CESPRI

Centro di Ricerca sui Processi di Innovazione e Internazionalizzazione

Università Commerciale "Luigi Bocconi"
Via R. Sarfatti, 25 – 20136 Milano
Tel. 02 58363395/7 – fax 02 58363399
http://www.cespri.it

Emanuele Bacchiocchi, Fabio Montobbio

EPO VS USPTO CITATION LAGS

WP n. 161 **September 2004**

Stampato in proprio da:
Università Commerciale Luigi Bocconi – CESPRI Via Sarfatti, 25 20136 Milano
Anno: 2004

EPO vs. USPTO Citation Lags

E. Bacchiocchi[a], F. Montobbio[b]

[a]University of Milan,
[b]University of Insubria and CESPRI - Bocconi University, Milan

September 30, 2004

Abstract

This paper estimates the diffusion and obsolescence of technological knowledge by technological field, country and type of institutions that generates it. We use two comparable samples of patents and patent citations from the NBER U.S. Patent Citations Dataset (based on patents from the US Patent Office) and from the EP Cespri Dataset (based on patents from the European Patent Office). Using a quasi-structural model, as proposed by Caballero and Jaffe (1993) and discussed in Hall et al. (2001), we test whether the observed processes of knowledge diffusion and obsolescence reflect the specific institutional mechanism generating them. Results show that at the USPTO there are more citations per patent due to the different rules governing citation practices and that their median lag is twice as large relatively to the citations at the EPO. We also find that the relative properties of the citation frequencies in different technological fields change according to the patent office considered.

Keywords: Patent citations, Knowledge flows, Spillovers, Diffusion, Patent Examiners
JEL Classification: O30, O33, O34

1 Introduction

Patents are a fundamental empirical source to measure research productivity and patent citations are increasingly used to evaluate the value of patents (e.g. to evaluate companies' patent portfolios) (Hall et al. 2004; Lanjouw and Shankermann, 2004; Fung and Chow, 2002; Jaffe and Trajtenberg, 2002; Haroff et al. 1999; Trajtenberg, 1990) and to track knowledge flows between different applicants or inventors (e.g. intensity and geographical and technological scope of knowledge spillovers) (Verspagen, 1997; Maruseth, Verspagen, 2002, Malerba Montobbio, 2003; Malerba et al. 2003). This paper estimates the diffusion and obsolescence of technological knowledge by technological fields, countries and type of institutions using patent citations[1] from two datasets from the US and European Patent Offices (USPTO and EPO, respectively). It tackles three issues that make the use and interpretation of citation counts extremely relevant for economic analysis.

The first issue is related to the time dimension of knowledge R&D spillovers. Patents and patent citations have been increasingly used as a channel to measure knowledge spillovers from R&D activity but relationships have been often assumed contemporary and the time dimension tends to be unexplored. This paper enquires whether the observed processes of diffusion and obsolescence using patent citations reflect the technological properties of knowledge production or rather, the specific institutional mechanism generating them.

The second issue concerns the effect on the statistical properties of the distribution of citations of the differences in the institutional process governing the decision (by inventors or patent examiners) to include a patent citation in the patent document. There are relevant differences between citation practices in the USPTO and EPO. In the US there is the 'duty of candor' rule which imposes all applicants to disclose all the prior art they are aware with. Therefore many citations at the USPTO come directly from inventors and applicants and are subsequently filtered by patent examiners[2]. At the European Patent Office there is not such a rule and patent examiners draft their report trying to include all the technically relevant information within a *minimum* number of citations (Michel and Bettels, 2001; Akers, 2000; Breschi and Lissoni, 2004). Hence at EPO patent citations are, with few exceptions, added by the examiners. As a result the analysis of diffusion and obsolescence of technological knowledge (e.g. for specific technologies and/or companies' portfolios) and knowledge spillovers may reveal different properties according to the patent dataset that is used.

The third issue is the truncation bias: recent cohorts of patents are less likely to be cited then the older ones, because the pool of potentially citing patent is smaller. This issue is addressed with a quasi-structural model as proposed

[1] Citations help examiners and applicants to assess the degree of novelty and inventive step of the claims of the patent. Once published citations provide a legal delimitation of the scope of the property right conveyed by the patent.

[2] Alcàcer and Gittleman (2004) using a random sample of 442,839 patents granted at the USPTO over the period 2001-2003 show that 40% of the cited-citing pairs are generated by patent examiners.

by Caballero and Jaffe (1993) and discussed in Hall et al. (2001). This model provides a flexible empirical tool to adjust raw citation counts. The extent of the truncation bias may differ according to the data source.

