Indirect Patent Citations

Gamal Atallah*
University of Ottawa and CIRANO

Gabriel Rodríguez
University of Ottawa

First Version: August 2002
This Version: March 4, 2003
(Still preliminary)

Abstract

Patent citations are extensively used as a measure of patent quality. However, counting citations does not account for the fact that citations come from patents of different qualities, and that some citations are of a higher quality than other citations. We develop a citation index which takes into account the cumulative quality of the citing patents. We apply this index to the 2,139,314 utility patents granted in the U.S. between 1975 and 1999. We study the properties of this index by year and by technological category, and analyse the links between patents.

Keywords: Patents, Patent citations, Patent quality.

JEL Classification: C25, O31, O34

*Address for Correspondence: Department of Economics, University of Ottawa, P. O. Box 450, Station A, Ottawa, Ontario, Canada, K1N 6N5. E-mail address: gatallah@uottawa.ca.

1 Introduction

Patent citations have been used extensively as a measure of patent quality[1]: the more a patent is cited, the more impact it has on other inventors, and the more important it is. In addition to being indicative of links between patented innovations, citations are a trail of spillovers, as well as a measure of the importance of a patent (Hall et al., 2001). The idea behind patent citations is to go beyond patent counts, deemed a very imperfect measure of innovative output, insofar as different patents have different degrees of originality and innovativeness.

However, patent citations suffer from a similar handicap. Counting citations simply transposes the problem of patent counts to the level of the citing, rather than the cited patents. Patent citations analysis relies mainly on counting the citations received by patents, and using these citations as a measure of quality. However, this also amounts to counting patents: only the counted patents are those citing a given patent, rather than those produced by a given inventor. A patent may be receiving its citations from high quality patents, which themselves received/will receive a high number of citations. Or it may be receiving its citations from low quality patents, which are hardly cited at all. It is reasonable to believe that everything else being equal, a patent is deemed of higher quality if it receives its citations from high quality patents rather than from low quality patents. Using patent citations as a measure of quality equates citations from high quality patents with citations from low quality patents. Hence we end up counting citations of varying qualities, rather than patents of varying qualities. But the fundamental problem of controlling for the quality of patents remains.

We could go one step further, taking into account the quality of the citing patents, not only the number of citations received by a patent. For instance, we could give more weights to citations received by high quality patents (themselves receiving more citations). Trajtenberg et al. (2002) apply such a measure, where the citations received by the citing patents are "discounted" compared to the citations received by the original patent, to all university patents applied for in 1975 and in 1980. This is one step in the right direction, but it is only a partial solution, because the citations received by those (citing) patents themselves come from patents of differing quality, and so on.

The recursive nature of the problem starts to emerge. One way of accounting for the quality of patents is to take into account the quality of

[1] Examples include Jaffe et al. (2000) and Trajtenberg (1990).

all the patents citing those patents, all the patents citing the patents citing those patents, and so on. The idea is to use all the information available at any point in time about the quality of the citing patents, the only limit being the absence of information about future citations (the well known citations' truncation problem).

In this paper we propose a measure of patent quality which takes into account explicitly not only the number of citations received by a patent, but the quality of all the patents involved in the "chain" of citations starting with that patent. We argue that this measure makes a better use of the information available about patent citations, and is a more precise measure of the quality of patents than simply counting citations. We construct this measure for the 2,139,314 utility patents granted between 1975 and 1999 in the U.S., and for which citation data is available.

The analysis is organized as follows. First, we propose the new measure of patent quality in two versions: one where all citations receive equal weights, and one where citations closer to the initial patent receive more weight. Second, we analyse the properties of this indicator, and how it compares with the standard measure obtained by counting the "direct" citations received by a patent. Third, using a count model, we study the relationship between the proposed quality index and two characteristics of patents: their grant year and the technological category in which they are classified. Finally, we analyse the connectedness and the "distance" between patents.

2 An Index of Indirect Patent Citations

Let N be the total number of patents. Let $Q_0(x)$ represent the number of citations received by patent x. Let $Q_1(x)$ represent the number of citations received by the patents citing patent x. We have that

$$Q_1(x) = \sum_{i=1}^{N} \alpha_i(x) Q_0(i) \tag{1}$$

with $\alpha_i(x) = 1$ if patent i is citing patent x, and $\alpha_i(x) = 0$ otherwise. $Q_0(x)$ is the standard measure of quality of a patent as measured by the number of citations it receives. $Q_1(x)$ is the total number of citations received by all the patents citing patent x. Hence $Q_1(x)$ is a (partial) indirect measure of the quality of patent x. The higher $Q_1(x)$, the higher is the average quality of the patents citing patent x, and the higher is therefore the importance of patent x.

