to patents and show

[11] See Griliches (1981) and Griliches and Cockburn (1988) for early examples of this literature.

that each time a firm' patent is cited, its market value increases by three percent. Nagaoka (2005) found that the effect of patent citations on a firm's market value is greater in industries where innovation follows a strong cumulative process such as the ICTs industry (see Hall (1998) for a comprehensive review of this literature).

4.3.2 Path of Knowledge Flows

Patent citations data prevail over Paul Krugman's pessimistic view that "Knowledge flows … are invisible [as] they leave no paper trail by which they may be measured or tracked" (Krugman 1991, p. 53). Numerous studies have successfully applied patent citations to identify the path of knowledge flows across geographic locations, sectors, technologies and time. The approach was popularised by Jaffe et al. (1993) who showed a localisation in knowledge spillovers. The 'citation function' mechanism the authors used was subsequently challenged (see the two 2005 AER comments by Thomson and Fox-Kean and Jaffe et al.). Nonetheless, their study provided the inspiration for much of the ongoing work that is far more specific in its scope (See Jaffe 1998 for a review on the use of patent citations as a proxy for knowledge spillovers).

4.3.3 The Economic and Technological Impact of Patented Innovations

Intrigued by the earlier indications of the positive relationship between citation frequencies and the value of patents, numerous studies have applied patent citations to analyses of the performance and quality of patented innovations across countries, technologies, firms and sectors, and time. Trajtenberg et al. (1997) analysed the performance of corporate patents versus university patents. Henderson et al. (1998)

analysed the innovative performance of US universities as the result of the Bayh-Dole Act. Jaffe et al. (1998) analysed the effects of changes in R&D expenditure on the quality of NASA and other US Federal labs on patents activity and performance. Trajtenberg (2001) studied the quality of innovations of the US versus small innovative economies throughout time. Sakakibara and Branstetter (2001) examined the effects of the 1988 Japanese patent reform, which widened the extent to which patents claims could be included in one patent. Hall and Ziedonis (2001) examined the patenting performance of firms in the semiconductor industry. Jaffe and Lerner (2001) examined the effects of the 1980's US initiative to encourage US National Laboratories to patent.

4.4 *Summary*

The empirical investigation undertaken in this thesis is mostly nested within the "citation as a proxy for patent value" and trace of knowledge literature. The following chapter provides a description of the patent citations data used in this thesis.

5 Patent Citation Data

This Chapter describes the citation data used in this thesis.

5.1 *Overview*

Patent citations are a complicated network. The complexity of citation retrieval limits the capacity to use them in large-scale research. To identify the number of citations a given patent receives, one needs to observe the complete set of existing patents. This is a huge research task. Only recently, with the assistance of advances in ICTs, was this objective finally achieved.

During the 1990s, a team of scholars computerised the items that appear on the US patent grant document. These items were computed into the ‘NBER U.S. Patent Citations Data’ (Hall et al. 2001). Many of the studies discussed in the previous chapter extracted samples of this dataset and with the completion of the project in 2001, an accelerated number of ‘citations’ studies has emerged. The empirical exercise followed in this thesis is based on this dataset.

The data includes all the utility US Patent Office (USPTO) granted patents from January 1963 to December 1999.[12] The data were retrieved on December 1999. Three million patents were granted during that period, reaching over 16 million citations.[13]

---

[12] The USPTO classifies patent into three categories:

- **i.** Utility patents – invention and discoveries of any new and useful process, machine, manufacture.
- **ii.** Design patents - invention and discoveries of any a new, original and ornamental manufacturing design.
- **iii.** Plant patents - invention and discoveries of distinct and new variety of plant.

The last two categories are minor and were therefore, excluded from the dataset.

5.2 *Variables Used*

Patent Number - the USPTO patent number

Citations Received - the number of citations a patent receives from later patents.

Citations Made – the number of citations a patent makes to previous patents.

Country – the country where the first inventor of the innovation resides.

Technological Category – the USPTO classifies the granted patents into a 3-digit patent class. In this dataset, this class system is aggregated into six technological fields:

1. Chemical (excluding drugs)
2. Computer and Communications
3. Drugs and Medical
4. Electrical and Electronics
5. Mechanical
6. Others (Agriculture, Fixtures, Furniture, etc)

Grant Date – the data a patent is granted at the USPTO

Grant Year – the year a patent is granted at the patent USPTO

Application Year – the year an application is lodged at the USPTO

Assignee Type - the USPTO classifies the patent assignee into seven classes:

1. Unassigned - inventors who have yet to assign the right of their invention.
2. US non-Government organizations
3. Non-US non-Government organizations
4. US individuals
5. Non-US individuals
6. US Federal Government
7. Non-US Government

[13] The citations retrieval only started with patent granted from 1975 onwards. Citations information on pre-1975 patents could not be retrieved.

Generality - the extent to which a patented innovation spreads through and contributes to the development of patented innovations in a range of different technological fields.

Originality - the extent to which a patented innovation is broad in its scope in the sense that the innovation is based on knowledge coming from patented innovations belonging to a wide range of technological fields.

6 The Value of Patented Innovations

This chapter is the starting point for the analysis of the value of patented innovations. The analysis is based on a methodology first proposed in Trajtenberg (2001). To ensure full comprehension of the methodology and use of the patent citation data, the chapter replicates the original econometric estimations. This is intended to set the scene and provide context for the analysis carried out in Chapter 7.

6.1 *Overview of Trajtenberg (2001)*

Trajtenberg (2001) appears to be the first to provide a comprehensive and systematic econometric analysis of the technological value embodied in patented innovations across countries and time, as approximated by patent citations.

Trajtenberg's objective was analysis of the performance of Israeli inventive output. Israel is a small open economy with a strong reputation for significant and impressive inventive capabilities. It is widely recognised that if a second Silicon Valley exist, Israel is its base.

Trajtenberg used USPTO granted patents as the indicator for successful innovation and their received citations as a proxy for value. In an attempt to develop an inventive benchmark of performance, Trajtenberg constructed two control groups. The first group included a 1/72 random sample of all US originated patents in that period (US Group). The second group included Finland, New Zealand, Spain and Ireland. These four countries were selected according to their GDP per capita figures and population size. In the 1990s these countries were broadly comparable to the Israeli figures. Patents originating in these countries were aggregated into one group (the Reference Group). The period analysed ranged from 1965 to 1996.[14]

Three sets of dummy variables were created. The first set was country dummies, which included a US dummy and a Reference dummy. The dummy for Israel was omitted to avoid a dummy trap and was represented by the intercept term. The second set of dummies were Technology Type dummies which were constructed in order to control for the possibility that patents of different technologies would have different citation tendencies. The technological dummies included: chemical ('Chemical'), electrical & electronics ('Elec'), computers & communication ('Cmpcmm'), mechanical, drugs & medical ('Mech'), and other[15] ('Other'). 'Other' was omitted to avoid the dummy trap. The third set of dummies were Grant Year dummies, constructed to control for time effect.

The estimation method was linear OLS regression, where the number of citations received by each patent was regressed on the three set of dummies. The estimated coefficients for the country dummies describe the average frequency with which patents originating in a specific country are cited, while controlling for the age and technological field of patents. As the number of patent citations is indicative of technological and economic 'value', the regression estimation showed the relative value of patents. Therefore, the division of the average frequency with which patents originating in country A are cited, divided by the average frequency with which patents originating in country B are cited, yields the relative strength of country A's patents versus country B's patents.

[14] The Israeli patents extend to 1998.
[15] 'Other' included patents belonging to various miscellaneous industries such as Agriculture, Food, Apparel & Textile, House Fixtures, Earth working & Wells.

Table 1 shows Trajtenberg's results. All the estimated parameters were statistically significant. By comparing the country dummies and intercept term, the results showed the US has the 'best' patents (~3.7 average citations), then Israel (~3 average citations) followed by the reference countries of New Zealand, Spain, Finland and Ireland (~2.3 average citations). Based on these estimates, Trajtenberg concluded that US originated patents were about 25% better than Israeli patents, while Israeli patents were about 25% better than Reference country patents.

Table 1 Trajtenberg (2001) regression estimates (page 383)

ncites	Coef.	Std. Err.	t	$P>\|t\|$
usa	0.6954136	0.0793592	8.763	0.000
refer	−0.6985195	0.0855526	−8.165	0.000
chemical	0.335095	0.0773475	4.332	0.000
cmpcmm	2.372321	0.1090868	21.747	0.000
drgsmed	1.61299	0.107602	14.990	0.000
elec	0.3790388	0.0845855	4.481	0.000
mech	−0.2321834	0.0745865	−3.113	0.002
_cons	2.988059	0.0842784	35.455	0.000
gyear	$F(33, 37{,}272)$ (34 categories)	=	142.390	0.000

6.2 *Replication of the Results*

I replicate Trajtenberg's (2001) to set benchmark for later work. The replication results are presented as Table 2.

Table 2 Replication of Trajtenberg Results

US	**0.865**
	(7.45)**
Reference	**-1.169**
	(9.36)**
Chemical	**0.481**
	(4.23)**
Cmpcmm	**4.811**

	(30.89)**
Drgsmed	**3.058**
	(19.95)**
Elec	**0.775**
	(6.30)**
Mech	**-0.315**
	(2.87)**
Constant	**4.198**
	(34.72)**
GYEAR2 F(33,37393)=67.098	

Observations 37434
R-squared 0.09
Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%

All the estimates are statistically significant and their signs match the original regression. Nevertheless, the values of the coefficients differ somewhat from the original estimations, although they do not fundamentally overturn the original results. According to the replicated coefficients, US patents are 20.6% better than Israeli patents as US patents receive ~5 citations on average whereas Israeli patents receive ~4.2 citations on average, and Israeli patents remain significantly better than patents originating in the reference countries as Israeli patents receive ~4.2 citations on average whereas reference patents receive ~3 citations on average. A plausible explanation for the different coefficients estimated is the use of a larger dataset in replication regression than the one used in Trajtenberg (2001) as the Hall et al. (2001) data, which is the one Trajtenberg (2001) is primarily relying on, would have been a work in progress at the time the original regression was estimated. In addition, the sampling technique used in Trajtenberg (2001) is not reported and the author could have quite possibly followed a stratified sampling technique, which would have led to different results.[16]

[16] Unfortunately, the original regression code no longer exists.

6.3 *Discussion*

Trajtenberg's (2001) approach is valuable and meaningful. Under some very sensible assumptions of the importance of citations as a value indicator, the OLS analysis provides a clear measure of the average value of patented innovations.

Nevertheless, I suggest that Trajtenberg's analysis could be very sensitive to a number of important factors that appear to have been overlooked. Firstly, as patent citations are not a normally distributed variable, OLS analysis may not provide robust results. Secondly, although aggregation of patents into a reference group may enhance the statistical power of the model, it could potentially lead to misspecified estimation. Thirdly, the nature of the citations data and institutional factors at the USPTO office may induce the possibility of a break in the data, which require a thorough empirical examination. Trajtenberg, arbitrarily chooses the year 1986 to divide the data into two samples to without any explanation or ex-ante theoretical justification and re-estimates the model. Although Trajtenberg's re-estimation results were primarily consistent with his original estimates, I suggest that this approach for testing for the robustness of the estimates over time is simplistic, atheoretical and does not accurately capture the possibility of breaks. Fourthly, unweighted patent citation could contain a considerable amount of noise (Jaffe et al. 2000) and would therefore require careful examination of the results. These are the topics of the next chapter.

7 Tests of Robustness

The objective of the chapter is to build on Trajtenberg's original methodology by testing for the robustness of the estimations. Chapter 6 suggested that the OLS analysis, although very valuable, could be sensitive to the statistical properties of the dependent variable, the structure of the underlying model, breakdates in the data and noise in citations counts. This chapter discusses and tests these factors in detail. The results indicate that these factors, if overlooked, may lead to inappropriate and/or invalid econometric estimations.

7.1 *The Statistical Nature of the Dependent Variable*

The first test of robustness stems from the statistical characteristics of patent citations. The analysis of patents quality carried out in Trajtenberg (2001) and replicated in Chapter 6 was based on an Ordinary Least Squares (OLS) model. The OLS model is a normal linear regression and is applied when the variable of interest, the dependent variable, is continuous and normally distributed around the mean. When the dependent variable fails to satisfy the above statistical characteristics, the OLS predicted outcomes could lead to inefficient, inconsistent and biased estimations (Long 1997).

The number of patent citations received is not a normally distributed variable, see Figure 1. The histogram clearly indicates that citations have an extremely skewed tail. The variable has a variance of 53.95811, skewness of 7.768 and a Kurtosis value of 291.6572.

Patent citations represent the occurrence of the event that a granted patent cites an existing patent in a fixed period. When this event occurs, a citation is added to the stock of citations of a patent. Thus, patent citations are the counts of such an event. This implies that theoretically, patent citations should be analysed using Count Data models, which can explicitly model the nonnegative characteristic of citations. I therefore test whether Trajtenberg's OLS predictions are robust and consistent when carried out in the context of Count Data model.