These issues have not been consistently addressed yet. To this aim this work compares different properties of the processes of obsolescence and diffusion of knowledge measured by patent citations in the two different datasets. The processes are evaluated by technological fields, countries and institutional types of applicants. Obsolescence and diffusion of knowledge vary across technological field. This is because in different technologies opportunities for innovations are different, there may be different probabilities of innovating given the specific stock of knowledge produced in the field and, finally, there are different mechanisms of appropriability and, therefore, use of intellectual property rights. Moreover we expect to observe variation across countries because countries may differ in firms' patenting practices and also in the value of their innovations. Finally for policy-making purposes it is important to perform an institutional comparison by looking at different institutional types of applicant. In particular we distinguish between government and non government (corporate) patents.

Results show that at the USPTO there are more citations per patent due to the different rules governing citation practices and that their approx median lag is twice as large relatively to the citations at the EPO. Strikingly, we also find that the relative properties of the citation frequencies in different technological fields change according to the patent office considered.

2 The data

We use the publicly available NBER U.S. Patent Citations Data, which contains the 2,923,922 USPO (granted) patents from 1963 to 1999 and 16,522,438 citations from (and to) USPO patents from 1975 to 1999 (Hall et al., 2001 and Jaffe and Trajtenberg, 2002), and the EP Cespri dataset, which contains the 1,391,350 EPO patent applications from 1978 to 2001 and 1,119,761 citations from (and to) EPO patents from 1978 to 2001[3]. From these datasets (from now on USPOD and EPCD) we selected two samples: the universe of all patents and patent citations between 1978 and 1998. In particular we consider all the citations from patents granted between 1979 and 1998 to patents granted between 1978 and 1997 (in the EPCD we use patent applications). This is done in order to have the same right and left truncation biases in the two datasets. Summary statistics are displayed in Table 1. Each patent is characterized by a date, a country (first inventor's address) a technological field (based on the International Patent Classification for EPCD and the USPTO classification system for the USPOD) and the institutional type of the applicant (government or non government) (Details for both datasets are provided in the Appendix).

[3]Bibliographic data on patent applications come from the Espace Bulletin CD-R produced by the EPO, patent citations come from the REFI tape.

Table 1: Statistics for EP and US patent and citation samples

	EP Dataset	USPO Dataset
Range of cited patents	1978-1997	1978-1997
Range of citing patent	1979-1998	1979-1998
Potentially cited patents	906,792	1,766,075
Potentially citing patents	984,148	1,734,687
Total citations	959,852[a]	8,080,276[a]
Citations per potentially citing patent	0.98	4.66
Citations per citing patent	1.86	5.59
Cited patents by fields,%[b] (potentially cited patents in parenthesis)		
Chemicals	27.45 (22.1)	17.93 (19.3)
Computers and Communications	10.58 (10.1)	17.60 (12.6)
Drugs and Medical	12.92 (9.5)	10.8 (9)
Electrical and Electronics	12.72 (13)	18 (17.5)
Mechanical	29.89 (35.3)	18.05 (21.2)
Others	6.43 (9.8)	17.62 (20.2)
Cited Patents by country,% (potentially cited patents in parenthesis, %)		
Germany	16.06 (20.1)	5.99 (7.8)
France	6.59 (7.9)	2.34 (3)
Italy	2.73 (3.2)	0.83 (1.2)
United Kingdom	7.57 (6.5)	2.64 (2.9)
Japan	21.82 (18.5)	19.6 (19.9)
United States	31.76 (29.1)	61.09 (54.7)
Sweden and Finland	2.17 (2.5)	0.94 (1.2)
Others	11.29 (12)	6.56 (9.1)
Cited Patents by institutional field,%[c] (potentially cited patents in parenthesis, %)		
not assigned	9.14 (10.6)	14.62 (16.8)
firms	87.46 (86.3)	83.93 (81.5)
non firms	3.40 (3.1)	1.45 (1.6)

a. Cells with the lag $T - t < 1$ have been removed,

b. see the Appendix for the sectoral concordance between EPCD and USPOD,

c. in the EPCD the group called 'firm' includes just companies while in the USPOD this group includes 'non government organization'. The group called 'non firm' in the EPCD includes university and public research centres while in the USPOD dataset is just 'government'.

At the USPTO there are more patents and in particular much more citations per patent. This is expected on the basis of the different institutional processes underlying the citation practices. In Table 1 the institutional, technological and country composition of citations are displayed and compared with the composition of the EPCD and USPOD patent samples. If the share of total citations of a country is higher than its fraction of total patents (in parenthesis), this indicates an above average citation intensity for that country. The same reasoning occurs for technological fields and institutional types. It's worthwhile noting that the US have a higher share of citations relatively to their share in the patent sample in both datasets. This probably reflects their position as world wide technological leaders. Moreover in both datasets the citation intensity is higher in Pharmaceutical and lower in the Mechanical sectors. This is due to the different nature of the technologies and probably their different average patent scope. However only in the EPCD there is a relatively higher propensity to cite in Chemicals and only in the USPOD there is a remarkably higher propensity to cite in Computers and Communications. This raises the issue on which other variables affect the citation intensity of a technological field beyond its technological characteristics. We discuss at lenght this issue in the following section.