However, we need to go further and take into account the quality of the patents citing the patents citing patent x. For that, we define

$$Q_2(x) = \sum_{i=1}^{N} \alpha_i(x) Q_1(i) \tag{2}$$

where $Q_2(x)$ is the number of citations received by the patents citing the patents citing patent x. The higher this number is, the higher is the quality of the patents citing the patents citing patent x, the higher is the importance of the patents citing patent x, and hence the higher is the importance of patent x. Note that the citations of a given order to patent x are equal to the sum of the citations of the previous order for all the patents citing patent x.

To use all the information available about citations, the cumulative measure of citations needs to take into account all the direct and indirect citations that have been made to a patent. This cumulative measure is

$$Q_T(x) = \sum_{j=0}^{M} Q_j(x) \tag{3}$$

where $Q_j(x) = Q_0(x)$ when $j = 0$ and $Q_j(x) = \sum_{i=1}^{N} \alpha_i(x) Q_{j-1}(i)$ when $j \neq 0$, and where M is such that $Q_{M+1}(x) = 0$.

Therefore, $Q_T(x)$, which we call *cumulative patent citations*, is the sum of all direct and indirect citations received by patent x. It incorporates the number of citations received by patent x, received by the patents citing patent x, received by the patents citing those patents, etc. $Q_0(x)$ are the direct citations received by patent x, while $Q_j(x)$ are the indirect citations of order j received by patent x. M is the number of orders with a non nil number of patents citing x indirectly. The citations of order M received by patent x come from patents which themselves receive no citation at all: $Q_{M+1}(x) = 0$. Hence, M is where the chain of citations ends for patent x. Patents have different values of M, depending on the length of their chain of citations. In the data used here, the maximum value of $M = 29$.[2] While it is true that this index ultimately relies on counting patents, it has the advantage of using all the information available to assess the quality of a patent, and takes into account as much as is feasible the quality of the citing

[2] One could divide the number of citations received at any order by the number of patents receiving those citations (Q_j/Q_{j-1}), in order to get a measure of the "average" quality of the citing patent at that order. While useful, this measure does not account for the volume of citations.

patents. Using cumulative citations to account for the value of a patent takes into account the fact that the value of a patent is affected by the quality of the citations it receives.

There may seem to be some sort of double counting here. For instance, patent B may cite patent A, while patent C cites both patent A and patent B. In this case, patent A would benefit from the direct citation by patent B, from the direct citation by patent C (both would count in $Q_0(A)$) and from the indirect citation wherein C cites B: this last citation counts in $Q_1(A)$. However, insofar as this indirect citation reflects a higher quality of patent B, this is consistent with the objective of the index.

One weakness of $Q_T(x)$ is that it does not take into account the position of the citation in the chain: a citation received directly by patent x (i.e. a citation of order 0) is given the same weight as a citation of order 1, and the same weight as a citation of order 15. It can be argued that direct citations should receive more weight than indirect citations, and that, between two indirect citations, the citation closer to the patent should receive more weight. To account for the closeness of citations to the cited patent, we derive a weighted measure of cumulative citations, which we call *weighted cumulative patent citations*:

$$Q_w(x) = \sum_{j=0}^{M}(1 - \frac{j}{30})Q_j(x) \tag{4}$$

We see that the closer a citation is to the patent, the higher is the weight it gets. Direct citations get a weight of 1. The value 30 is chosen because in the sample studied here (see below) the maximum value of $M = 29$: this way the furthest citation gets a small but positive weight.

3 Analysis of Cumulative Citations

The indices derived above are now applied to citation data of utility patents granted in the U.S. between 1975 and 1999 (see Hall et al., 2001 for a description of the data)[3]. Table 1 and figure 1a (all figures except 1b are in logarithmic scale) present the average values of Q_T by year[4] and by technological category. The six technological categories are: 1) Chemical; 2)

[3] We eliminated from the data all patents prior to 1975. This is because the citation data (for citing patents) starts only with patents granted in 1975; using the whole sample would introduce a bias, since citations made by patents between 1963 and 1974 would not be incorporated.

[4] Throughout this paper we use the grant year as the relevant date for patent analysis. While the application year is another candidate, and may well have some advantages over

Computers and Communications; 3) Drugs and Medical; 4) Electrical and Electronics; 5) Mechanical; and 6) Others.