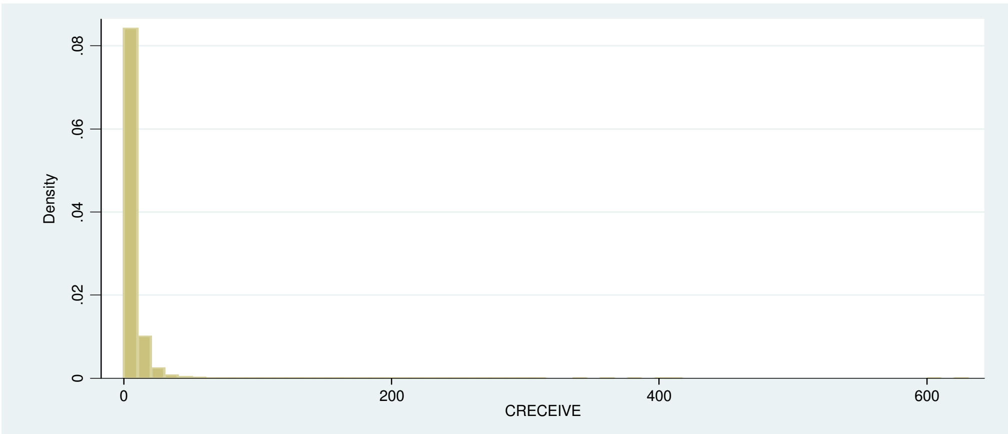

Figure 1 The histogram of the number of citations received by patents.

7.1.1 Count Data Models

Count data models are now widely used in microeconometrics.[17] In many micro applications, the underlying data take non-negative random integer values and follow a count data process that necessitates the use of specialised estimation techniques. The two 1984 papers by Gourieroux, Monfort and Trognon, and Hausman, Hall, and Griliches provided the early groundbreaking methodological techniques for the analysis of micro data of this kind. Cameron and Trivedi (1986), Winkelmann (1994), Long (1997) and Cameron and Trivedi (1998) provide a good overview of the standard

[17] See, for example, the special issue of the *Journal of Applied Econometrics* (1997) that is devoted to the analysis and applications of count models in economics.

models available for the regression analysis of count data. In the presentation and analysis of this topic, I shall follow their texts closely.

The Poisson regression model appears to be the most frequently referred to count data model. The model is based on a Poisson distribution. The distribution describes the probability associated with the number of events occurring in a specific time interval and is derived from the stochastic Poisson process, which assumes the independence in the occurrence of the underlying events.

Let y be the variable of interest. In our case, y is a random variable indicating the number of citations a patent receives from subsequent patents over time. Patent citations follow a Poisson distribution if

$$\Pr(y|\mu) = \frac{\exp(-\mu)\mu^{y}}{y!} \quad \text{for y=0,1,2…}$$

This equation implies that the probability of a certain citation count depends on the parameter μ, the mean number of times a patent is cited per unit of time, and y, the citation count of interest. As the mean citations increase, the probability of low citation counts decreases and the distribution shifts to the right and approaches the normal distribution. The Poisson distribution assumes equidispersion of the mean and variance, which imposes equality of the two: $\text{var}(y) = \mu$.

The Poisson Regression Model estimates the expected value of a dependent variable given a number of independent variables: $\mu_i = E(y_i|x_i) = \exp(x_i\beta)$. The variations in the estimations of μ are then due to variations in the values of the explanatory variables x_i. This property is known as the observed heterogeneity.

The estimated μ_i provides the expected number of citations per patent conditional on certain characteristics that are of interest. This conditional mean is then used to find the probability of various citation counts, y: $\Pr(y_i | x_i) = \frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!}$.

Frequently, the Poisson distribution does not model the economic data well. The common explanation is that the strong assumption that all heterogeneity in the conditional mean of the variable of interest is observed is invalid (Long 1997). Under a Poisson regression model, the variation of μ is the result of different values of the explanatory variables. When analysing the probability of the occurrence of an event, the estimated rate at which the event occurs (μ) would be identical for all observations conditional on similarity in the set and values of the explanatory variables. The failure to count the unobserved heterogeneity results in inequality between the conditional mean and the conditional variance, which violates the assumption underpinning the Poisson model and hinders its validity.

When the conditional variance exceeds the conditional mean, the data is said to be overdispersed, whereas if the conditional mean exceeds the conditional variance, the data is said to be underdispersed. The most common observation is an overdispersion in the data. When this occurs the standard errors of the Poisson regression estimates are biased downwards with very small p-values (Cameron and Trivedi, 1986). The Negative Binomial model is the common model used when overdispersion characterises the data of interest.[18] The model accounts for the unobserved heterogeneity that could not be explained by the regressors as it adds a random error term to the regression structure. The estimated conditional mean then becomes:

$$\tilde{\mu}_i = \exp(x_i\beta + \varepsilon_i) = \mu_i \exp(\varepsilon_i) = \mu_i\varepsilon_i$$

The expected value of the error term is assumed to equal one, which implies that the expected value of the $\tilde{\mu}_i$ still equals μ_i.

$$E(\tilde{\mu}_i) = E(\mu_i\delta_i) = \mu_i E(\delta_i) = \mu_i$$

However, the conditional variance is allowed to differ and becomes:

$$\text{var}(y_i \mid x_i) = \mu_i(1+\alpha\mu_i) = \exp(x_i'\beta) + \alpha[\exp(x_i'\beta)]^2 = \mu_i + \alpha\mu_i^2$$

where α is the variance of the error term and is known as the dispersion parameter.

The equation above implies that the Negative Binomial Model is a generalisation of the Poisson model. That is, when $\alpha = 0$ the Negative Binomial model is reduced to a Poisson model. Interested readers can refer to Long (1998) and Cameron and Trivedi (1998) for further discussion.

7.1.2 Estimation

In order to compare the Negative Binomial with the Poisson model, I fit the citations predictions of the two distributions against the actual citations received in the period, see Figure 2. The graph clearly indicates that the fitted Poisson model over-predicts the counts four, five, six, seven and eight and under-predicts zeros, ones and twos, whereas the Negative Binomial model fits the data much more accurately, see Figure 2. The estimated overdispersion parameter is 1.226. I therefore choose the Negative Binomial Model to estimate Trajtenberg (2001) regression, Table 3.

[18] For a model that deal with underdispersion, see Cameron and Johansson (1997).

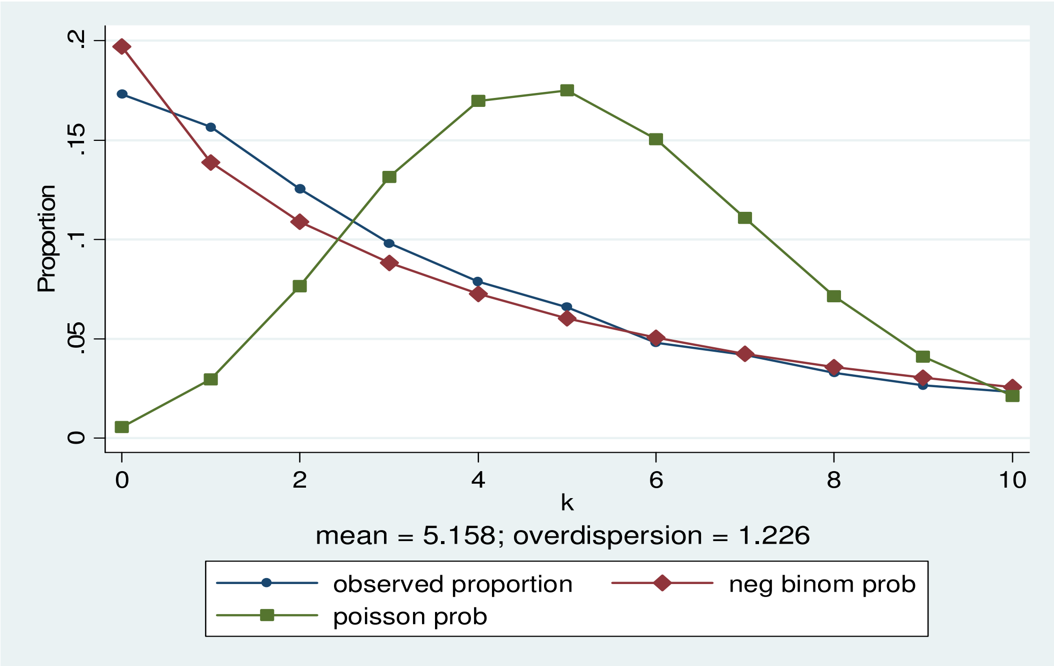

Figure 2 Comparison of the Poisson and Negative Binomial distributions and the observed distribution of citations received.

The computation and interpretation of the intercept term differs between OLS and Count Data models. Whereas under the OLS, the intercept is the value of the population regression line when all regressors are equal to zero and thus represents the relative quality of Israeli patents, the nonlinearity of Count Data models implies that the intercept term cannot be interpreted in the same way. I therefore run three separate regressions, identical in structure and data to OLS, but include only two country dummies in each regression. This allows straightforward inspection of the relative performance of each of the examined country dummies and a test for the robustness of the original predictions. The first regression is for the US vis-à-vis Israel, the second regression is for US vis-à-vis Reference country, the third regression is for Israel vis-à-vis Reference country.[19]

[19]The year dummies are included and are statistically significant but are not shown.

Table 3 Negative Binomial Regression

	1	2	3
US	0.407	0.163	
	(29.54)**	(9.42)**	
e^b	1.5026	1.1742	
%	50%	18%	
Israel	0.244		-0.163
	(12.95)**		(9.42)**
e^b	1.2797		0.8516
%	28%		-15%
reference		-0.244	-0.407
		(12.95)**	(29.54)**
e^b		0.7815	0.6655
%		-22%	-33%
chemical	0.099	0.099	0.099
	(5.78)**	(5.78)**	(5.78)**
cmpcmm	0.862	0.862	0.862
	(36.46)**	(36.46)**	(36.46)**
drgsmed	0.518	0.518	0.518
	(22.59)**	(22.59)**	(22.59)**
elec	0.157	0.157	0.157
	(8.41)**	(8.41)**	(8.41)**
mech	-0.072	-0.072	-0.072
	(4.29)**	(4.29)**	(4.29)**
Constant	-1.449	-1.205	-1.042
	(21.19)**	(18.21)**	(15.30)**
Observations	37434	37434	37434
R^2	0.0352	0.0352	0.0352

Absolute value of z statistics in parentheses
* significant at 5%; ** significant at 1%
e^b = exp(b) = factor change in expected count for unit increase in X
% = percent change in expected count for unit increase in X

7.1.3 Interpretation of the Results

The estimated coefficients require a different interpretation compared to the coefficients estimated under the linear OLS model. Whereas under the OLS analysis, the patent inventive technological performance was obtained directly by the country coefficient, the patent inventive technological performance under count data analysis requires further computations of the estimated country dummy coefficients.

The count data coefficients provide the factor change in the expected patent citations for patented invention originating in different countries. If ω represent the country of interest and $E(y|x,x_k)$ represents all the other variables in the regression model, then $\frac{E(y|x,x_k+\omega)}{E(y|x,x_k)}$ is the factor increase or decrease in expected patent citation when this specific country's patents are tested. The factor change estimations are derived by taking the exponential value of the country dummy, while holding all other variables in the model constant: $\frac{E(y|x,x_k+\omega)}{E(y|x,x_k)} = \exp \beta_k \omega$.

The factor change estimations are interesting as they provide indications of the relative strength of the patented inventions in a specific country vis-à-vis other analysed countries. A factor change of the country's dummy that is greater than one implies that the change in the quantity of expected citations is positive for that country, whereas if it is less than one the change is negative. Note that the factor change is constrained to be positive as we are taking the exponential value of the dummy coefficient. The factor change estimations can also be then used to identify the percentage change of the expected patent citations for patented invention originating in specific countries. The percentage change is $100 * \frac{E(y|x,x_k+\omega) - E(y|x,x_k)}{E(y|x,x_k)}$ which is $100*(\exp \beta_k \omega - 1)$.

7.1.4 Results

All the parameters estimated via the Negative Binomial model are statistically significant and their sign match the original Trajtenberg results, see Table 3. The Negative Binomial analysis of citation reveals that the OLS estimates are robust to the statistical nature of patent citations. The 'best' patented innovations according to the Negative Binomial model remain US originated ones. Regression 1 shows that the US dummy has a factor change of 1.50 whereas the Israel dummy is slightly behind with a 1.27 factor change. This implies that given a patent is invented in the US, the expected citation it receives would increase by a factor 1.50, corresponding to a 50% rise in citation stock, whereas if it is an Israeli patent, the expected citations would increase by a factor of 1.27, corresponding to a 27% increase.

The Negative Binomial estimations also support the advantage of Israeli patents over Reference country patents. Regression 3 shows that the Israeli dummy has a factor change of 0.85 whereas the reference dummy has a 0.66 factor change. This implies that given a patent is originated in Israel, the expected citation it receives decreases by a factor of 0.85, whereas if it originates in a Reference country, the expected citations would decrease by a factor of 0.66.

7.1.5 Summary

The results indicate that Count Data estimates are similar to the ones by OLS, as the Negative Binomial regression does not fundamentally overturn the OLS results. Both regressions point to a strong relative advantage of US originated patents vis-à-vis the other country patents, and a relative advantage of Israeli patents vis-à-vis the Reference patents. A plausible explanation for this finding is that the large number of the citation observations[20] pushes the count variable towards a continuous variable which can be analysed in the context of a linear regression.

[20] Reaching almost 40,000 observations.

7.2 *Breakdates in the Data*

The second test of robustness stems from the possibility of breakdates in the citations data. If breaks exist, they may affect the robustness of model parameters and the predictions derived. The current literature on applied patent citations appears to neglect this matter.[21]

7.2.1 Theoretical Explanation for the Existence of Breaks

I review material on structural adjustments that have occurred at the USPTO office and documentation on the retrieval of the Hall et al. (2001) data. The material reveals that the existence of breaks is highly possible. There are three reasons for the possibility of a break.