3 Model specification and econometric framework

The specification and estimation of an econometric model allows the researcher to infer the diffusion process while distinguishing and controlling for the multiple effects on the citation rates[4]. We describe the random process underlying the generation of citations with a quasi-structural approach, as described in Hall et al. (2001). The diffusion process is modelled as a combination of two exponential processes, one for the knowledge diffusion and the other for the natural process of obsolescence. The general formulation of the model is

$$\begin{aligned} p(k,K) &= \alpha(k,K)\exp\left[-\beta_1(k,K)(T-t)\right] \\ &\quad \times \left(1-\exp\left[-\beta_2(k,K)(T-t)\right]\right) \end{aligned} \tag{1}$$

where $p(k,K)$ is the likelihood that any particular patent k, granted at time t, is cited by some particular patent K, granted at time T. The parameters β_1 and β_2 represent the rate of obsolescence and of diffusion, respectively, and both exponential processes depend on the citation lag $(T-t)$.

The coefficient α does represent a multiplicative factor, as the constant term in a simple linear regression model. However, as indicated by the dependence of α from (k,K), such proportionality factor $\alpha(k,K)$ is allowed to vary with attributes of the citing and cited patents. The estimate of a particular $\alpha(k,K)$, indicates the extent to which a patent k is more or less likely to be cited, with respect to a base characteristic patent, by a patent K.

From the formulation above, β_1 and β_2 single out the main features of the diffusion process. The lag at which the citation function is maximized, i.e. the modal lag, is approximately equal to $1/\beta_1$, while the maximum value of the citation frequency is approximately equal to β_2/β_1. Such features of the model have important implications for both the estimation and interpretation of the results. In fact, an increase in β_1 simply shifts the citation function to the left, while an increase in β_2, leaving β_1 unchanged, increases the overall citation intensity, at every value of $(T-t)$. As a consequence, variations in β_2 with β_1 unchanged are not separately identified from variations in the constant term α. Following Jaffe and Trajtenberg (1996), thus, we prefer allowing variations in α leaving β_2 constant for all observations.

The fact that the constant term α and the two structural parameters β_1 and β_2 depend on k and K indicates that they depend upon particular features of both cited and citing patents. From the empirical point of view, however, modelling single pairs of patents (citing and cited), might conduct to dealing with very small expected values. Therefore we aggregate patents in homogeneous groups and model the amount of citations to a particular group of cited patents by a particular group of citing patents. We want to have a finer understanding of the statistical properties of the citations received (forward citations), since

[4]The model follows the specification in Jaffe and Trajtenberg (1996), which adapts the original formulation proposed by Caballero and Jaffe (1993). Both the papers have been republished in Jaffe and Trajtenberg (2002).

this is the usual way of assessing the value of patents. The following characteristics of the cited patent k might affect its citation frequency (see the Appendix for relative details of the USPOD and EPCD):

- t, the application or priority date,
- p, the first inventor's country,
- c, the technological field,
- i, the institutional type.

Moreover the following attributes are considered for the citing patent K.

- T, the application or priority date,
- g, the first inventor's country,

The amount of citations to a specific group of cited patents by a specific group of citing patents is: c_{tpicTg}. Hence a treatable formulation of the model, where the various different effects enter as multiplicative parameters, becomes

$$\begin{aligned} E(c_{tpicTg}) &= (n_{tpic})\,(n_{Tg})\,\alpha_t\alpha_p\alpha_i\alpha_c\alpha_T\alpha_g \exp\left[-\left(\beta_1\right)\beta_{1p}\beta_{1i}\beta_{1c}\beta_{1g}\,(T-t)\right] \\ &\quad \times (1-\exp\left[-\beta_2\,(T-t)\right]) \end{aligned} \tag{2}$$

or equivalently, in the estimable form

$$\begin{aligned} p_{tpicTg} &= \frac{c_{tpicTg}}{(n_{tpic})\,(n_{Tg})} = \alpha_t\alpha_p\alpha_i\alpha_c\alpha_T\alpha_g \exp\left[-\left(\beta_1\right)\beta_{1p}\beta_{1i}\beta_{1c}\beta_{1g}\,(T-t)\right] \\ &\quad \times (1-\exp\left[-\beta_2\,(T-t)\right]) + \varepsilon_{tpicTg} \end{aligned} \tag{3}$$

where n_{tpic} and n_{Tg} represent the total amount of potentially cited and citing patents for each of the particular (*tpic*) and (Tg) groups, respectively. The model (3) can thus be estimated by nonlinear least squares under the well known hypotheses on the residuals terms ε_{tpicTg}.