The average value of Q_T for all patents is 204,493.1 citations.[5] In general citations decline over time, which reflects the cumulative nature of the index: earlier patents have more time to accumulate indirect citations than later ones (the truncation problem). However, for a few years and for some technological categories, the patents of a year have more citations than the patents of the previous year. This is true at the technological category level, however; it is never true at the level of all patents.

Category 2 is constantly above the yearly average, while categories 5 and 6 are constantly below. A notable feature is that patents of category 1 had the most citations until 1977, and then decline faster than other categories, so that for later years they have the worst performance. From 1978 on, category 2 took the lead, and by far, in terms of citations. Hence even though the overall average of category 1 is the highest, this is due mainly to good performance between 1975 and 1977, and hides a less than average performance later on. Part of this evolution can be explained by the change in the share of each category in patents over time. The three traditional fields (Chemical, Mechanical and Others) declined over the last three decades (Hall et al., 2001), explaining the decline in cumulative citations in those fields. Computers and Communications, as well as Drugs and Medical, have increased, while Electrical and Electronics has been steady as a percentage of total patents. Hence part of the exceptional performance of the Computers and Communications category is due to the increase in patents in that category, which increases the pool of citing patents. Moreover, the "general purpose technology" character of patents in that category also increases their quality as measured by cumulative citations.

It is useful to compare the results obtained using cumulative citations with those obtained using only direct citations. Table 2 and figure 1b show Q_0 (direct citations) per year and per category. Q_T declines more uniformly than Q_0, reflecting the built-in time bias of Q_T. According to Q_0, categories 2, 3, and 4 have above average citations, while the three other categories

the grant year (see Hall et al., 2001), a patent can start receiving citations only after it is granted.

[5] After the calculations have been made, an error has been discovered in the initial data. The patent #5489070 is reported to have cited itself, which is impossible. The error has been corrected for that patent, but was not corrected for the patents cited by that patent, nor for the patents cited by those patents, etc. However, these errors are unlikely to have a significant impact on the results, as the patent affected by the original data entry error has made and received a small number of citations.

are below average. However, according to Q_T, only categories 1 and 2 are above average. With Q_T the overall average is somewhat inflated because of the very high cumulative citations obtained in the early years by category 1. These different rankings of category 1 emphasize that category 1 patents receive less direct citations than patents from other categories (2, 3, and 4, for instance); however, these citations are made -cumulatively- by high quality patents, which receive a large number of citations. Also, the only category doing well (above average) on both measures, direct citations (Q_0) and cumulative citations (Q_T), is category 2: this is the only category where patents receive on average a large number of direct citations from high quality patents.

Reading Tables 1 and 2 jointly, a major difference between categories 5 and 6 can be observed. Using direct citations, category 5 has the worst overall performance for most of the time, and category 6 patents receive more direct citations on average than category 5 patents, until the 1990s. However, using cumulative citations, category 6 is the worst performer for all years. Hence while category 5 obtains less citations (until recently) than category 6, those citations are made by higher quality patents.

Table 3 and figure 1c present the average values of the weighted index, Q_w by year and category. The overall values are significantly lower than for Q_T. For instance, the average for all patents is 126,373.7, less than two-thirds of the average of the unweighted index. This decline reflects the lower weight given to cumulative citations. However, as the comparison of figures 1a and 1c shows, the overall shape of the distribution of citations between categories is not much affected by the introduction of the weights. Hence the simple addition of cumulative citations does not seem to introduce a significant bias in comparing the relative quality of patents, at least when the analysis is performed at the aggregate level. The results regarding the comparison of individual patents may be more sensitive to the weights.

Table 4 details the orders of citations by technological category. This allows us to see which orders of citations carry more weight in the final index. The distribution of citations between orders has an inverted U-shape, which means that as the order of citations increases, the number of citations first increases and then declines. Overall the bulk of citations is between Q_8 and Q_{15} (in this range average cumulative citations per patent are in the 5 digits). For category 3 the bulk of citations comes from lower orders, namely between Q_4 and Q_{10}, while for category 1 the bulk is spread between Q_8 and Q_{17}. Category 3 has a very good performance for lower orders, but as soon as it reaches Q_{11} and on has the worst performance. This means that patents in that category receive a high number of cumulative citations from high quality patents for a few orders, but then these patents are not highly cited later on. The chain of citations seems quite short for that category. Overall, the tail for lower orders is much thicker than the tail for higher orders.