7.2.1.1 Truncation Effect

The first reason for a break is the truncation effect. The truncation effect is inherent in citation analysis and occurs when the citation data is collected. Receiving a citation is a lengthy process. The longer the patent exposure is, the likelier it is to be cited. As the Hall et al. (2001) data was collected in 1999, patents granted in 1998 (for example) would have had only a one year exposure, which would severely affect their citations stock. Research shows that citations still arrive even after ten years of exposure, and many years pass until the number of citations received actually matures (Hall et al. 2001). The truncation effect is apparent in Figure 3, which sorts the average number of patent citations received according to the patent grant year. The graph shows a drastic

[21] For example, Trajtenberg (2001) tested for the robustness of the results over time but did so by just arbitrarily choosing a year to divide the data without any explanation or ex-ante theoretical justification for dividing the data. Information motivating the possibility of a break was not provided or suggested.

decline in the average number of citations as the patent grant year approaches the data collection date.

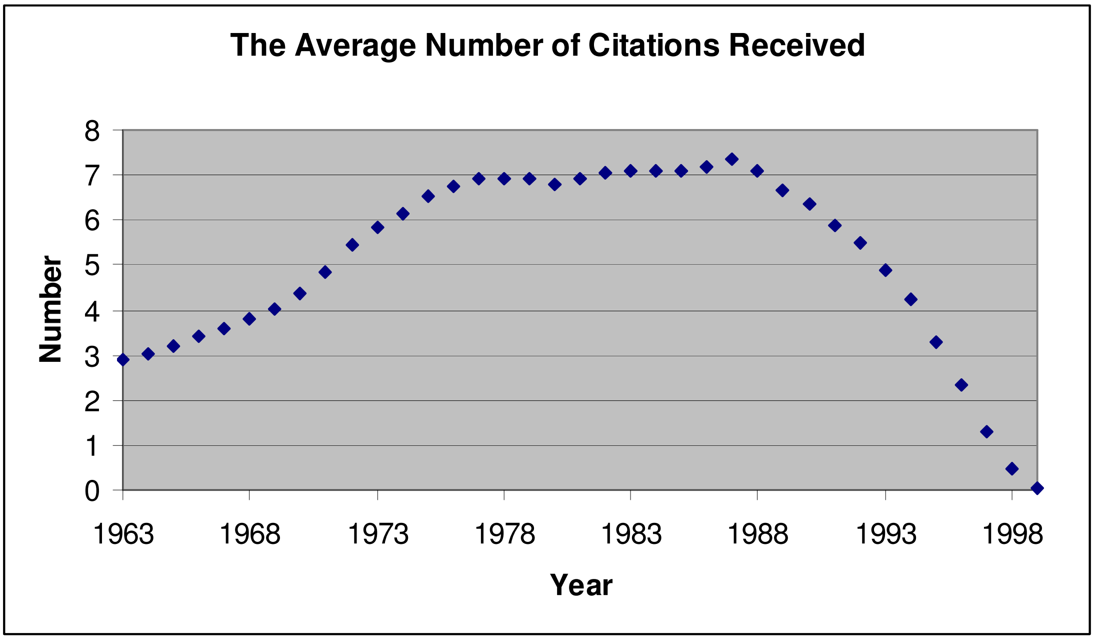

Figure 3 The Average Number of Citations Received

The most common approach to dealing with patent citation truncation in the literature is to "take quite a wide time window to get significance coverage of forward citations" (Hall et al. 2001, p. 17). This is the reason grant year dummies were added in the Trajtenberg (2001) analysis.

7.2.1.2 Patent Explosion

The second reason for a break in the data is patent explosion. Figure 4 shows the number of patents granted at the USPTO from 1963 to 1999. The graph shows that the number of patents granted more than tripled in that period. From 45,000 in 1963, they tripled to over 150,000 in 1999.

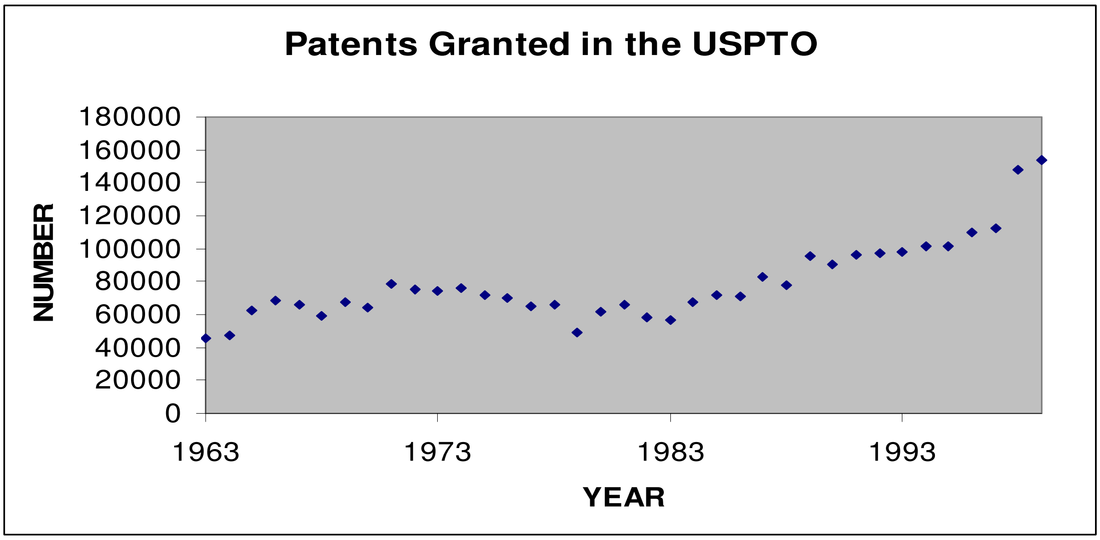

Figure 4 Patents granted in the USPTO from 1963 to 1999

The major jump in patent grants occurs in the mid eighties. The reasons for the acceleration in patenting (patent explosion) are US Congress adjustments to the US patent system. Jaffe and Lerner's (2004) recent award winning book documents these changes and their harmful consequences on the quality of US granted patents. The main adjustment occurs in 1982. Prior to 1982, patent disputes were settled in a district court. These courts differed in their interpretation of patent law, leading to considerable consequences on their rulings. This led to the development of perceived friendly and less friendly courts, whereby firms would strategically decide in which court to lodge the claim, resulting in a severe undermining of the US Patent Office. In 1982, US Congress established a centralised patent appeal court, the Court of Appeals for the Federal Circuit (CAFC), in an attempt to restore order in the chaotic patent system and to strengthen patent holders rights. The CAFC

1. Increased the incentive to patent by firstly making certain new technologies patentable
2. Lowered patent grant standards
3. Made patent rights more durable

The result was a significant increase in patent applications and grants. The Patent Office struggled to find qualified and knowledgeable patent examiners that could deal with the dramatic increase in patents applications and the new technologies that they cover.

A few years after introducing the CAFC, Congress converted the USPTO from an agency that runs on tax revenues to a 'profit-centre' funded by fees. It is commonly believed that the CAFC and the changes in tax structure had led to a dramatic increase in the grant of trivially obvious and/or dubious patents (Jaffe and Lerner 2004).

The implication of the increase of patent grants on patent citation is very simple. The higher the number of granted patents, the more patents that are cited. This implies that patents granted after the Congress adjustments may have higher citation tendencies simply because there are more patents to cite. The increase in citations made by each newly-granted patent is apparent in Figure 5, which sorts the average number of patent citations made according to when the citing patent was granted. A strong positive trend is visible, which could suggest that there exists a point in time where the average number of citations received by each patent had shifted due to the Congress adjustments discussed in above. If such a break in citations received exists and is due to these Congress adjustments, it is likely to occur prior to the introduction of the adjustments as citations go back in time.

7.2.1.3 Citations Retrieval

The third reason for a break in the data is due to data collection. The Hall et al. (2001) data begins the identification of citations made by each new granted patent only for 1975 granted patents onwards. Citations made by patents granted prior to 1975 could not be retrieved as the USPTO did not store computerised patent file information in that period. The result is evident in Figure 3 that clearly shows a steep rise in the number of citations received after 1975. Patents granted prior to 1975 still receive citations from patents granted after 1975 but not from patents granted prior to 1975.

Furthermore, the computerisation of the USPTO patent file process in the 1970s and 1980s, would have made the search for 'prior art' by examiners much easier and more efficient, which is another reason for the increase in citations made by each new-granted patent apparent in Figure 5 and the possibility of a break in the data.

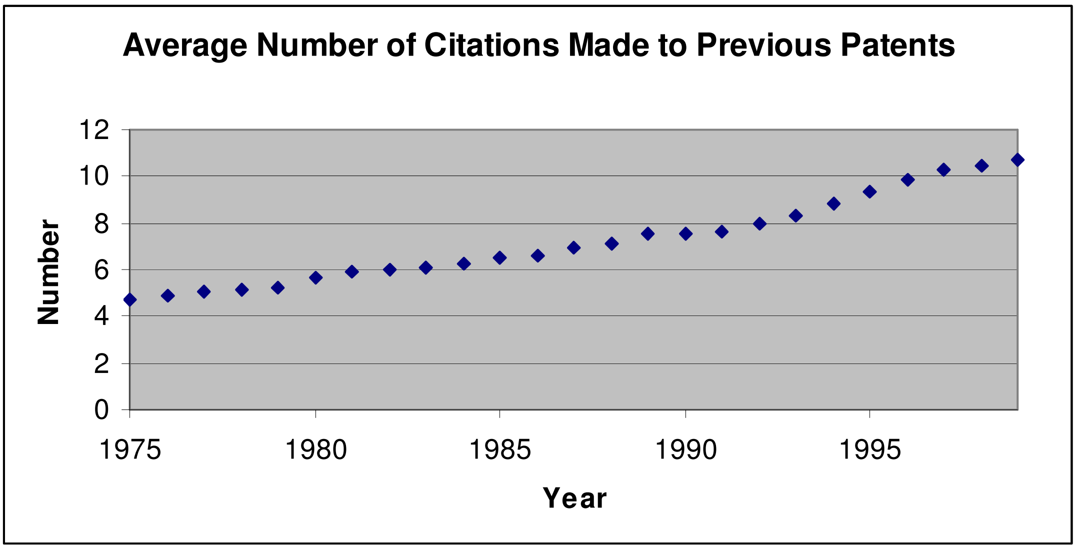

Figure 5 Average Number of Citations Made to Previous Patents

7.2.2 Detection of Breakdates

The detection of breakdates in the data requires the use of an econometric test. Arbitrarily dividing the data into sub-periods according to some hypothesized break period, although it might be practical, is too simplistic in this case as we do not have prior knowledge on the exact date the breaks occur.[22] Since the distribution of patent citations is sensitive to the period and size of the sample analysed, I propose the use of

[22] See Hansen (2001) for a discussion and an example of the pervasiveness of breaks in economic data.

a non-parametric approach for the detection of breaks, which avoids making assumptions about the distribution of the underlying variable.

7.2.2.1 Regression Tree

I suggest the use of a Regression Tree test as the non-parametric procedure for the detection of breaks. The development of Regression Trees goes back to Breiman et al. (1984). The application for dating structural break whose dates of occurrence are unknown was proposed in Cappelli and Reale (2005). In the presentation of this technique, I shall follow their approach closely.

Regression trees are a useful technique to discover and explore hidden information that might exist in large datasets. They can be used to predict the values of the dependent variable for a range of structured relationships observed in the data. Their strength is their relative methodological simplicity in implementing least squares partitions and their engaging visual presentation of the partition process, which provides a hierarchical representation of the data, allowing straightforward inspection for estimation errors.

Regression trees make use of a least squares methodology in minimising the squared errors between the observation and their mean value. The construction of a regression tree often is achieved with the division of the data into a 'training set', which is used to construct the regression tree, and a validation set that is used to trim the tree.

Consider a random vector (Y_i, X_i) consisting of n cases (Y_n, X_n). Let $f(X)$ be the predictor of the dependent variable, Y, for given values of the independent variable, X. Under the conditional expectations of the dependent variable, given the measurement vector $E(Y|X = x)$, $SS(h) = E(Y - f(X))^2$ describes the measurement error or the mean squared errors implied by $f(X)$ under the Least Squares Regression Trees (LSTR). The LSTR operation fits a group mean which represents the value of $f(X)$ that minimises the sum of squares of all Ys for the n cases that fall under this group. Breiman et al. (1984) provide a detailed discussion on the properties of this process.

The data splitting follows a recursive process. A split is a binary question which induces partition of the Y observations into a left descendent if, for example, $(x_{ij} < g)$ or into a right descendent if $(x_{ij} \geq g)$, for all of g ranging in the domain of x_i. The best split of the data is the one that leads to maximisation in the reduction of the deviance of the sum of squares. The best split, selected by the algorithm, is the one that leads to the minimisation of the split deviance and hence maximisation of the difference between the sum of squares at node, h $(SS(h))$ and the within-group deviance of the right and left descendents: $SS(h_r) + SS(h_l)$ .Formally, the algorithm iteratively splits the data to obtain maximisation of $\Delta SS(\mu, t)$: $\Delta SS(\mu, t) = SS(h) - [SS(h_i) + SS(h_r)]$ so that $\Delta SS(\mu^*, t) = \max_{\mu \in \Psi} \Delta SS(\mu, t)$, where the set of premised splits is Ψ.

The result is recursive partition of the data until no further gain of $\Delta SS(\mu, t) = SS(h) - [SS(h_i) + SS(h_r)]$ can be achieved. At this stage, overfitting the data into a large number of nodes is a common difficulty, which is avoided by following a pruning method that trims the tree based on a measurement criterion such as the popular AIC, BIC (Schwarz, 1978) or the RIC criteria based on Shi and Tsai (2002). Alternatively, a predetermined rule can be followed which limits the number of attained nodes, using a condition that stops the algorithm from further partitioning the data when specific conditions are met.