Variations in any particular $\alpha\,(k)$ (i.e. the multiplicative coefficients related to cited patents) should be interpreted as differences in the propensity to be cited, with respect to the base category. As an example, let consider an estimated coefficient $\alpha\,$(k=Computers and Communications) = 2.094; this means that patents belonging to the category "Computers and Communications" have a more than double probability (across all lags) to receive a citation in the next years vis à vis patents belonging to the base field.

Equivalently, estimates of multiplicative coefficients related to citing patents, $\alpha\,(K)$, indicate differences in the propensity to cite compared to a base category. One coefficient for each category, thus, will be omitted from the estimation procedure and will be constrained to unity.

Similar interpretation has to be given to variations in β_1 coefficients, which represent differences in the rate of decay or obsolescence across categories of cited and citing patents. Higher values of β_1, with respect to the base category, means a faster obsolescence, which corresponds to a downward and leftward shift in the citations function.

Table 2. Statistics for regression variables

	EPCD				USPOD			
	Mean	St. Dev	Min	Max	Mean	St.Dev	Min	Max
Number of citations	3.97	18.95	0	776	33.4	233.86	0	13661
Potentially cited patents	262.36	579.7	1	6626	588.77	1335.22	1	13433
Potentially citing patents	7414.97	5843.27	277	25813	11903.73	17359.69	320	76976
Citation Frequency (10^6)	2.61	12.58	0	1632.65	4.86	15.25	0	1619.43
Lag in years[a]	7.33	4.82	1	20	7.33	4.82	1	20
Regression weights	907.84	1111.34	16.64	13078.11	1442.3	2232.51	17.89	29690.93

a. Cells with the lag $T - t < 1$ have been removed.

In Table 2 there are the statistics for the regression variables. The data consist of one observation for each feasible combination of values of t, p, i, c and L and g. For the cited patents we have 20 years, 3 institutional types, 6 technological fields, and 8 countries and for the citing patents we have 20 years and 8 countries. We consider only citations with a lag between the citing and cited patent greater than or equal to 1. Hence the total amount of observations is: n_obs=[(20*21)/2]*8*8*6*3=241920. In each dataset there are some cells with zero citations and some cells with missing values. We have zeros when c_{tpicTg} is zero and $(n_{tpic})(n_{Tg})$ is positive. Missing values are generated when also $(n_{tpic})(n_{Tg})$ is zero. In the EPCD 144481 observations have zero citations (59%) and there are 15360 missing (6.3 %). These are due to the scarcity of patents by universities or public research centers in Germany and Italy between '78 and '82 and Sweden and Finland mainly between '78 and '86 In the USPOD 81454 obs. have zero citations (33%) and 24616 observation are missing. Missing values come from the scarcity of patents by universities or public research centres in Germany, Italy and Sweden and Finland.

One more consideration about the specification of the model concerns the difficulties in estimating citing and cited time effects together with the citation lag; in fact, citation lags enter the model non-linearly and the identification of all effects is not precluded a priori. However due to the great number of parameters to be estimated we prefer to calculate the fixed effects grouping cited years into 5-year intervals, as in Jaffe and Trajtenberg (1996)[5]. We estimate the model using weighted non-linear least squares. The weights are needed in order to deal with heteroskedasticity. Since each observation is obtained dividing the number of citations by the product of the total amount of potentially citing and potentially cited patents corresponding to a given cell, it has been weighted by $(n_{tpic}n_{Tg})^{1/2}$, following Jaffe and Trajtenberg (1996) and Hall et al (2001).

[5]Grouping cited year is a reasonable assumption in that we believe that the fertility of invention do not change substantially over time. Estimated results, not reported in the present paper, confirm such assumption.

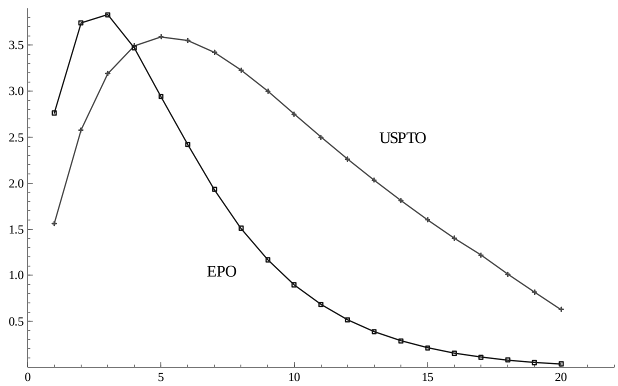

Figure 1: Fitted frequency ($\times 10^6$) of citation from EPCD and USPOD.