Table 5 shows the citations of different orders by year. Citations of most orders decline over time, although there are occasional increases, especially during the 1990s. A notable feature is the very slow decline of Q_2 over the period 1975-1987. Earlier years have their peak order of citations later, with 1975 and 1976 having their peak at Q_{12}. Later years have their peak earlier, as the patents citing them have not had much time to get cited. The years 1996 and onward have their peak at Q_0. Overall the highest average is for citations of order Q_{12}. The concave shape of citations by year indicate that as the order increases, citations increase at a decreasing rate, and decrease at an increasing rate.

4 Empirical Evidence on the Effects of Categories and Years on Direct and Indirect Citation Indexes

4.1 The Methodology

The data used in this paper falls in the category of count models. For a survey about the different classes of econometric specifications applied to count models, see, for example, Cameron and Trivedi (1986), Gurmu and Trivedi (1994). We use the negative binomial model, which relaxes the Poisson assumption that the mean equals the variance that is considered a shortcoming of the Poisson regression model. The negative binomial model arises from a formulation of cross-section heterogeneity (Greene, 2002). Following Greene (2002), we can show the derivation of a form of the negative binomial distribution. In this model, the parameter λ_i is related to the regressors x_i. The basic equation of the model is

$$\Pr(Y_i = y_i | x_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \tag{5}$$

for $y_i = 0, 1, 2, ...$; and the most used formulation for λ_i is the so named loglinear model which is expressed by

$$\ln \lambda_i = x_i'\beta. \tag{6}$$

>From the last expression we can see that

$$\begin{aligned} E[y_i|x_i] &= Var[y_i|x_i] \\ &= \lambda_i \\ &= e^{x_i'\beta}. \end{aligned}$$

To obtain a negative binomial distribution, we need to generalize the Poisson model by including an individual, unobserved effect into the conditional mean:

$$\begin{aligned} \ln \mu_i &= x_i'\beta + \epsilon_i & (7) \\ &= \ln \lambda_i + \ln u_i, & (8) \end{aligned}$$

where the disturbance term considers either specification error as in the classical regression model or some kind of cross-sectional heterogeneity commonly arising in microeconomic data. The distribution of y_t conditioned on x_i (and ϵ_i) is yet Poisson with conditional mean and variance given by

$$f(y_i|x_i, u_i) = \frac{e^{-\lambda_i u_i}(\lambda_i u_i)^{y_i}}{y_i!}. \tag{9}$$

The unconditional distribution $f(y_i|x_i)$ is the expected value of $f(y_i|x_i, u_i)$ over the domain of u_i, which is equivalent to

$$f(y_i|x_i) = \int_0^\infty \frac{e^{-\lambda_i u_i}(\lambda_i u_i)^{y_i}}{y_i!} g(u_i) du_i. \tag{10}$$

>From the last expression, it is clear that the choice of the density for u_i defines the unconditional distribution. A few suggestions appear in the literature such as the normal-Poisson mixture proposed by Greene (1995, 1997) and Terza (1995). One shortcoming is that there is no closed form for this specification, although approximations are possible. One distribution that offers a mathematical convenience is the Gamma distribution. It is usually assumed for $u_i = \exp(\epsilon_i)$. It is also assumed that $E[\exp(\epsilon_i)] = 1.0$ because of the unidentification of the mean when a constant is included in the regression. Using this normalization, we have

$$g(u_i) = \frac{\theta^\theta}{\Gamma(\theta)} e^{-\theta u_i} u_i^{\theta - 1}. \tag{11}$$

Consequently, the density for y_i is

$$
\begin{aligned}
f(y_i|x_i) &= \int_0^\infty \frac{e^{-\lambda_i u_i}(\lambda_i u_i)^{y_i}}{y_i!} \frac{\theta^\theta}{\Gamma(\theta)} e^{-\theta u_i} u_i^{\theta-1} du_i \\
&= \frac{\theta^\theta {\lambda_i}^{y_i}}{\Gamma(y_i+1)\Gamma(\theta)} \int_0^\infty e^{-(\lambda_i+\theta)u_i} u_i^{\theta+y_i-1} du_i \\
&= \frac{\theta^\theta {\lambda_i}^{y_i}\Gamma(\theta+y_i)}{\Gamma(y_i+1)\Gamma(\theta)(\lambda_i+\theta)^{\theta+\lambda_i}}.
\end{aligned}
$$

Defining $r_i = \lambda_i/(\lambda_i+\theta)$, the last expression can be written as

$$
f(y_i|x_i) = \frac{\Gamma(\theta+y_i)}{\Gamma(y_i+1)\Gamma(\theta)} r_i^{y_i}(1-r_i)^\theta,
$$

which is one form of the negative binomial distribution as it appears in Cameron and Trivedi (1986). The conditional mean is λ_i and the conditional variance is $\lambda_i(1+(1/\theta)\lambda_i)$.