Cappelli and Reale (2005) show that if as a covariate we use a strictly ascending or descending frequency of numbers, then the Regression Tree would identify structural breaks on the mean. Regression Trees provide a number of advantages over other existing structural breaks tests, such as the Chow and the Bai and Perron tests. In the case of the Chow test, the Regression Tree can detect large number of breaks whereas the Chow test can is limited to one break at a time. Furthermore, the Chow test requires a predetermined break date to carryout the test, whereas Regression Tree does not. In the case of the Bai and Perron test, Regression Trees are much quicker in dealing with large datasets and do not show tendencies to underestimate the number of breaks whereas the Bai and Perron test does (Rea et al. 2006)

7.2.2.2 Results

Using the Regression Trees methodology presented in previous Chapter, I test whether structural breaks exist in the data used by Trajtenberg (2001).[23] The number of citations received is the variable of interest and the tree algorithm is computed to identify all admissible splits in citations count during that period, see Figure 6.

The values above the node indicate the split point and values beneath the terminal node indicate the mean number of citations received in that specific sub-period. The tree identifies two decision nodes, indicating the occurrence of two breaks and three regimes. The first is the pre-1971 period, the second is the 1972 to 1993 period and the third is post-1993 period.

[23] I use the *tree package* in the R statistical computation software to estimate the tree.

The 1993 break is easy to explain. It is solely due to the truncation effect as indicated by the time the break occurs and the value underneath the terminal node. The post-1993 patents receive less than three citations on average, lower than the citations expected under the two other regimes. Patents granted in 1995 would have had less exposure time compared to a 1985 patents, and thus, the lower citations expected.

The 1971 break is more curious, as it occurs prior to the 1982 adjustments and prior to the beginning of citation count in 1975. It is quite possible that the break captures both effects since the two are reinforcing each other. The 1982 adjustments led to a patent explosion and the rise of the average citations made by post-1982 patents and the consequent rise in citations received by pre-1982 patents. The tree suggests that the rise of citations received goes back to 1972. The computerisation changes of 1975 would have contributed and accelerated the high citations tendency as it became easier to search for ‘prior-art’. These two effects work in same direction and are picked up by the tree as significantly increasing the citation tendencies during the 1972 to 1993 period. As expected, the 1972 to 1993 period has the highest citations received, 6.17 per patent.

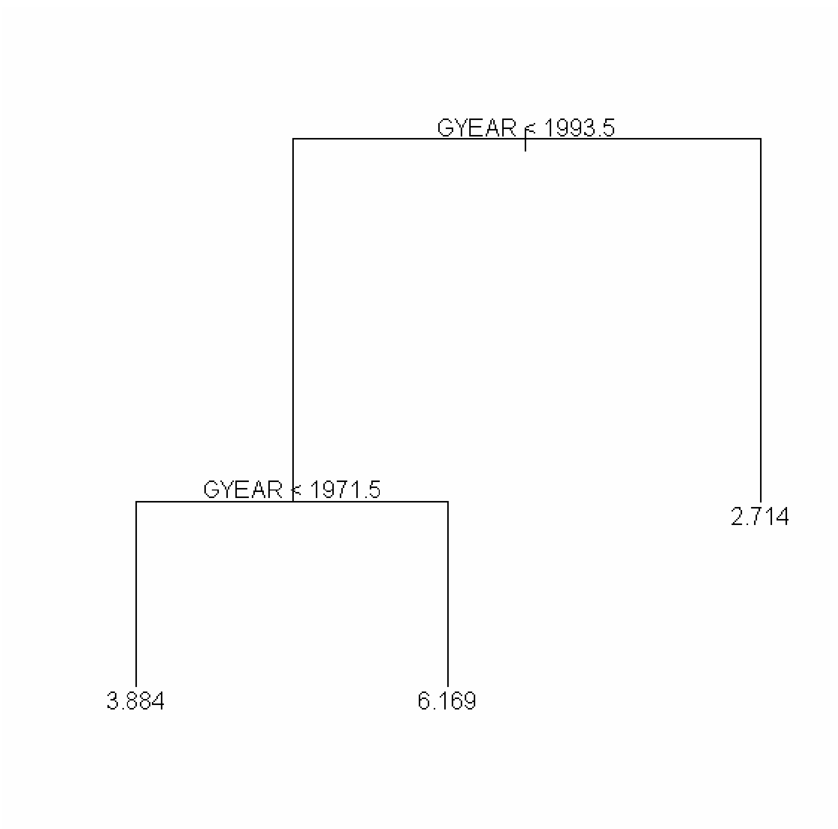

Figure 6 Regression Tree of the number of citations received partitioned across grant years. Values above the node indicate the split point and values beneath the terminal node indicate the mean number of citations.

The value of the pre-1971 terminal node is interesting. Although the count of citations only began in 1975, these period patents receive 3.9 citations on average, which are significantly higher than the ~ 2.7 citations during the truncation bias period. This confirms that citation is a lengthy process with significant time lags. Although it is difficult to know the precise cause for the 1970 break, the discussion in Jaffe and Lerner (2004) and data computation procedures would suggest that it is likely to be due to changes in the Patent Office. Beyond that, we cannot tell the extent of the 1982 effects on as opposed to the 1975 data computation effects on the break. The only way to identify whether the breaks reflect a real transformation in the data is to re-estimate the model for each period.

7.2.2.3 The Implication of the Breaks on the Estimations

I test the impact of the breaks on Trajtenberg (2001) estimations by running three OLS regressions, one for each sub period, Table 4. The results show that the Trajtenberg (2001) estimations are highly unstable across periods. The expected citations significantly change across three regimes, see Table 4. Only the 1972 to 1993 period (the majority of the data) resembles the original results. The estimates are all statistically significant and their sign match the original regression. However, for the pre-1971 period, the reference country patents are on par with Israeli patents as the reference dummy is statistically insignificant, which different from the results obtained in Trajtenberg (2001). Furthermore, in pre-1971 regression many of technological dummies become statistically insignificant and the remaining significant technological dummies change their signs compared to the original regression.

In the post-1993 period, the citations are highly truncated as the average citations estimated by the country dummies is significantly lower than in is suggested in the original Trajtenberg (2001) regression. The estimations in Table 4 therefore imply that the Trajtenberg (2001) results are not robust to the presence of breaks.

Table 4 Three OLS regressions, one for each sub period.

	<=1971	1972- 1993	>=1994
US	1.163	0.843	0.642
	(3.92)**	(5.61)**	(4.26)**
reference	0.074	-1.358	-0.737
	-0.22	(8.43)**	(4.67)**
chemical	-0.207	0.723	0.015
	-1.08	(4.84)**	-0.09
cmpcmm	-0.405	6.103	3.677
	-1.24	(26.92)**	(21.48)**
drgsmed	2.017	4.25	1.31
	(4.76)**	(20.67)**	(7.25)**
elec	-0.835	1.048	0.908
	(4.04)**	(6.26)**	(5.19)**
mech	-0.542	-0.331	0.042

	(2.95)**	(2.28)*	-0.24
Constant	3.142	5.148	1.667
	(10.05)**	(32.21)**	(11.10)**
Observations	6034	24623	6777
R-squared	0.02	0.07	0.15

Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%

Table 5 Expected Citations

	<=1971	1972- 1993	>=1994
US	4.305	5.991	2.309
Israel	3.142	5.148	1.667
Reference	3.216*	3.79	0.93

* insignificant at 5%

7.3 *The Structure of the Model Estimated*

The third robustness test examines the structural form of the Trajtenberg regression model. The aggregation of patents originating in Finland, Ireland, Spain and New Zealand into one control group, instead of four independent dummy regressors, could lead to a significant loss of information and thus, to a loss of accuracy of the results. It is therefore imperative to test whether the aggregation of the dummies yields robust and consistent results. Treating the Trajtenberg aggregated reference dummy model as a restricted model if nested within the unrestricted disaggregated model, hence a log likelihood ratio test could be used to test for the effects of aggregation. As the aggregated model imposes that the patented innovations originate in one 'big' country, it is a simpler version of the disaggregated model and typically would have a lower maximum likelihood value. The log likelihood test asks whether the reduction in likelihood value is statistically significant and is carried out by looking at the difference between the log likelihood values of the restricted and unrestricted models. If $-2(V_{Restricted} - V_{Unrestricted})$, where $V_{Restricted}$ is the log likelihood value of the restricted model and $V_{Unrestricted}$ is the log likelihood value of the unrestricted model, is greater than the χ^2 value (with n degrees of freedom, where n is equal to the difference in the number of free coefficients estimated in each case) then the null is rejected and the unrestricted model is be said to be significantly better than the restricted model.

Table 6 shows the results of the unrestricted regression.[24] With the exception of the Irish dummy, all the country dummies are statistically significant and their signs match Trajtenberg (2001) estimated parameters. However, the size of the coefficients varies significantly across countries. US patented innovations remain the 'best' with a ~5 averaged citations, while Israel maintains its advantage over the remaining countries with 4.2 averaged citations, with the exception of Ireland.

The insignificance of the Irish dummy is picked by the log ratio test., which strongly rejects the null hypothesis. The computed log ratio value is 63.96, which under t-statistics of three degrees of freedom at the five percent significance level corresponds to a P-Value of 0.000 and the rejection of the test. This implies that the aggregation of country dummies leads to a significant loss of information, which therefore suggests that Trajtenberg's (2001) regression structure does not capture the full information contained in the data. That is, aggregation of country dummies into one 'big' country results in a significant loss of valuable information for the analysis.

Table 6 Results of the unrestricted regression

US	0.855 (7.37)**
NZ	-1.2 (4.98)**
FI	-0.969 (6.74)**
ES	-1.919 (11.28)**
IE	0.016 -0.06
Constant	4.231 (34.93)**

[24] The year and technology type are included but are not shown.

Observations	37434
R-squared	0.09

Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%

7.4 *The Noise in Patent Citations*

The fourth test of robustness is more a test of accuracy. It stems from the possibility that citation count may contain a large degree of noise. Just like in the case of bibliometric analysis of scientific citations whereby the importance of an academic paper is determined by the number of citations it receives and the importance of the citing paper (e.g. the importance of journal where the citing paper is published), it is valuable to test the accuracy of the estimates when citations are weighted according to some 'value' index. To do so, I compare Trajtenberg's estimations to two independent patent citations-based 'value' measures, developed and tested by Henderson et al. (1997).[25] This is the first direct comparison of regression results estimated with citations count dependent variable against citations-weighted dependent variable. The estimation procedure is OLS regression.

7.4.1 Independent Citation-Based Value Indexes

The first 'value' measure is the Generality Index, which indicates the extent to which an innovation spreads through and contributes to the development of innovations in a range of different technological fields.

The index is computed according to the Henderson et al. (1997) F/Generality computation and is a Herfindahl index of concentration.

Let $Generality_i = 1 - \sum_{k=1}^{Ni} \left(\frac{NCITING_{ik}}{NCITING_i} \right)^2$, where $NCITING$ is the number of citations a

[25] The two measures are positively correlated with the citations frequencies.

patent receives, k is the index of patent classes, N_i is the number of different classes.

This implies that $\left(\frac{NCITING_{ik}}{NCITING_i}\right)$ is the percentage of citations received by patent *i* that belong to patent class *k,* out of N_i patent classes. The N_i patent class is based on a 3-digit patent class, consists of 417 classifications.[26]

The Generality index takes values between zero and one, where low values represent high concentration and consequently low patent Generality. If many subsequent patents that belong to the same technological class as the cited patent cite the patent, the Generality measure will be low. Conversely, if many patents from a wide range of fields cite the patent the Generality index will be high.

The second 'value' measure is the Originality Index, which is the extent to which an innovation is broad in its scope in the sense that the innovation is based on knowledge coming from innovations from a wide range of technological fields. The Originality index is by the Henderson et al. (1997) B/Originality computation. The Originality index is a Herfindahl index of concentration. It is similar in computation to the Generality measure but is based on the number of class a patent cites rather than receives. Therefore, $0 \le ORIGINAL_i \ge 1$, where higher values of originality imply broader innovations as they cite many previous innovations from a wide range of technological fields.

7.4.1.1 Adjusting for a Bias in the Indexes

The computation of the Generality and Originality indexes are based on the concentration of citations made or received by patents. Hall (2000) highlights the possibility of a bias in these two indexes due to the way the data are counted. The

[26] This is based on the 1999 patent classification.

concern arises because patents with zero or one citations, although having a non-zero probability of receiving citations in subsequent years, are removed from the analysis. Since patents may have a higher or lower tendency to cite or to be cited simply because of the cohort to which they belong, the author shows that if the citations follow a multinomial distribution, the ‘concentration’ measures would be biased upwards, leading to lower estimations of the ‘value’ of the patented innovations. Hall (2000) suggests a computation that leads to an unbiased estimation. I follow the author’s methodology and adjust the value of Originality and Generality for the possibility of a bias.

7.4.2 Estimation

I estimate two value adjusted OLS regressions, see Table 7.[27] The two are similar in structure to Trajtenberg’s (2001) OLS regression but have Generality and Originality indexes as the dependent variables. The country dummies then represent the average expected Generality or Originality of patented innovations originated in those countries and are contrasted to the value predicted by the simple citations regression.

The results indicate that regressing on a simple citation aggregation could be somewhat misleading in predicting the aggregate values of patented innovations. The Generality regression shows that US patented innovations do not appear to be any better than Israeli innovations as the US dummy is statistically insignificant. However, the Originality regression maintains the US advantage over the Israeli innovations. The US dummy is three basis points above the Israeli dummy, implying that US patent innovations are eight percent more Original than Israeli patented innovations. In the case of Israel vis-à-vis the Reference country, the Generality and Originality

[27] The year dummies are included but are not reported.

estimations confirm Trajtenberg results. Israeli patents are superior to Reference country patents.