4 Empirical results

The results from the estimation of equation (3) are reported in Table 3. All fixed effects have been estimated relative to a base value of unity; for each effect thus, one group is omitted from the estimation and constrained to unity. Significant tests for the estimates of any particular $\alpha(k)$, being a proportionality factor, focus on the null hypothesis $H_0 : coeff = 1$. The null hypothesis of significant tests for both β_1 and β_2, however, remains the standard $H_0 : \beta_i = 0$, $i = 1, 2$.

Results show that citations in the EPCD have shorter life and the rate of decay is twice the one observed for USPOD ($\beta_1 = 0.396$ and $\beta_1 = 0.189$ for the EPCD and USPOD respectively). The modal lag is approx 5.3 for the USPOD[6] and 2.7 for the EPCD. For the two datasets average fitted values of equation (3) are plotted in Fig. 1. The likelihood that a EPO patents is cited becomes half of its estimated maximum after about 6-7 years while for USPTO patents this occurs after 14-15 years. Moreover after 20 years, the estimated probability for a EPO patent to be cited is almost zero, for a USPTO patent it is one fourth of its maximum value.

The goodness of fit of the model, measured as adj-R^2, highlights the difficulty of such double-exponential model to fit zero probabilities. The adj-R^2 for the USPOD and EPCD datasets corresponds to 0.45 and 0.22 respectively. The low goodness of fit for the European data can be easily explained by observing that the percentage of zeros is almost double with respect to the US data (59% against 33%).

Time Effects. The estimated citing year effects, at the USPOD, do not show any upward trend. All estimated coefficients appear to be greater than

[6] This confirms approximately the results of Jaffe and Trajtemberg (1996 and 1999) even if our estimated $\beta_1 = 0.189$ is slightly lower.

one but in many cases they are not significantly different from one. For the EPOD instead, the α_T display a steep downward trend. As the amount of potentially citing and cited patents increases over time in both datasets, the amount of citations per patent grows faster in the USPOD than in the EPCD. This creates the observed decline in the coefficients for the EPCD and the absence of a trend for the USPOD. For the cited time effects a substantial absence of fertility changes characterizes both datasets. To substantiate this conjecture we calculated the differences in level and trend of the raw amount of backward citations per citing patent in the two data sets (note that in the two datasets we have the same left truncation bias because we do not consider citations that goes to patents granted, or applied for, before 1978). In the EPCD backward citations per patent are 1.16 in 1979, they reach the maximum in 1994 at 2.10, declining slightly afterwards. In the USPOD backward citations per patent are 1.26 in 1979 and they grow more steeply reaching the maximum in 1995 at 8.28.

Technological Fields. Two types of variation relative to the technological fields are considered in the model: variations in the fixed effects α_c and in the obsolescence parameter β_1. The base field is 'Chemicals' for both the USPOD and EPCD. Somehow surprising results appear to be quite different for the two datasets (see Table 3, Figure 2 and Figure 3). We confirm the results for the USPOD of Jaffe and Trajtenberg 1996 and Hall et al. 2001 showing that Computers and Communications, Electrical and Electronics have a higher rate of decay and reach their modal lag earlier with respect to the other technological field. Conversely Pharmaceuticals reach their modal lag at least two years later but produce significant patents for a longer period. However at the EPCD the picture seems to be quite different. In particular it is reversed the behavior of citation in Computers and Communications and Pharmaceuticals with the former display a relative lower rate of decay and the latter reaching the modal lag slightly earlier. Moreover in Europe the Chemicals display very high early citations but the most rapid obsolescence.

These differences may occur for at least three reasons. In principle there may be some noise due to the different patent classifications on which the technological fields are built. As explained in the Appendix, differences between the two datsets may emerge because the matching between the US NBER categories and the reaggregation of 30 technological classes based on European IPC codes may be imperfect. However we do not think this can be the only explanation of these diverging sectoral patterns. If the technological fields have the same composition in terms of amount and type of technological classes, it might also be that these technological classes have a different size in terms of amount of patents in the different Patent Offices. Therefore a second explanation may come from the different technological specialisiation across Patent Offices. Finally, these sectoral differences may arise also if at the EPO and USPTO we had excactly the same sectoral aggregates with exactly the same weight in terms of patents. In this case the different relative properties of diffusion and decay of the technological fields would depend simply upon differences in the relative amount of citations received in each technological class. This last possibility may reflect different qualities of patents in the same fields in the different patent offices.

Institutional Types. For the European data, patents assigned to Universities or Public Institutions and to Companies are respectively 40% and 18% more likely to be cited than the 'Not Assigned' patents. For the US data instead (as in Jaffe and Trajtenberg, 1996), non government patents are cited significantly more than government ones, although they have a slightly higher rate of decay. These differences are probably affected by the different classifications in the two datasets. For example a relevant role is probably played by university patents that seem to have higher likelihood to be cited according to Jaffe and Trajtenberg (1996). These patents in the USPOD belong to the non government group while in the EPCD they are in the non firm group.