The method of estimation used is Quasi-Maximum Likelihood (QML) which is robust in the sense that it produces consistent estimates of the parameters of a correctly specified conditional mean, even if the distribution is incorrectly specified. Further details on QML estimation are provided by Gourieroux, Monfort and Trognon (1994a, 1994b). See also Wooldrige (1990) for a good summary of the use of QML techniques in estimating parameters of count models. Finally, standard errors have been corrected for overdispersion. It means that a consistent estimate of the covariance is obtained by imposing the Generalized Linear Models (GLM) condition that the true variance of the dependent variable is proportional to the variance of the distribution used in specifying the log likelihood. Further details can be found in McCullaugh and Nelder (1989) and Fahrmeir and Tutz (1994).

4.2 Empirical Results

As explained above, we use count models to identify the effects of categories and years on Q_0 and Q_T[6]. More specifically, the dependent variable is either Q_0 or Q_T. For the set of explanatory variables, we consider different steps in the estimation procedure. Firstly, we estimate the model using a dummy variable for each year in the sample, which allows us to have estimations by categories. Secondly, we perform similar estimations but using dummy variables for each category as the explanatory variables. In this case, we have

[6] We thank William Greene for advice on this issue.

estimations for each year of the sample. Finally, we consider an estimation using dummy variables for each year and for each category in the sample[7]. In the estimations, the variable $Category_i = 1$ if observation i belongs to the category i; and 0 otherwise. In a similar fashion, $Year_i = 1$ is the observation i corresponding to the year i. Because all estimations include a constant, we introduce the total number of dummies less two dummies. In this case, the coefficients related to the dummy variables are interpreted as values with respect to the dummy variable used as a *"base"*. In the case of categories dummy variables, we use the sixth category as the base-category. In the case of year dummy variables, it is the year 1975 which is used as the year-base variable.

Results of the estimations by year are shown in Table 6. The following comments are of note here. For all cases, the LR index or Pseudo-R^2 is higher for equations where Q_T is the dependent variable, suggesting that the categories dummies have more explanatory power with Q_T than with Q_0. Observing the effects of each category on the dependent variable, we can establish a ranking between categories with respect to the base-category variable. Overall, we observe that the ranking of categories is more stable (from year to year) when Q_0 is the dependent variable. In fact, according to these estimations, the third category (*Drugs and Medical*) presents higher effects on the dependent variable. Next come categories 2 (*Computers*) and 4 (*Electrical and Electronics*). On the other hand, notice that categories 1 (*Chemicals*) and 5 (*Mechanical*) present negative coefficients, which means that these categories have a lower effect on the dependent variable compared to the category-base. Magnitudes of these effects have changed for some years as for example category 5 presents positive effects in 1990 and after 1992. The category with higher effects (category 3) presents negative effects from 1998. The "ranking" of the effects change from 1997, where we observe that category 2 is now the category presenting the highest values and categories 4 and 5 are the following.

When the dependent variable is Q_T a less stable "ranking" is observed. Overall, it is the category 2 which presents the higher values of the effects on the dependent variable with respect to the category-base. After this category, we observe categories 3 and 4 or vice versa as the highest values. The fourth and fifth positions are occupied by categories 1 and 5. Notice that from 1996 and on, category 3 loses importance in the number of total citations while category 4 presents a better performance. We also observe

[7] Notice that, in this case, no estimation for Q_0 was possible given a singularity problem in the regressors matrix.

that categories 1 and 3 present negative effects.

Summarizing the above results we can establish that either categories 2 and 3 present the higher effects on the number of citations. It is interesting to note the differences using the two different dependent variables. In fact, both variables suggest different categories as being the most important to explain the number of citations. In the case of Q_0, it is the category *Drugs and Medical* which presents the most important effect on the number of citations. However, when we use Q_T as the dependent variable, it is the category *Computers* that presents the most important effects. This result means that the category of *Drugs and Medical* can explain the direct number of patent citations , but it is the category of *Computers* which is able to explain the total number of patent citations. It would indicate that this last category is more informative in terms of the value of a patent in comparison with the other category. More importantly, these results indicate that the patents related to *Computers* received more total citations compared to the category of *Drugs and Medical*. A similar dichotomy regarding the ranking of categories 5 and 6, depending on whether we use Q_0 or Q_T, was noted in the descriptive analysis in section 3, and is also observed here: category 6 performs better on direct citations until the 90s, while category 5 performs better on cumulative citations for the whole period.