The finding that citation-weighted value indexes may provide different predictions on the value of patented innovation imply that regressing on citation frequencies requires careful examination for the consistency and robustness of the results. Simply regressing on citation count may not provide the most accurate predictions.

Table 7 Two 'value' weighted OLS regressions

	Generality	Originality
US	0.013	0.03
	-1.89	(4.98)**
Reference	-0.032	-0.034
	(4.26)**	(5.31)**
chemical	0.103	0.104
	(15.26)**	(16.34)**
cmpcmm	0.097	0.089
	(11.04)**	(11.06)**
drgsmed	-0.024	0.037
	(2.60)**	(4.50)**
elec	0.052	0.045
	(7.25)**	(6.68)**
mech	0.027	0.015
	(4.15)**	(2.42)*
Constant	0.467	0.385
	(64.96)**	(61.54)**
Observations	25278	25723
R-squared	0.02	0.03

7.5 *Summary*

The chapter discussed and tested four factors that appeared to have been overlooked in Trajtenberg's original work. The results indicate that the original estimations are sensitive to three factors: structural breaks, structure of the OLS model and noise in citation frequencies. The most important finding is that breaks exist in the Hall et al. (2001) citation data. Furthermore, these breaks significantly change the estimated results. A failure to address this issue in applied work may lead to a loss of significant information and robustness.

8 Long Run Trends in the Value of Innovations

This Chapter extends the discussion to the literature on patent citations as a proxy for the geography of knowledge spillovers. It begins by providing the conceptual, theoretical and empirical background for economic inquiry into the geographic flow of knowledge. It then takes advantage of a cointegration test and Graphical Modelling and shows a clear single trend and strong association in the value of patented innovations originating in the G-5 countries over time.

8.1 *Overview*

So far, this thesis discussed and explored the literature on patent citations as a proxy for measuring the ex-post value of patented innovations. This chapter extends this discussion to the literature on patent citations as a proxy for knowledge spillovers. This emerging body of work explores the usefulness of patent citations to trace the flow of knowledge across institutions and organisations, countries and time. In the presentation and discussion of this topic, I shall closely follows texts by Branstetter (1998), Griliches (1979, 1990, 1992), Jaffe and Trajtenberg (2002) and Romer (1994).

The concept of knowledge spillovers is intimately tied to the nature of knowledge assets. Knowledge is nonrivalrous and only partially excludable (Jones 2002). These attributes imply that the inventors of new knowledge can only partially appropriate and capture the entire economic value of the knowledge embodied in their inventions. The available legal mechanisms that are designed to protect the property rights of the inventors (copyrights and patents are common examples) are imperfect. Ultimately, some knowledge inevitably spills out, leading to a knowledge externality.

8.2 *Why Do Economists Study Knowledge Spillovers?*

The theoretical foundations to the economic inquiry into the geographic flow of knowledge spillovers can be seen in the work of Grossman and Helpman (1990, 1991, 1995). The authors present spillover-based endogenous growth models to explain the rate of economic trade and growth. A key novelty of these models is their explicit treatment of the idea that inventors, who may earn monopoly rents for their inventive output, discover inventions endogenously (Romer, 1994). This idea is not accommodated in the neoclassical growth literature, which typically treats knowledge as a public good (Romer 1994). This is one of the reasons the Grossman and Helpman type theoretical models are sometimes referred to as Schumpeterian growth models. They emphasise the prediction of Schumpeter in the late 1930s that temporary monopoly returns on inventions are required to ignite the innovative process, which determines the rate of growth. More specifically, the outcome of the inventive activity yields benefits both to the inventors in the form of returns on investment and to society by adding new knowledge into the aggregate spillover pool. As the inventive activity increases, more knowledge is created and spills over into the aggregate knowledge pool. This reduces the investment costs for inventing new knowledge and prohibits diminishing returns from arising (Branstetter 1998).

The economic inquiry into the geographic patterns knowledge spills is because Grossman and Helpman models may yield different predictions depending on the assumptions surrounding the spillover flow (Branstetter 1998). If knowledge diffuses beyond borders (internationally), then the normal competitive advantage opportunities dictate trading and the corresponding growth rates. However, if knowledge diffuses locally (intranationally), than an economy, which might only have a slight technological advantage over the other economies, would eventually dominate the world production (Branstetter 1998). The authors provide the following example:

> Suppose it is country A that begins with more research experience. Then initially this country's researchers have a competitive advantage in the research lab, and they perform all of the world's R&D at time 0. But then additional knowledge accumulates in country A, while in the absence of international knowledge spillovers, the knowledge stock remains fixed in country B. So, country A's competitive lead in R&D widens and there is even greater reason for this country to conduct all the world's research in the next period. The initial lead is selfreinforcing and eventually country A comes to dominate production in the high-technology sector. (Grossman and Helpman 1995, Chapter 2)[28]

Why might knowledge spillovers exhibit intranationally rather than international tendencies? Bransteter (1998) suggests three key answers to this question:

1. Interaction - inventors might find it easier to communicate their ideas if they are geographically proximate.
2. Language - inventors might find it easier to transfer knowledge when they communicate in the same language.
3. Transactional costs – fewer regulatory barriers when knowledge is transferred intarnationally rather than internationally.

8.3 *Empirical Findings*

The theoretical implications of trade and growth type theories discussed above necessitated empirical investigations into identifying whether and to what extent do geographic knowledge spillovers exist. Before discussing this body of work in depth, it is important to understand what economists regard as knowledge spillover.

---

[28] This quote also appears in Branstetter, 1998.

Strictly speaking, a knowledge spillover occurs when:

> firm A is able to derive economic benefit from R&D activity undertaken by firm B without sharing in the cost firm B incurred in undertaking its R&D (Branstetter 1998, p. 495).

This implies that knowledge that is bought out by other inventors or firms merely represents a knowledge transfer rather than a knowledge spillover. Griliches (1979) is a bit more specific and distinguishes between two different types of knowledge spillovers:

1. Pecuniary (embodies) spillover, which occurs when the inventors cannot appropriate all the surplus of their invention. This type of externality could occur when the inventors cannot perfectly price discriminate (Branstetter 1998).
2. Nonpecuniary (disembodied) spillover, which is the impact of new knowledge on the discoveries of new inventions.

It is the second type, Nonpecuniary (disembodied) spillover, that is the knowledge spillover discussed in the trade and endogenous growth models and has preoccupied the applied microeconomics literature (Branstetter 1998). This type of spillover corresponds precisely to the nonrivalry and low excludability attributes, which make them extremely difficult to trace and even harder to quantify. This research frustration can be seen in Paul Krugman's pessimistic statement that "knowledge flows … are invisible [as] they leave no paper trail by which they may be measured or tracked" (Krugman 1991, p. 53).

Most existing empirical work only indirectly tests for geographic knowledge spillover. The common approach is to use some form of production function, with either aggregated industry or country data, to represent a single profit maximising firm. The production function typically has the normal ingredients plus a knowledge asset on the right hand side. Examples of influential papers of this kind include Coe and Helpman (1995) and Coe et al. (1995). Although their results indicate significant international spillovers, the production function approach imposes numerous assumptions, whose validity has been highly challenged in later work, predominantly by the scholars in the micro-productivity literature; see Branstetter (1998). Griliches (1992) suggests a cost function approach as an alternative, although this too suffers from considerable pitfalls due to the assumed underpinning of the structural models. A more micro approach for estimating knowledge spillovers was proposed by Griliches (1979) and tested in Jaffes' 1983 thesis. The approach uses common technological groups of firms, clustered by the patent type of their granted patented inventions and tests whether the firms' activity is correlated to the overall cluster to which it belongs. The findings show significant correlation between the aggregate knowledge pool and the firm data.[29]

Broadly speaking, the literature discussed above provides evidence in favour of knowledge spillovers as a gap between the private and social rate of return to knowledge is identified. However, this literature does not explain the mechanism for the transmission of the spillovers. More recently, Jaffe, Trajtenberg and colleagues have used patent citations data to identity and measure the direct flow of knowledge, where citations represent a 'paper trail' of codified technological knowledge spillovers. Their survey of the use of US inventors on the citations of patents revealed that "aggregate citations flows can be used as [direct] proxies for knowledge spillover intensity…between categories of organizations or between countries (Jaffe et al. 2000, p. 218). Caballero and Jaffe (1993) were the first to lay the methodological foundations to this research. In the context of a general equilibrium model, the authors developed a 'citations function' that captures the patent citation process in terms of both knowledge

[29] The reader can refer to Griliches (1979, 1990) for further discussion on this approach.

obsolescence and knowledge diffusion as time elapses. Their findings show a clear decline in the spillover potency across time as the knowledge obsolescence rate is shown to be endogenously determined by the number of ideas rather than exogenously by time.

With Caballero and Jaffe's (1993) work setting the scene, later work explored and specifically linked patent citations to the geography of knowledge spillover. Jaffe et al. (1993) tested for the geography of spillovers by examining the relationship between the location of the citing and cited patent, relative to an expected relationship due to the technological activity in the regions. That is, the expected likelihood given the existing concentration of technological activity. The authors were the first to show a significant localisation in the geography of patent citations (Jaffe and Trajtenberg 2002). Patents tend to cite more heavily patents that originate in the region than would be expected given the distribution inventive resources in those regions. Thus, technological knowledge is utilised more readily where it originates (Branstetter 1998).

As the NBER U.S. Patent Citations Data project (Hall et al. 2001) evolved, providing richer data, scholars were able to further examine and more carefully quantify the patterns the spillover flow identified in Jaffe et al. (1993). Jaffe and Trajtenberg (1996) used a nonlinear citation function to predict the nationality of inventors. They showed "that there is a clear time path to the diffusion of knowledge, in which domestic inventors' citation probabilities are particularly high in the early years after an invention is made" (Jaffe and Trajtenberg 1999, p. 106). Jaffe and Trajtenberg (1999) went further and explored the citations patterns of inventors from the G-5 countries, while accounting for, *inter alia*, changes in the citation tendencies and the truncation bias. Their results showed that inventors of the same country are 30 to 80 percent more likely to site each other.

The citation approach for tracing knowledge spillovers is still evolving and continuously challenged (see the two 2005 AER comments by Thomson and Fox-Kean and Jaffe et al.). Yet, two clear messages appear to have emerged:

1. Knowledge spillovers tend to show a strong 'intranational' tendency.
2. As time goes by, this intranational fades away towards an international tendency.

Jaffe and Trajtenberg (2002) explain the intuition behind these findings concisely:

…whatever initial advantage geographic proximity may offer in terms of knowledge transmission and as stimuli for further knowledge creation, the very 'ethereal' nature of knowledge dictates that such advantage should diminish with time (Jaffe and Trajtenberg, 2002, p. 12).

8.4 *Building on the Literature*

The implications of the findings of the geographic spillover citation-based literature imply that the diffusion of knowledge is a lengthy process and if it exists, ought to appear in long-run information. Based on this observation, the question I want to ask in this chapter is whether a statistical relationship can be identified in the value of knowledge embodied in patented inventions originating in the G-5 countries in the long run? This question should not be thought of as a question on the occurrence of a knowledge spillover, but as simply an inquiry into the statistical relationship in the value of inventive output across the G-5 countries over the long horizon. This relationship, if found, could be argued to be a necessary condition for the existence of technological knowledge spillovers, but not sufficient.

Why might we expect to identify a long run relationship in the value of knowledge embodied in patented inventions across the G-5 countries? To answer this question, consider a quality increase in the technological inventive output of one of the countries. In the presence of knowledge spillovers, foreign inventors, working on related technological inventions, can benefit from this new technological knowledge for their own research and discoveries. If they are advantaged by gaining access to this knowledge, we would expect this new knowledge to eventually stimulate and enhance the quality of their own inventive activity and output.

The theoretical validity to the reasoning above is implicitly found in endogenous growth theory. Consider the Romer (1990) model (Chapter 2). The model asserts that the advanced economies in the world form one integrated gigantic economy, whose economic performance is determined, *inter alia*, by their access to knowledge from the common aggregate knowledge pool, represented by the ϕ parameter (Jones 2002, see Chapter 2). The discovery of better knowledge in one economy raises the knowledge pool and pushes forward the 'world' technological frontier and consequently the 'world' economy growth rate.

Patent citations can be used as a proxy for the value of patented innovations originating in the G-5 countries. I follow the maintained assumption in the patent citation value literature (see Chapter 4) that "patents are a proxy for 'bits of knowledge' and patent citations are a proxy for a given bit of knowledge being useful…" (Jaffe and Trajtenberg 1999, p. 108). By treating patent citations as a proxy for the value of codified technological knowledge embodied in patents, the averaged aggregate patent citations flow of patents originating in the G-5 economies represent a series of innovative output value data.

I propose that the long run relationship in the value of patented innovations originating in the G-5 countries should be tested through two distinctively different set of tests that complement each other. The first is a time-series cointegration test to identify whether a stochastic common long run quality trend relationship can be identified in the quality of patented innovations series of the G-5 countries. With n most advanced economies in the world, n nonstationary (innovative output quality) series, and $n-k$ significant cointegrating vectors, there will be k common stochastic (value)trends (Greasley and Oxley, 2000).

The second test is a multivariate-based Graphical Modelling analysis that uses graphs to identify the conditional association in complicated statistical series. If the cointegration test identifies common stochastic trends in the data, Graphical Modelling can further represent the nature of these statistical relationships. This is the topic of the following section.

8.4.1 Data and Estimation

I use the patents and corresponding citations of the G-5 countries, US, Japan, Germany France and UK for the 1965 to 1995 period. I then calculate the weekly average citations per granted patents each week for each of the five countries. Each average is assumed to represent the observed weekly quality of a country's innovative output. This provides five variables, US, Japan, Germany France and Britain, each containing more than 1,600 observations.