Country Effects. For what concerns the country effects (Tab. 3 and Figure 3) results are less dissimilar between the two datasets. The US situation appears quite clear: at the USPOD patents granted to American inventors are more likely to cite and to be cited at every lags. Moreover, the gap respect to the other countries is consistent and in the order of 30% and more, in both citing and cited cases. In the EPCD Japanese patents have the higher probability to be cited. Japanese patents have the highest rate of decay in both the EPCD and the USPOD. This might reflect the country specific patenting and citing practice as emphasized by Ordover (1991) among others. Before recent reforms the so called "Sashimi system" was characterized by a narrower patent scope and limited number of claims (one single independent claim before 1988). This patent structure increases the amount of patents and citations.

Table 3: Estimated results

	United States		EU	
	coeff.	$t - statistic$ H_0: $coeff$=1	coeff.	$t - statistic$ H_0: $coeff$=1
citing year effect ($base$=1979)				
1980	1.191	3.28	0.859	-2.28
1981	1.233	4.04	0.872	-2.19
1982	1.178	3.27	0.878	-2.14
1983	1.139	2.66	0.776	-4.44
1984	1.095	1.89	0.755	-5.02
1985	1.077	1.56	0.717	-6.09
1986	1.093	1.86	0.705	-6.44
1987	1.107	2.12	0.646	-8.42
1988	1.102	2.03	0.607	-9.93
1989	1.083	1.68	0.576	-11.23
1990	1.068	1.38	0.552	-12.29
1991	1.081	1.63	0.556	-12.04
1992	1.131	2.51	0.547	-12.40
1993	1.183	3.36	0.532	-13.09
1994	1.226	3.97	0.524	-13.44
1995	1.344	5.51	0.480	-15.89
1996	1.249	4.27	0.434	-19.02
1997	1.125	2.36	0.375	-24.02
1998	0.882	-2.80	0.292	-34.75
cited time effect ($base$=1978−1982)				
1983-1987	1.049	8.36	0.986	-1.24
1988-1992	1.040	4.31	0.948	-3.06
1993-1997	0.967	-2.76	0.972	-1.16
institutional nature ($base$=$not\ assigned$)				
companies	1.348	34.17	1.181	8.37
Univ or public	0.839	-7.72	1.397	10.12
technological field ($base$=$chemical$)				
computer & communication	2.094	65.75	0.836	-12.46
drugs & medical	1.336	27.98	1.243	14.04
electrical & electronic	1.407	32.89	0.771	-19.43
mechanical	0.990	-1.01	0.592	-53.61
others	0.943	-6.35	0.395	-54.67

Table 3: Estimated results, continued

cited patent country				
(base=United States)				
Germany	0.505	-66.47	0.544	-49.38
France	0.517	-43.71	0.602	-27.98
Italy	0.453	-30.16	0.643	-15.35
Great Britain	0.600	-36.99	0.980	-1.16
Japan	0.700	-60.55	1.281	18.70
Sweden and Finland	0.604	-21.89	0.749	-8.92
Other	0.615	-52.76	0.796	-14.86
citing patent country				
(base=United States)				
Germany	0.433	-156.02	0.717	-51.83
France	0.492	-88.97	0.784	-26.73
Italy	0.417	-66.31	0.711	-24.65
Great Britain	0.633	-61.14	1.052	5.62
Japan	0.607	-178.27	1.089	13.98
Sweden and Finland	0.584	-47.71	0.735	-19.74
Other	0.537	-150.18	0.873	-18.37
β_1	0.189	121.67	0.396	71.77
β_2	3.29E-06	21.86	9.27E-06	15.12
rate of obsolescence				
by technological field				
(base=Chemical)				
computer & communication	1.045	7.61	0.878	-12.91
drugs & medical	0.812	-33.54	0.977	-2.48
electrical & electronic	1.140	19.89	0.924	-8.16
mechanical	1.064	8.89	0.863	-18.52
others	0.970	-4.54	0.797	-13.26
byinstitutional nature				
(base=not assigned)				
companies	1.105	16.82	1.008	0.69
univ. or public	1.052	2.88	1.069	3.40
by cited patent country				
(base=United States)				
Germany	0.974	-2.54	0.875	-12.56
France	0.965	-2.32	0.893	-7.51
Italy	0.964	-1.25	0.900	-4.42
Great Britain	0.940	-4.92	0.974	-2.26
Japan	1.037	6.88	1.074	8.87
Sweden and Finland	0.949	-2.42	0.902	-4.15
Other	0.984	-1.84	0.924	-7.11

5 Conclusion

In recent years patent citations have been increasingly used to analyze the value of patents and to measure knowledge flows. This paper represents an initial comparison between the statistical properties of citations from two of the main Patent Offices world wide: the European Patent Office and the US Patent Office. It enquires whether the statistical properties of diffusion and decay of technical knowledge, as measured by patent citations, are robust to the specific institutional process that generates them. We have found some intriguing results that in part re-confirm findings in Jaffe and Trajtenberg (1996, 1999) and Hall et al. 2001 and, in part, raise new additional questions.