Table 7 presents the results of the estimations by category. In this case we are interested in the contribution of each year-dummy variable within each category. The first observation is related to the effect of the truncation of the data. Due to truncation, estimates are essentially negative given that we are using the year 1975 as the base-year. Hence, as we are advancing toward the end of the sample, the effects are, not surprisingly, negative. Estimations using Q_0 as the dependent variable show positive coefficients for categories 1 and 2 until 1988 and 1993, respectively. This indicates that even when 1975 is the year base, these categories present higher direct citations for these years. What is also observed is the fact that until around 1990, the effects of years are very close to zero, indicating an effect very similar to the base-year used in the estimations. The negative effects are more clearly observed and are highly negative from 1991 approximately.

When Q_T is used as the dependent variable, similar observations are obtained. However, the effects are always negative due to truncation, the effect of which is even stronger with Q_T than with Q_0. It is also possible to observe the evolution of these effects, namely, if there are abrupt changes in the negative effects from year to year. What we observe in this respect is that there are less important changes in these estimates for categories 2 and 3. In the other categories, abrupt changes in these coefficients from year to year are observed.

Table 8 presents the results using both sets of dummy variables used separately until now. We present only the result for Q_T as the dependent variable. Observation of the dummy of categories confirms that category 2 (*Computers and Communications)* shows the higher effects on total number of citations. Categories 3 (*Drugs and Medical*), 4 (*Electrical and Electronics*), 1 (*Chemicals*) and 5 (*Mechanical*) follow in that order. The dummy variables associated with each year show, as before, the evolution of the effects on the total number of citations of each year. According to the results, there are some years where the change in the (negative) effect is slower than in other years. Such is the case, for example, for 1978-1979, 1981-1982, 1983-1984, and 1991-1992. From 1996 until the end of the sample, the effects are more abrupt.

5 Links Between Patents

Another way of looking at the relationship between patents is by analysing the links between them. Patents can be viewed as elements of a network, with the citations constituting the links between them. A patent is linked directly to another patent through a direct citation, and indirectly through an indirect citation. As shown above, indirect citations can be of different orders, and hence a patent can be said to be more or less closely related to another patent (and through different channels, i.e. different citations' chains). The longer the chain of citations of a patent, the more is the impact of that patent and the patents citing it (and so on) spread through time. Hence a longer chain of citations is indicative of continuity of the impact of an innovation.

While the full analysis of the patents network goes beyond the scope of this paper, it is useful to examine to what extent patents are linked directly and indirectly to other patents. The analysis performed above of the citations by their different orders by year constitutes a preliminary step in that direction. However, the above analysis was performed in terms of averages and in term of relationships. In this section we look more closely at the relationships between patents by analysing the number of patents obtaining citations of every order.

Table 9 presents the number of patents having non-zero citations for each order and for each year. For any given year, the number of patents receiving citations decreases as the order of the citation increases (this is true by construction of the index). Moreover, for any given order, the number of patents receiving citations of that order decreases as the patents are more recent. Citations of higher orders are rather selective. Whereas more than 75% of all patents receive at least 1 citation (of order 0, i.e. a direct citation), less than a quarter of all patents receive citations of order 6 or higher, and less than 1% receive citations of order 17 or higher. Only 3 patents receive citations of order 29. To what extent the length of the chain of citations of a patent is important to its quality may depend on the type of analysis performed.

This can also be seen from the histograms in figures 2 through 4. Whereas the mass of patents have a Q_0 between 5 and 12 (figure 2), the mass of patents have a Q_T (figure 3) and a Q_w (figure 4) close to zero, with a very thin tail.

6 Concluding Remarks

In this paper cumulative citations and the length of citation chains were used to analyse patents at an aggregate level. From the point of view of more specific evaluations of patents, the index can be used to assess the quality of any individual patent or a portfolio of patents of an innovator or a firm, and compare it to a relevant benchmark (e.g. the fixed-effects approach; see Hall et al., 2001).

Because the analysis of cumulative citations favours most the older patents, the analysis may be particularly useful for analysing the relative quality of patents of similar periods, or over a fixed window. Although this problem is also encountered with the standard measure of (direct) citations, it is accentuated here because of the stronger time bias of cumulative citations.