I begin the analysis with a cointegration test to identify common trends in the value of a country's innovative output. The advantage of following such a rigorous process is the ability to examine short-term dynamics without losing "long-run information" (Engle and Granger, 1987).[30]

The cointegration test is only appropriate when the variables of interest are nonstationary and are integrated of the same order, I(d). I use the Augmented Dickey-Fuller (ADF) test of unit root to determine the stationarity of the quality variables.

8.4.1.1 Test for Stationarity

Consider a simple AR(1) process: $Y_t = \rho Y_{t-1} + \varepsilon_t$, where Y_t is the variable of interest and ε_t is a white-noise process. If ρ equals to one, the variance of the series approaches infinity and Y_t is said to be a nonstationary process.

The ADF test (Dickey and Fuller, 1981) is carried out by estimating $\Delta Y_t = \beta_0 + \alpha Y_{t-1} + \gamma t + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + \varepsilon_t$, where ΔY is a change in the variable of interest, $\alpha = \rho - 1$ and ε_t is an independently and identically distributed white-noise process. If the estimated coefficient of the lagged variable, α, is significantly less than zero, the null hypothesis of a unit root can be rejected and Y_t is said to be a stationary process.

When the series are difference stationary, stationarity is induced by differencing the variable of interest. A variable that needs to be differenced (d) times to reach stationarity is regarded as a I(d) process or as integrated of order (d).

[30] The methodological explanation in this section follows Lee (1997); Maddala (1992) and Maddalla and Kim (2000) and the Eview manual.

For all the five quality series, the unit root test fail to reject the null of hypothesis of nonstationary in the levels of the variables, but is rejected when the five variables are first differenced, Table 8. The five countries have a nonstationary, I(1), innovation quality series.

Table 8 Unit Root Tests (Augmented Dickey-Fuller Statistics) for the 1965 to 1995 period

	Level	1st Difference
Germany	-0.75	-18.67*
	0.83	0.00
US	-0.97	-24.21*
	0.76	0.00
France	-2.22	-20.83*
	0.20	0.00
Japan	-1.49	-21.06*
	0.54	0.00
Britain	-1.54	-17.28*
	0.51	0.00

The first number is the t-Statistics and the number below is the corresponding P-Values. The lag length was chosen on the basis of Akaike's Information Criteria. * denotes rejection of the hypothesis at the 0.05 level.

Engle and Granger (1987) show that nonstationary variables might share a long-term equilibrium relationship. Consider the following example: if two series Y_t, X_t are both I(d) and there exists a linear combination e.g., $Y_t - \beta X_t$ which is stationary, then the variables, Y_t and X_t, do not drift a way from each other and are described as being cointegrated with a cointegrating parameter β. That is, the variables progress through time in a similar pattern and there exists a long-run equilibrium between the two variables. The test used here to determine whether the group of five I(1) variables are cointegrated or not is based on Johansen's (1988) maximum likelihoods technique.

8.4.1.2 Testing for Cointegration

Johansen's method involves the estimation of the following k^{th} order Vector Autoregressive (VAR) equation: $y_t = A_0 + A_1 y_{t-1} + ... + A_k y_{t-k} + \beta x_t + \varepsilon_t$ where

y_t = a k-vector of nonstationary I(1) variables

A_0 = an $(n \times 1)$ vector of intercept terms

A_i = $(n \times n)$ matrices of coefficients

ε_t = the $(n \times 1)$ vector of error term

This VAR could also be written in terms of the following Vector Error Correcting Mechanism (VECM): $\Delta y_t = A_0 + \pi y_{t-k} + \sum_{i=1}^{k-1} \beta_i \Delta y_{t-i} + \varepsilon_t$ where $\pi = -(I - \sum_{i=1}^{k-1} A_i)$ and $\beta_i = -(I - \sum_{i=1}^{k-1} A_j)$. The rank of the matrix π represents the number of cointegrating vectors r among the variables in the vector y. Subtracting the number of cointegrating vectors r from the number of variables in the vector y yields the number of long run stochastic trends (Stock and Watson 1988). Johansen's method employs two likelihood-ratio test statistics to test the null hypothesis of at most r cointegrating vectors in y_t. The first one, $\lambda(r)_{trace} = -\mathrm{T} \sum_{j=r+1}^{n} \ln(1-\varphi_j)$, is a λ_{trace} statistics, which reports the H_0 of at most r cointegrating vectors against the H_1 of more than r. The second test $\lambda_{\max}(r, r+1) = -\mathrm{T} \ln(1-\varphi_{r+1})$ is a $\lambda_{\max}$ criteria, which is similar to the null of λ_{trace}, but its alternative is $r+1$ cointegrating relationship.

Table 9 shows that both the Max criteria and the Trace statistics reject the three or fewer cointegrating vectors in favour of four, pointing to the existence of one single common stochastic trend for the five innovation quality series. This implies that all the five series follow one stochastic knowledge frontier. The issue now is whether some of the G-5 countries play a leading role in the stochastic trend identified above. This could be revealed through a test of statistical association, where the associated links tested concern the interrelationship surrounding the fluctuations between the five series. I test for association with a Graphical Modelling (GM) test. The advantage of implementing Graphical Modelling on the series is twofold. Firstly, if patterns of association emerge between the countries' value series then we would have an insight into the nature of the fluctuations of common stochastic trend. Furthermore, if association is not picked up by the Graphical Modelling analysis, then it would be taken as seriously casting doubt on the validity the cointegration results.

Table 9 Johansen Cointegration Test Results

H(0)	H(1)	Maximal Eig	Trace
r = 0	r = 1	0.17*	860.39*
r ≤ 1	r = 2	0.16*	557.38*
r ≤ 2	r = 3	0.13*	257.11*
r ≤ 3	r = 4	0.03*	55.49*
r ≤ 4	r = 5	0	1.86

r = the number of cointegrating vectors.
The lag length was chosen on the basis of Swartz's Information Criteria.
* rejection of the null at the 5% level.

8.4.1.3 Testing for Association

Graphical Modelling is a multivariate-based analysis that uses graphs to translate the relationship of complicated systems into statistical meaning (Edwards 1995; Jordan 1999). The issue of association pursued here uses graphs to identify the statistical relationship across all five innovative output quality variables.

Two different graphical representations are used in practice, directed and undirected graphs, which vary in the rules that are applied to read the graphs. I use an undirected graph to estimate the conditional dependence relationship between the quality series. In introducing GM theory and presenting the basic definitions of conditional independence, I closely follow the texts by Edwards (1995) and Wasserman (2003).

8.4.1.3.1 Conditional Independence

The conditional probability of event A, given event B, is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad (1)$$

When multiplying both sides of (1) by $P(A)$ the Multiplicative Rule of Probability is obtained:

$$P(A \cap B) = P(A|B)P(B) \qquad (2)$$

When event A and event B are independent events, then $P(A|B) = P(A)$ and, therefore, (2) becomes:

$$P(A \cap B) = P(A)P(B) \qquad (3)$$

Equation (3) can also be represented mathematically as $A \perp B$.

Extending the discussion to three random variables, we can say that the random events A and B are *conditionally independent* of C when the information contained in B does not provide further information about A once C is known.

Mathematically it can be represented as

$$A \perp B|C \quad \text{or} \quad B \perp A|C \qquad (4)$$

8.4.1.3.2 Undirected Graphical Model

Graphical models provide a visual representation for the $A \perp B|C$ relationship.

An undirected graph is a structure of $G = (V, E)$, where the set of V are vertices (nodes) and the set of E are edges (lines) which connect these vertices. The variables X and Y are *adjacent*, $X \sim Y$, if there is an edge connecting them. A *path* is a sequence of $X_0,...,X_n$ if $X_{i-1} \sim X_1$ for every i. A *complete graph* is a graph where every pair of its vertices is connected by an edge. The pairwise Markov property holds that any pair of random variables that are non adjacent are conditionally independent given all other variables in the model. Identification of conditional independence among the variables in the model can simply be found by observing the separated set of variables. If A and B are separated by a set C, then $A \perp B \mid C$, which is the *global Markov property.*

Consider Figure 7, which represents the four variables model, A, B, C and D. Visual examination shows the non-existence of edges between the vertices A and D and B and D and therefore, variable A is conditionally independent of D given C, $A \perp D \mid C$ and that B is conditionally independent of D given C, $B \perp D \mid C$.

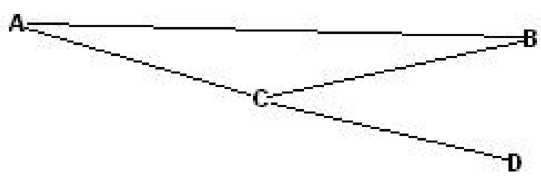

Figure 7 Graphical Modelling: an example

8.4.1.3.3 Results and Summary

I use the *mimR package* in the R statistical computation software to construct an undirected graphical model, Figure 8. The result is a complete graph as there is an edge between every pair of vertices. The graph indicates that all the five value series are conditionally dependent. None of the five quality variables can be analysed in isolation form the remaining set, which implies that dominating countries do not exist. This reinforces the result of the cointegration test of a common stochastic trend shared across all the five patented innovations value series. Each of the five economies appears to follow a common technology frontier trend generated by this ‘whole’ “global” economy. This result strongly supports the implied assumption in Romer (1990) and the observation of international long run tendencies in the spillovers of technological knowledge in the patent citation literature.

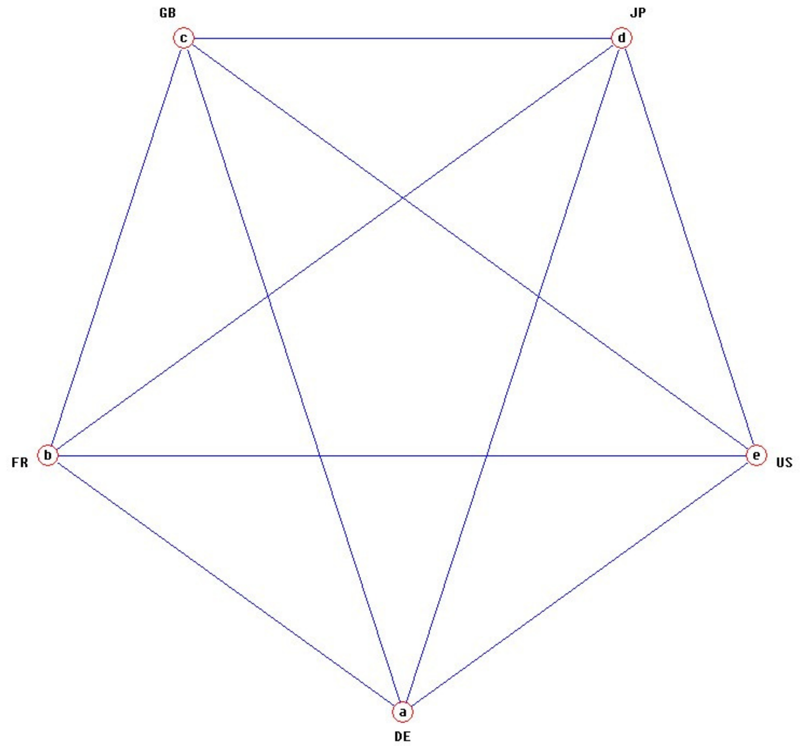

Figure 8 Graphical Modelling for the time-series quality US, French, British, Japanese and German averaged aggregate citations flow.

9 Conclusion

The objective of this thesis is to advance our understanding of the value of innovations using information contained in patent citations. This chapter synthesises the findings of this thesis.

In Chapter 2, I began by reviewing the reasons economists are intrigued by innovative activity. The literature indicates that the creation of innovative output is a key component in the growth process of advanced economies. Concepts such as technological ideas, R&D and scientific discoveries are found at the heart of modern macro and micro economic inquiry into the determinants of economic prosperity and social well-being.

In Chapter 3, I discussed the difficulties in finding empirical counterparts for the dynamic process of innovative activity and supported the use of patent statistics as an imperfect measure of innovative output. Patents are objective, based on a voluntary economic system, and contain highly detailed information about each granted innovation. Upon a comprehensive analysis of the methods and empirical proxies for the elicitation of the value embodied in patents, patent citations, a prior-to-new-art link, appeared as the most objective and systemic indicator for the ex-post technological value patents represent.

In Chapter 6, I developed a framework for the analytical examination of patent values based on Trajtenberg's (2001) approach. This approach yields a clear measure of the average value of patented innovations. However, upon investigation in Chapter 7, I found that this approach is somewhat simplistic and requires further robustness tests. I emphasised four econometric and theoretical factors that are of importance in carrying this empirical analysis: the statistical properties of the dependent variable, the structure of the underlying model, breakdates in the data and noise in citations counts. The results indicated that three out of the four factors, if overlooked, could lead to non-robust econometric estimations; these are the structure of the underlying model, breakdates in the data and noise in citations counts.

Upon reviewing the properties of citation arrival and the literature of the US Patent Office, I proposed that structural breaks that may impair the robustness of the econometric estimations would exist. Atheoretical Regression Trees revealed that two break and three sub-periods exist in the data. Adjusting the estimations to three individual periods confirmed a strong sensitivity of the results to period examined. Furthermore, the structural form of the model proposed in Trajtenberg (2001) was then tested and found to be sensitive. The aggregation of patented innovation into one big control group leads to a significant loss of information, and thus to a loss of accuracy of the results.