First, we show the well-known result that at the USPTO there are more citations per patent due to the different rules governing the citation practices. Second, citations at the USPTO have longer life and a lower rate of decay. The approximate median lag is twice as large relatively to the citations at the EPO. Third, we do not find a strong evidence of an increased citing year effect in US as in Jaffe and Trajtenberg (1996, 1999) and Hall et al. (2001). Moreover for the same coefficients we find a clear downward trend in Europe. This effect is due to the increased amount of potentially citing and cited patents and to the relatively more stable amount of citations per patent at the EPO.

Strikingly, we found that the relative properties of the citation frequencies in different technological fields change according to the patent office considered. This raises the questions on why this occurs. First there may be differences in the relative size of the technological classes composing the broad technological fields that we use in this paper. Second the different relative properties of diffusion and decay of the technological fields in the two datasets may reflect the different technological significance of patents. This might be due to the strategic patenting choices of firms and to the relative technological specialization of the United States and Europe. Further work is required to substantiate these conjectures.

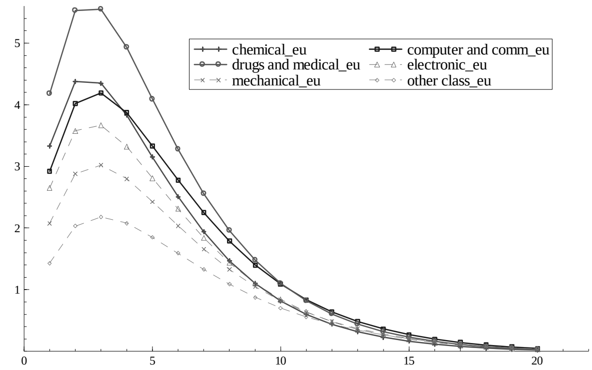

Figure 2: Fitted citation function for class of patents from the EPCD dataset.

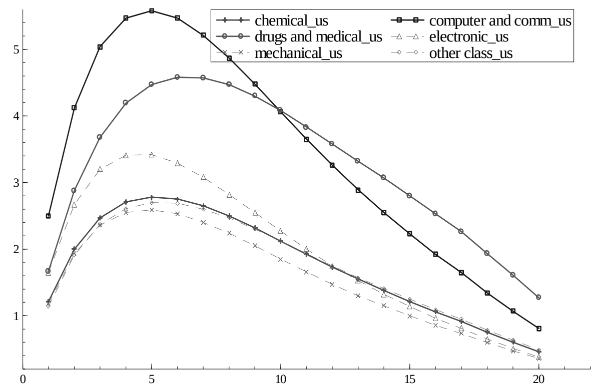

Figure 3: Fitted citation function for class of patents from the USPOD dataset.

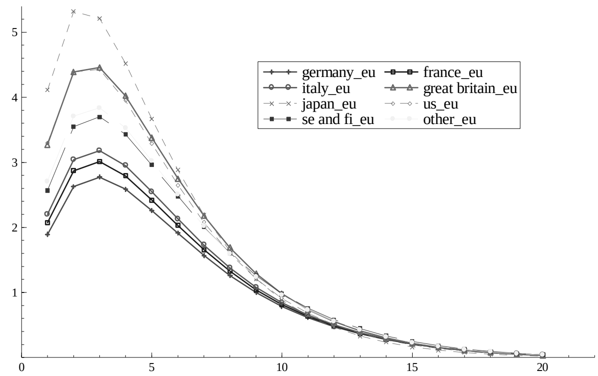

Figure 4: Fitted frequency of citation to patents originating in different countries; resuts from the EPCD dataset.

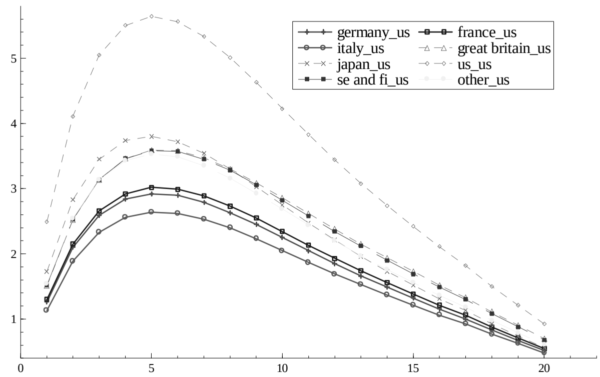

Figure 5: Fitted frequency of citation to patents originating in different countries; results from the USPOD dataset.