One dimension of the analysis of patent citations is the originality and generality of patents. This requires analysing the technological categories to which the (directly) citing patents belong. When indirect citations are incorporated, such an analysis becomes more difficult, as indirect citations are very large in number. However, it should still be possible to extend these analyses to indirect citations.

The idea of cumulative citations can be extended to other types of citations, such as the study of scientific networks. For any type of citation, accounting for the quality of citations, not only their numbers, provides a more accurate evaluation of the quality of patents.

References

[1] Cameron, C., and P. Trivedi (1986), "Econometric Models Based on Count Data: Comparison and Applications of Some Estimators and Tests," *Journal of Applied Econometrics* **1**, 29-54.

[2] Fahrmeir, L., and G. Tutz (1994), *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer.

[3] Gourieroux, C., A. Monfort, and C. Trognon (1984a). "Pseudo-Maximum Likelihood Methods: Theory," *Econometrica* **52**, 681-700.

[4] Gourieroux, C., A. Monfort, and C. Trognon (1984b). "Pseudo-Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica* **52**, 701-720.

[5] Greene, W. (1995), *LIMDEP, Version 7.0: User's Manual*. Bellport. N.Y. Econometric Software, 234-241.

[6] Greene, W. (1997), "Frontier Production Functions," In Pesaran, M. and P. Schmidt (Eds.), *Handbook of Applied Econometrics: Volume 2: Microeconomics*. London: Blackwell Publishers.

[7] Greene, W. (2002), *Econometric Analysis*, Prentice Hall.

[8] Gurmu, S., and P. Trivedi (1994), "Recent Developments in Models of Events Counts: A Survey," Manuscript, Department of Economics, Indiana University.

[9] Hall, B., Jaffe, A., and Trajtenberg, M. (2001), "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools," Manuscript, September.

[10] Jaffe, A. B., Trajtenberg, M., and Fogarty, M. S.(2002), "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors," *American Economic Review* **90 (2),** 215-18.

[11] McCullagh, P. and J. A. Nelder (1989), *Generalized Linear Models*, 2nd Edition, Chapman & Hall.

[12] Terza, J. (1995), "Estimating Count Data Models with Endogenous Switching and Sample Selection," Working Paper IPRE-95-14, Department of Economics, Penn State University.

[13] Trajtenberg, M. (1990), "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *RAND* **21 (1),** 172-87.

[14] Trajtenberg, M., Henderson, R., and Jaffe, A. B. (2002), "University versus Corporate Patents: A Window on the Basicness of Invention", in Jaffe, A. B., and Trajtenberg, M. (eds.), *Patents, Citations and Innovations: A window on the knowledge economy*, The MIT Press, Cambridge, M.A.

[15] Wooldridge, J. M. (1990), "Quasi-Likelihood Methods for Count Data," in M. Hashem Pesaran and P. Schmidt (Eds.) *Handbook of Applied Econometrics*, Volume 2, Blackwell, 352-406.

Table 1. Average Q_T by Year and Category

Year	Category					
	1	2	3	4	5	6
1975	7831408.4	5639149.6	486918.3	1384580.6	1147392.9	111185.7
1976	5553340.3	3423221.4	301677.2	967191.4	483177.6	65391.3
1977	1877162.4	1856307.8	170805.4	573110.6	453306.6	48955.8
1978	425235.6	1135114.2	132516.1	297281.7	181576.4	76295.9
1979	177170.8	652723.1	90959.5	288338.9	103400.7	11054.0
1980	103097.3	415692.6	46858.8	94660.2	48256.4	6966.5
1981	36911.5	243825.0	47661.2	69473.9	24786.4	4659.3
1982	23571.9	169710.9	27704.0	36266.8	16207.7	4273.9
1983	11966.8	103758.3	14811.3	22524.6	9618.3	2431.5
1984	4364.1	70363.5	15373.5	12723.2	5580.2	1546.4
1985	1974.6	35525.3	10101.6	7883.4	3050.6	1079.2
1986	1189.7	16812.4	7185.7	4224.1	1493.1	675.0
1987	698.6	9231.6	3401.0	2150.5	1121.8	496.3
1988	453.2	4183.9	1819.4	1158.2	587.8	265.5
1989	191.5	1684.2	891.8	470.7	343.3	123.6
1990	101.5	892.1	470.8	316.0	146.7	81.9
1991	71.9	429.2	322.8	143.7	66.2	46.2
1992	31.7	241.0	114.3	76.0	40.9	27.0
1993	16.3	111.4	49.7	38.7	20.3	14.7
1994	11.3	54.2	20.5	18.7	10.6	8.6
1995	4.8	23.2	8.6	9.7	5.6	4.6
1996	6.3	9.7	3.5	4.4	3.0	2.3
1997	1.0	4.0	1.3	2.1	1.4	1.1
1998	0.3	1.8	0.4	0.5	0.4	0.4
1999	0.0	0.0	0.0	0.0	0.0	0.0