From the tests of robustness, it was also apparent that patent citations are a noisy signal of technological value. I identified an inconsistency in Trajtenberg's (2001) results of technological significance compared to the results obtained with weighted indexes of patent value compiled by patent citations. The Herfindahl index of concentration Generality regression showed, contrary to Trajtenberg (2001) estimates, that US patented innovations do not appear to be any better than Israeli patented innovations as the US dummy is statistically insignificant.

A further valuable contribution of this thesis has been the integration of predictions by endogenous growth theory, the applied knowledge spillover literature and patent citations to test for geographic long run trends and association in the value of technological knowledge across time. The geographic knowledge spillover literature shows that knowledge spillovers tend to show a strong 'intranational' tendency in the flow of knowledge. However, as time goes by, this 'intranationality' fades away towards an international tendency. Based on this observation, I ask whether a statistical relationship can be identified in the value of knowledge embodied in patented inventions originating in the G-5 countries in the long run. To answer this question I use patent citations in the context of a cointegration test to identify whether G-5 innovative output fluctuates and moves in a similar pattern over the long time horizon. The results show that there is a single long run stochastic trend in the quality of inventive output between the five most advanced economies in the world. Furthermore, applying a multivariate Graphical Modelling association estimation, I show that countries' quality series are completely associated and domination does not appear to exist. This finding supports the implicit assumption in the endogenous growth literature of a technological knowledge frontier common to the advanced countries in the world.

Overall, this thesis provides an original contribution to the economics of innovation literature in a number of areas:

i. A comprehensive description of the various proxies available to elicit the value embodied in patented innovations.
ii. A direct comparison of regression results of the value of patent innovations based on citations count as the dependent variable versus citations-weighted as the dependent variable

iii. An introduction and application of Regression Tree and Graphical Modelling tests to advance empirical analysis of the value of patented innovations.

iv. Estimation of the fluctuations and associations of the values of patented innovation originating in the G-5 countries.

This thesis will provide a valuable reference both for researchers in the area of economic of innovation, and for policy analysts in developing and analysing innovation, technology and science and R&D based policies.

10 References

Aghion, P., and Howitt, P., 1992. “A Model of Growth through Creative Destruction.” *Econometrica* **60**, pp. 323-351

Aghion, P., and Howitt, P., 1998. “Endogenous Growth Theory.” The MIT Press, Cambridge

Albert. M. B., Avery, D., Narin, F. and McAllister, P., 1991. “Direct validation of Citation Counts as Indicators of Industrially Important Patents.” Research Policy **20,** pp. 251-259

Barro, R.J., and Sala-i-Martin, X., 1992. “Convergence.” *Journal of Political Economy* **100**, pp. 223-251.

Barsberg, L. B., 1987. “Patents and the Measurement of Technological Change: A Survey of the Literature.” Research Policy **16**, pp. 131-141.

Bertran, J. F., 2004. “Patents, Citations and the Market Value of Patents.” PHD Thesis. University of Rochester, New York

Black, J., 2002. “A dictionary of Economics.” Oxford, Oxford University Press.

Boldrin, M., Levine, D., 2002. “Perfectly Competitive Innovation.” Staff Report 303. Federal Reserve Bank of Minneapolis.

Bound, J., Cummins, Z., Griliches, Z., Hall, B., and Jaffe, A., 1984 “Who Does R&D and Who Patents.” In Griliches, Z., ed. (1984). R&D, Patents and Productivity. University of Chicago Press.

Branstetter, G. L., 1998. “Looking for International Knowledge Spillovers; A Review of the Literature with Suggestions for New Approaches.” in Encaoua et al. (eds.), *The Economics and Econometrics of Innovation*. Kluwer Academic Publishers, Netherlands pp. 495-518.

Breiman, L., Friedman, J., Olshen, R. and Stone, C., 1984. “Classification and Regression Trees.” Chapman & Hall, New York

Caballero R. J., and Jaffe., A., 1993. “How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth,” in *NBER Macroeconomics Annual 1993*, Cambridge: The MIT Press, pp. 15-86.

Cameron, C., and Johansson, P. (1997) “Count Data Regression using Series Expansions with Applications”, *Journal of Applied Econometrics* **12,** pp. 203-223

Cameron, A. C., and Trivedi, P. K., 1986. “Econometric models based on count data: comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* **1,** pp. 29-54.

Cameron, A. C., and Trivedi, P. K., 1998. “Regression Analysis of Count Data.” Cambridge: Cambridge University Press.

Campbell, R. S., and Nieves, A. L., 1979. “Technology Indicators Based on Patent Data: The Case of Catalytic Converters – Phase 1 Report: Design and Demonstration.” Battelle, Pacific Northwest Laboratories.

Cappelli, C., and Reale, M., 2005. “Atheoretical Regression Trees for Detecting Breaks in Lake Mean Water Levels: an Application to Lake Michigan-Huron data.” Working Paper

Carpenter, P. M., Narin, F., and Woolf, P., 1981. “Citations Rates to Technologically Important Patents.” *World Patent Information* **3(4),** pp. 160-163.

Cockburn, I., and Griliches, Z., 1988. “Industry Effects and Appropriability Measures in the Stock Market’s Valuation of R&D and Patents.” *American Economic Review* **78**, pp. 419-423.

Coe, D., and Helpman, E., 1995. “International R&D Spillovers”, *European Economic Review,* **39(5)**

Coe, D., Helpman, E., and Hoffmaister, A. 1995. “North-South R&D Spillovers”, *NBER Working paper* 5048.

Dernburg, T. and Gharrity, N., 1961. “A Statistical Analysis of Patent Renewal Data for Three Countries.” *Patent, Trademark and Copyright Journal* **5,** pp. 340-361

Dickey , D. A., and Fuller, W. A., 1981. “Likelihood ratio Tests for Autoregressive Time Series with a Unit Root.” *Econometrica* **49**, pp. 1057-1072.

Duguet, E., and Kabla, I., 2002. “Appropriation Strategy and the Motivations to Use the Patent System: an Econometric Analysis at the Firm Level in French Manufacturing, in D. Encaoua et al. (ed.), The economics and Econometrics of Innovation, Kluwer Academic Publishers, Boston

Edquist, C., 1997. “Systems of Innovation: Technologies, Institutions, and Organizations” Pinter, London

Edwards, D., 1995. “Introduction to Graphical Modelling.” Springer-Verlag, New York

Ellis, P., Hepburn, G. and Oppenheim, C., 1978. “Studies on Patent Citation Networks.” *Journal of Documentation* **34(1),** pp. 12-20

Engle, R. F., and Granger, C. W. J., 1987. “Co-integration and Error Correction: Representation, Estimation and Testing.” *Econometrica* **55**, pp. 251-276

Freeman, C., 1974. “The Economics of Industrial Innovation.” 2nd ed London, Pinter

Freeman, C. 1994. “The Economics of Technological Change.” *Cambridge Journal of Economics* **18(5)**, pp. 463-514

Gay, C., Le Bas, C., Patel, P., and Touch, K., 2005. "The Determinants of Patent Citations: an Empirical Analysis of French and British Patents in the US." *Economics of Innovation and New Technology* **14(5),** pp. 339-350

Giuri, P., and Mariani, M., 2005. "Everything You Always Wanted to Know About Inventors (But Never Asked): Evidence from the PatVal-Eu Survey." *Laboratory of Economics and Management*. Working Paper.

Glaeser, E., Kallal, H. D., Scheinkman J. A., and Shleifer, A., 1992. "Growth in Cities". *Journal of Political Economies* **100(6),** pp. 1126-1152

Gourieroux, C., Monfort, A., and Trognon, A., 1984. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica* **52,** pp. 701-720

Granger, C. W. J., 1969. "Investigating causal relations by econometric models and cross-spectral methods." *Econometrica* **37(3),** pp. 424-438

Greasley, D., and Oxley, L., 2000. "British Industrialization 1815-1860: A Disaggregate Time Series Perspective." *Explorations in Economic History* **37**, pp. 98-119.

Grilliches, Z., 1979. "Issues in Assessing the Contribution of R&D to Productivity Growth." *Bell Journal of Economics* **10(1),** pp. 92-116

Griliches, Z., 1981. "Market Value, R&D, and Patents." *Economic Letters* **7**, pp. 183-187

Griliches, Z., 1984. "R&D, Patents and Productivity." (ed.) University of Chicago Press.

Griliches, 1990. "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature* XXVIII, pp. 1661-1707.

Griliches, Z., 1992. "The Search for R&D Spillovers." *Scandinavian journal of economics* **94,** pp. 29-47.

Grossman, G., and Helpman, E., 1990. "Comparative Advantage and Long-Run Growth", *American Economic Review,* **80(4)** pp. 796-815

Grossman, G., and Helpman, E., 1991. "Innovation and Growth in the Global Economy." MIT Press, Cambridge.

Grossman, G., and Helpman, E., 1995. "Technology and Trade", in *Handbook of International Economics,* vol. 3, edited by Gene Grossman and Kenneth Rogoff.

Guellec, D., and Potterie, P. B., 2000. "Applications, Grants and the Value of Patent." *Economic Letters* **69(1)**, pp. 109-114

Guellec, D., and Potterie, P. B., 2002. “The Value of Patents and Patenting Strategies: Countries and Technology Areas Patterns” *Economics of Innovation and New Technology* **11(2),** pp. 133-148

Hall, B., 1998. “Innovation and Market Value,” In Barrell, Ray, Geoffrey Mason, and Mary O'Mahoney (eds.), *Productivity, Innovation and Economic Performance*, Cambridge, Cambridge University Press, 2000.

Hall, B., 2000. “A Note on the Bias in the Herfindahl Based on Count Data.” *Revue d'Economie Industrielle* **110**, pp. 149-156

Hall, B., Griliches, Z., and Hausman, J., 1986. “Patents and R&D: Is There a Lag?” *International Economic Review* **27(2),** pp. 265-283

Hall, B. H., Jaffe, A., and Trajtenberg, M., 2001. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." *NBER Working Paper 8498*

Hall, B., Jaffe, A., and Trajtenberg, M., 2005. “Market Value and Patent Citations” *Rand Journal of Economics* **36,** pp. 16-38

Hall, B., and Trajtenberg, M., 2005. "Uncovering General Purpose Technologies using Patent Data.” In Antonelli, C., D. Foray, B. H. Hall, and W. E. Steinmueller (eds.), New Frontiers in the Economics of Innovation and New Technology. Edward Elgar.

Hall, B., and Ziedonis, H. R., 2001. “The Patent Paradox Revisited: an Empirical Study of Patenting in the US Semiconductor Industry 1979-1995.” *RAND Journal of Economics* **32(1),** pp. 101-128

Hansen, E. B., 2001. “The New Econometrics of Structural Change: Dating Breaks in US Labor Productivity.” *Journal of Economic Perspectives* **15(4),** pp. 117-128

Hansen, J., 1992 “Innovation, Firm Size, and Firm age.” *Small Business Economics* **4(1)**, pp. 37-44

Harnoff, D., Narin, F., Scherer, F. M., and Vopel, K., 1999. “Citation Frequency and the Value of Patented Inventions”. *The Review of Economics and Statistics* **81(3)**, pp. 511-515

Harhoff, D., Scherer F. M., and Vopel, K., 2003a. “Exploring the Tail of Patented Invention Value Disturbances” in *Economics, Law and Intellectual Property* edited by Ove Granstrand (2003) Kluwer Academic Publishers Boston.

Harhoff, D., Scherer F. M., and Vopel, K., 2003b. “Citations, Family Size, Opposition and the Value of Patent Rights.” Research Policy **32**, pp. 1343-1363

Hausman, J., Hall, B., and Griliches, Z., 1984. “Econometric Models for Count Data with an Application to the Patents-R&D Relationship.” *Econometrica* **52**, pp. 909-937

Heckman, J., 2006. “Contributions of Zvi Griliches.” *NBER Working Paper 12318*

Henderson, R., Jaffe A., and Trajtenberg, M., 1999. "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting 1965-1988." *Review of Economics and Statistics* LXXX **(1)**, pp. 119-127

Højsgaard, S., 2004. The mimR Package for Graphical Modelling in R, *Journal of Statistical Software* ***11(6)***

Howitt, P., 1998. “On Some Problems in Measuring Knowledge-Based Growth.” In Dale Neef (1998) *The Knowledge Economy*. Butterworth-Heinemann Boston

Jaffe, A., 1983. “Using Patent Class Data to Measure Technological Proximity and Research Spillovers Among Firms.” Unpublished ms., Harvard University

Jaffe, A., 1986. “Technological Opportunity and Spillovers of R&D: Evidence from Firms’ Patents, Profits and Market Value.” *American Economic Review* **76(5),** pp. 984-1001.

Jaffe, A., 1998. “Patents, Patent Citations, and the Dynamics of Technological Change”. *NBER Report*

Jaffe, A., Fogarty M. Banks. B., 1998. “Evidence from Patents and Patent Citations on the Impact of NASA and other Federal Labs on Commercial Innovation” *The Journal of Industrial Economics* **46(2)**. Inside the Pin-Factory: Empirical Studies Augmented by Manager Interviews: A Symposium pp. 183-205

Jaffe, B., and Lerner, J., 2004. “Innovation and Its Discontents: How Our Broken Patent System is Endangering Innovation and Progress, and What to Do About It.” Princeton University Press, Princeton.

Jaffe A., and Lerner, J., 2001. “Reinventing Public R&D: Patent Policy and the Commercialization of National Laboratory Technologies.” *RAND Journal of Economics* **32(1)** pp. 167-198

Jaffe, A., and Trajtenberg, M., 1996. "Flows of Knowledge from Universities and Federal Laboratories: Modelling the Flow of Patent Citations over Time and across Institutional and Geographic Boundaries.” *Proceedings of the US National Academy of Sciences* **93(99),** pp. 12671-12677

Jaffe, A., and Trajtenberg, M., 1999. "International Knowledge Flows: Evidence from Patent Citations", *Economics of Innovation and New Technology* **8**, pp. 105-136.