Appendix

In both datasets *Countries* are defined on the basis of the address of the first inventor in the patent application. We have used 8 countries and country groups: 1. Germany, 2 France, 3. Italy, 4. United Kingdom, 5. Japan, 6. United States, 7. Sweden and Finland, 8.others.

The *Technological Fields* are the US NBER categories as in Hall et al (2000) that can be found in the USPOD. For the EPCD we used 30 technological classes based on the Annex III-A of OECD (1994). This classification aggregates all (primary) IPC codes (version 7 used at the EPO) into 30 technological classes. A concordance table has been created by the authors that reaggregates the 30 classes into the USPOD Fields The USPOD fields are: 1. Chemical, 2. Computers & Communications, 3. Drugs & Medical, 4. Electrical & Electronic, 5. Mechanical, 6. Others. Below we report the 30 classes and, in parenthesis, the USPOD field that has been assigned to each class by the authors: 1. Electrical engineering (4), 2. Audiovisual technology (4), 3. Telecommunications (2), 4. Information Technology (2) 5. Semiconductors (4), 6. Optics (5), 7. Control Technology (5), 8. Medical Technology (5), 9. Organic Chemistry (1), 10. Polymers (1), 11. Pharmaceutics (3), 12. Biotechnology (3), 13. Materials (1), 14. Food Chemistry (1), 15. Basic Materials Chemistry (1), 16. Chemical Engineering (1), 17. Surface Technology (5), 18. Materials Processing (5), 19. Thermal Processes (6), 20. Environmental Technology (6), 21. Machine Tools (5), 22. Engines (5), 23. Mechanical Elements (5), 24. Handling (5), 25. Food Processing (6), 26. Transport (5),27. Nuclear Engineering (4), 28. Space Technology (5), 29. Consumer Goods (6), 30. Civil Engineering (6).

The *institutional nature* of the assignee could not be built exactly in the same way for the two datasets. In particular in the EPCD the group called 'firms' includes just companies while in the USPOD this group includes 'non government organization'. The group called 'non firm' in the EPCD includes university and public research centres while in the USPOD dataset is just 'government'.

Finally we have chosen the closest *dates* available to the actual timing of invention for both datasets. These are the priority date for the EPCD and application date for the USPOD.

References

Akers N. (2000), 'The referencing of prior art documents in European patents and applications', World Patent Information 22, pp. 309-315

Alcàcer J., Gittleman, M. (2004); How do you know what you know? Patent examiners and the generation of patent citations, NBER Summer Institute Presentation.

Breschi S., Lissoni F. (2004), Knowledge networks from patent data: Methodological issues and research targets, in Moed, Henk F.; Glänzel, Wolfgang; Schmoch, Ulrich (Eds.), Handbook of Quantitative Science and Technology Research. The Use of Publication and Patent Statistics in Studies of S&T Systems. Springer. Berlin.

Fung M.K., William W. Chow, (2002); Measuring the intensity of knowledge flow with patent statistics Economics Letters Volume 74, Issue 3, 353-358.

Fung M. K. (2003); Technological proximity and co-movements of stock returns. Economics Letters. v. 79(1),131-136.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2004). Market Value and Patent Citations. Rand Journal of Economics. forthcoming.

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

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2000). Market Value and Patent Citations: A First Look," NBER Working Paper W7741.

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

Jaffe, A.B., M. Tratjenberg (1999). International Knowledge Flows: Evidence from Patent Citations. Economics of Innovation and New Technology, Vol. 8 (1999): 105-136.

Jaffe, A.B., M. Trajtenberg (2002). Patents, Citations and Innovations: A Window on the Knowledge Economy M.I.T. Press.

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

Malerba F., Mancusi M., Montobbio F. (2003); Innovation and Knowledge Spillovers: Evidence from European Data. Working Paper n. 39. University of Insubria, presentation at the American Economic Association meeting, San Diego 5 Gennaio 2003.

Malerba F., Montobbio F. (2003); Exploring Factors Affecting International Technological Specialization: the Role of Knowledge Flows and the Structure of Innovative Activity, Journal of Evolutionary Economics, v. 13, n. 4, pp. 411-434

Maurseth, P.B., Verspagen, B, (2002). Knowledge Spillovers in Europe: A Patent Citations Analysis, Scandinavian Journal of Economics, vol. 104(4), p. 531-45.

Michel J., Bettels B. (2001), Patent citation analysis: A closer look at the basic input data from patent search reports, Scientometrics 51, pp. 185-201

OECD (1994) Using Patent Data as Science and Technology Indicators – Patent Manual, OECD, Paris.

Ordover (1991); A patent system for both diffusion and exclusion. Journal of Economic Perspectives, 5(1)

Verspagen, B., (1997); Estimating International Technology Spillovers Using Technology Flow Matrices, Weltwirtschaftliches Archiv, 133, 226-248.