Jaffe, A., and Trajtenberg, M., 2002 “Patents, Citations and Innovations: A Window on the Knowledge Economy.” MIT Press, Cambridge.

Jaffe, A., Trajtenberg, M. and Fogarty, S. M., 2000. “The Meaning of Patent Citations: Report on the NBER/Case-Western Reserve Survey of Patentees.” *NBER Working Paper 7631*

Jaffe, A., Trajtenberg, M., and Henderson, R., 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations" *Quarterly Journal of Economics* **108**, pp. 577-598

Jaffe, A., Trajtenberg, M., and Henderson, R., 2005. "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment - Comment" *American Economic Review* **95(1)**, pp. 461- 464

Johansen, S., 1988. "Statistical-Analysis of Cointegration Vectors." *Journal of Economic Dynamics & Control* **12,** pp. 231-254

Jones, C., 1995. "R&D-Based Models of Economic Growth" *Journal of Political Economy* **103,** pp. 759-784

Jones, L., and Manuelli, R., 1990. "A Convex Model of Equilibrium Growth." *Journal of Political Economy* **98**, pp. 1008-1038

Jones, E. L., and Manuelli, R. E., 2005. "Neoclassical Models of Endogenous Growth: The Effects of Fiscal Policy, Innovation and Fluctuations" *Handbook of Economic Growth* **1(1)** pp. 13-65

Jordan, M. I., 1999. "Learning in Graphical Models." MIT Press, Cambridge
Kleinknecht, A., 1996. "Determinants of Innovation: the Message from New Indicators." New York, McMillan

Klenow, J. P., 1998. "Ideas Versus Rival Human Capital: Industry Evidence on Growth Models." *Journal of Monetary Economics* **42**, pp. 3-23

Ko, M., and Bryson, K., 2002. A Regression Tree Based Exploration of the Impact of Information Technology Investments on Firm Level Productivity. ECIS 2002, Poland.

Krugman, P., 1991. "Geography and Trade." MIT Press

Kuznets, S., 1962. "Inventive Activity: Problems of Definition and Measurement." in the *Rate and Direction of Inventive Activity (1962)*, Princeton University Press, Princeton

Lanjouw, J. O., 1998. "Patent Protection in the Shadow of Infringement: Simulation Estimations of Patent Value." *The Review of Economic Studies* **65**, pp. 671-710

Lanjouw, J. O., 1993. "Patent Protection: of What Value and for How Long" NBER Working Paper No. 4475

Lanjouw, J., Pakes, A., and Putnam, J., 1998. "How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data." *Journal of Industrial Economics* **46(4)**, pp. 405-432

Lanjouw, J. O. and Schankerman, M., 1999 "The Quality of Ideas: Measuring Innovation with Multiple Indicators." NBER Working Paper No. 7345

Lanjouw, J. O., and Schankerman, M., 2004. "Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators." *The Economic Journal* **114** pp. 441-465

Lieberman, B. M., 1987. "Patents, Learning By Doing, And Market Structure in the Chemical Processing Industries" *International Journal of Industrial Organization* **5** pp. 257-276

Lipsey, R., Carlaw, K., and Bekar, C., 2005. "Economic Transformations: General Purpose Technologies and Long-Term Economic Growth" Oxford University Press, Oxford.

Long, J. S., 1997. "Regression Models for Categorical and Limited Dependent Variables." Thousand Oaks, CA: Sage

Lucas, R. E., 1988. "On the Mechanism of Economic Development." *Journal of Monetary Economics* **22** pp. 3-42

Lundvall, B. A., 1992. "National Innovation Systems: Towards a Theory of Innovation and Interactive Learning" Pinter, London.

Machlup, F., 1962. "The Production and Distribution of Knowledge in the United States." *Princeton University Press*, New Jersey.

Magee, G. B., 1999. "Technological Development of Foregin patenting: Evidence from 19^{th}-Century Australia". *Explorations in Economic History* **36**, pp. 344-359

Maddala, G. S., 1992. "Introduction to Econometrics", 2^{nd} ed, Macmillan Publishing Company, New York.

Maddala, G. S., and Kim, I., 1998. "Unit roots, Cointegration and Structural Change." Cambridge University Press, Cambridge.

Mankiw, N. G., Romer, D., and Weil, D. N., 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* **107**, pp. 407-307.

Marsh, D., 2004. "An Investigation into the Determinants of innovation in the New Zealand Biotechnology Sector." PhD Thesis. University of Waikato. Hamilton New Zealand

Marshall, A., 1890. **"**Principles of Economics." Macmillan, London

Maurseth, P. B., 2005. "Lovely but Dangerous: The Impact of Patent Renewal." *Economics of Innovation and New Technology* **14(5)** pp. 351-374

McAleer, M., Chan, F., and Marinova, D., 2006. "An Econometric Analysis of Asymmetric Volatility: Theory and Application to Patents" forthcoming in the *Journal of Econometrics*.

Mokyr, J., 1999. "Is There a Theory of Economic History?" Retrieved April 16, 2006 from http://www.econ.yale.edu/alumni/reunion99/mokyr.htm

Mowery, D.C., Nelson, R.R., Sampat, N. B., and Ziedonis, A. A., 2001. "The Growth of Patenting and Licensing by U.S. Universities: an Assessment of the Effects of the Bayh-Dole Act of 1980." *Research Policy* **30**, pp. 99-119.

Nagaoka, S., 2005. “Patent Quality, Cumulative Innovation and Market Value: Evidence from Japanese Firm Level Panel Data.” *Institute of innovation Research*

Narin, F., Noma, E., Perry, R., 1987. “Patents as Indicators of Corporate Technological Strength.” *Research Policy* **16**, pp. 143-155.

Neslon, R., and Winter S. G., 1982. “An Evolutionary Theory of Economic Change” Belknap Press Cambridge Massachusetts.

Nordhaus, W. D., 1969. “Invention, Growth and Welfare: A Theoretical Treatment of Technological Change.” MIT Press, Cambridge

OECD., 1997 “National Innovation Systems.” OECD Publications, Paris.

Pakes, A., 1986. “Patents as Options: Some Estimates of the Value of Holding European Patent Stocks.” *Econometrica* **54(4)** pp. 755-784

Pakes, A., and Schankerman, M., 1984. "The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources", in Zvi Griliches, ed., 1984. “*R&D, Patents and Productivity”* University of Chicago Press, Chicago.

Pakes, A., and Schankerman, M., 1986. "Estimates of the Value of Patent Rights in European Countries During the Post-1950 Period" *Economic Journal* pp. 1052-1076

Pakes, A., and Simpson, M., 1989. “Patent Renewal Data,” *Brookings Papers on Economic Activity* Microeconomic Annual, pp. 331-401.

Pavitt, K., 1988. “Uses and Abuses of Patent Statistics.” In the *Handbook of Quantitative Studies of Science and Technology*. Edited A. F J. van Raan. Elsevier Science Publishers B. B. North Holland.

Putnam, D. J., 1996. “The Value of International Patent Rights.” PhD. Thesis. Yale University. Yale.

Rea, W., Reale, M., Cappelli, C., and Brown. J., 2006. “Identification of Changes in Mean with Regression Trees: An Application to Market Research”. Working Paper. University of Canterbury

Reale, M., 1998. “A Graphical Modelling Approach to Time Series.” PhD Thesis Lancaster University.

Rebelo, S., 1991. “Long-run policy analysis and long-run growth.” *Journal of Political Economy* **99**, pp. 500-521.

Reitzig, M., 2003. What determines patent value? Insights from the semiconductor industry” *Research Policy* 32, pp. 13-26

Romer, P., 1990. “Endogenous Technological Change.” *Journal of Political Economy* 98, 71-102.

Romer, P., 1994. “The Origins of Endogenous Growth.” *The Journal of Economic Perspectives* **8(1)** pp. 2-22.

Rosenberg, N., 1971. “The Economics of Technological Change.” *Penguin Books*, Harmondsworth, London.

Rosenberg. N., 1974. “Science, Invention and Economic Growth.” *The Economic Journal* **84(333),** pp. 90-108.

Schwarz, G., 1978. “Estimating the Dimension of a Model.” *Annals of Statistics* **6**, pp. 461–464

Solow, M. R., 1994. “Perspectives on Growth Theory.” *Journal of Economic Perspectives* **8(1)**, pp. 45-54.

Sokoloff, K. L., 1988. “Inventive Activity in Early Industrial America: Evidence from Patent Records, 1790-1846” *Journal of Economic History* **48(8),** pp. 813-850

Sokoloff, K. L., and Khan, B. Z., 1989. “The Democratization of Invention During Early Industrialization: Evidence from the U.S., 1790-1846” NBER Working paper 10

Stock, J. S., and Watson, M., 1988. “Variable Trends in Economic Time Series.” *Journal of Economic Perspective* **2(3)**, pp. 147-174.

Thompson, P., and Fox-Kean F., 2005. “Patent Citations and the Geography of Knowledge Spillovers: A Reassessment.” *American Economic Review* **95(1)** pp. 450-460

Tong, X., and Frame, D., 1994. “Measuring National Technological Performance with Patent Claims Data.” *Research Policy* 23, pp. 133-141.

Trajtenberg, M., 1990a. “Economic Analysis of Product Innovation - The Case of CT Scanners.” Harvard University Press, Cambridge, Massachusetts

Trajtenberg, M., 1990b. "A Penny for Your Quotes: Patent Citations and the Value of Innovations." *The Rand Journal of Economics* **21(1),** pp. 172-187

Trajtenberg, M., 2001. "Innovation in Israel 1968-97: A Comparative Analysis Using Patent Data." *Research Policy* **30**, pp. 363-389

Trajtenberg, M., Henderson, R., and Jaffe, A., 1997. “University versus Corporate Patents: A Window on the Basicness of Invention." *Economics of Innovation and New Technology* **5(1)**, pp. 19-50

Sakakibara, M., and Branstetter, L., 2001. "Do Stronger Patents Induce More Innovation." *RAND Journal of Economics* **32(1)**, pp. 77-100

Sanders. S. S., Rossman, J., and Harris, L. J., 1958. “The Economic Impact of Patents” *Patent, Trademark and Copyright Journal* **2(2)**, pp. 340-362

Schankerman, M., 1998. "How Valuable is Patent Protection? Estimates by Technology Field" *RAND Journal of Economics* **29(1),** pp. 77-107

Scherer, F, M., 1965. "Firm Size Market Structure, Opportunity, and the Output of Patented Inventions." *The American Economic Review* **55(5),** pp. 1097-1125

Scherer F. M., and Harhoff, D., 2000. "Technology Policy for a World of Skew-Distributed Outcomes" *Research Policy* **29,** pp. 559-566

Schmookler. J., 1966. "Invention and Economic Growth." Harvard University Press. Cambridge Massachusetts.

Shi, P., and Tsai, C. L., 2002. "Regression Model Selection – a Residual Likelihood Approach." *Journal of Royal Statistics Society* **13**, pp. 586-598.

Smith, A., 1976. "An Inquiry into the Nature and Causes of the Wealth of Nations." *Clarendon Press*, Oxford

Smith, K., 1991. "Economic Returns to R&D: Methods, Results and Challenges" *Science Policy Support Group Review* **3**

Sullivan, R. J., 1994, "Estimates of the Value of Patent Rights in Great Britain and Ireland. *Economica* **61(241)** pp. 1852-1876.
US Patent and Trademark Office., 2006. "General Information Concerning Patents" retrieved August 20
http://www.uspto.gov/web/offices/pac/doc/general/index.html#patent

Wasserman, L., 2003. "All of Statistics: A Concise Course in Statistical Inference." Springer, New York.

Wikipedia., 2006. Retrieved August 16 from
http://en.wikipedia.org/wiki/Computed_axial_tomography

Winkelmann, R., 1994. "Econometric Analysis of Count Data." Springer-Verlag, New York.

11 Appendix

United States Patent — **4,203,158**
Frohman-Bentchkowsky, et. al. — **May 13, 1980**

Electrically programmable and erasable MOS floating gate memory device employing tunneling and method of fabricating same

Inventors: **Frohman-Bentchkowsky; Dov** (Haifa, IL); **Mar; Jerry** (Sunnyvale, CA); **Perlegos; George** (Cupertino, CA); **Johnson; William S.** (Palo Alto, CA).
Assignee: **Intel Corporation** (Santa Clara, CA).
Appl. No.: **969,819**
Filed: **Dec. 15, 1978**

Related U.S. Application Data

Continuation-in-part of Ser No. 881,029, Feb. 24, 1978, abandoned.

Intl. Cl. :	**G11C 11/40**
U.S. Cl.:	365/185; 307/238; 357/41
Current U.S. Class:	365/185.29
Field of Search:	365/185, 189; 307/238; 357/41, 45, 304

References Cited | [Referenced By:]

U.S. Patent Documents

3,500,142	Mar., 1970	Kahng	365/185
4,051,464	Sept., 1977	Huang	365/185

Primary Examiner: Fears; Terrell W.
Attorney, Agent or Firm: Blakely, Sokoloff, Taylor & Zafman

Abstract

An electrically programmable and electrically erasable MOS memory device suitable for high density integrated circuit memories is disclosed. Carriers are tunneled between a floating conductive gate and a doped region in the substrate to program and erase the device. A minimum area of thin oxide (70 A-200 A) is used to separate this doped region from the floating gate. In one embodiment, a second layer of polysilicon is used to protect the thin oxide region during certain processing steps.

16 Claims, 14 Drawing Figures

Figure 9 Patent Grant Document, Source: Trajtenberg (2001)