Table 2. Average Q_0 by Year and Category

Year	Category					
	1	2	3	4	5	6
1975	6.17	8.65	10.20	6.78	5.80	6.40
1976	6.44	9.25	9.59	6.82	5.97	6.58
1977	6.57	10.10	9.10	7.23	5.95	6.73
1978	6.75	10.64	8.56	7.27	5.87	6.57
1979	6.76	10.11	9.27	7.32	5.90	6.42
1980	6.46	10.62	9.30	7.17	5.75	6.24
1981	6.77	10.86	9.15	7.28	5.85	6.22
1982	6.63	11.28	10.02	7.21	5.91	6.26
1983	6.72	11.56	10.14	7.26	5.96	6.24
1984	6.72	12.66	10.14	7.24	5.70	6.13
1985	6.72	11.91	10.09	7.40	5.71	6.18
1986	6.67	11.75	10.91	7.27	5.80	6.07
1987	6.59	12.07	11.46	7.38	5.80	6.08
1988	6.27	11.81	10.40	7.12	5.63	6.00
1989	5.82	11.18	9.69	6.79	5.20	5.37
1990	5.33	11.18	9.20	6.63	4.97	4.97
1991	4.84	10.26	8.64	6.14	4.58	4.66
1992	4.43	10.06	7.83	5.69	4.24	4.23
1993	3.73	9.17	6.52	5.23	3.72	3.69
1994	3.17	7.92	5.47	4.37	3.13	3.08
1995	2.37	6.05	3.85	3.50	2.50	2.40
1996	1.61	4.43	2.40	2.47	1.74	1.63
1997	0.85	2.45	1.09	1.40	0.99	0.90
1998	0.32	0.87	0.33	0.51	0.39	0.34
1999	0.03	0.06	0.02	0.05	0.03	0.03

Table 3. Average Q_w by Year and Category

Year	Category					
	1	2	3	4	5	6
1975	4519803.6	3751025.0	360053.8	893274.0	689296.0	75975.4
1976	3301421.6	2375064.2	224252.3	630834.8	307360.6	45703.6
1977	1163900.6	1321068.9	128564.1	387255.8	291125.2	35189.5
1978	275141.0	825035.1	100327.4	204867.6	120319.8	50827.9
1979	117894.6	479488.4	70413.9	199276.3	70694.4	8485.5
1980	70997.0	311111.4	37209.0	68925.6	34249.5	5383.8
1981	26247.6	184956.6	37497.3	51281.9	18090.8	3701.1
1982	17232.1	130304.1	22271.7	27343.2	12148.9	3407.1
1983	8986.1	81469.5	12118.0	17475.8	7382.1	1992.6
1984	3428.8	56228.2	12635.1	10024.1	4359.0	1287.0
1985	1611.5	28817.9	8401.8	6352.6	2444.5	903.6
1986	985.7	13887.1	6047.8	3467.2	1229.2	570.4
1987	593.4	7755.0	2929.5	1808.7	927.1	423.4
1988	393.1	3591.5	1587.9	991.7	494.4	232.3
1989	169.0	1473.7	786.9	413.1	293.6	111.4
1990	91.3	795.2	422.3	279.4	129.7	73.3
1991	65.5	387.6	291.5	130.6	60.6	42.4
1992	29.6	220.7	106.1	69.8	37.8	25.2
1993	15.6	104.2	46.8	36.4	19.2	14.1
1994	10.5	51.4	19.7	17.9	10.2	8.3
1995	4.7	22.2	8.4	9.4	5.4	4.5
1996	5.4	9.5	3.5	4.4	2.9	2.3
1997	1.0	3.9	1.3	2.0	1.4	1.1
1998	0.37	1.61	0.38	0.56	0.44	0.43
1999	0.027	0.062	0.022	0.049	0.033	0.028

Table 4. Average Q_i by Category

	Category					
	1	2	3	4	5	6
0	4.7	6.8	5.8	5.0	4.3	4.4
1	20.4	42.7	33.1	24.0	17.0	17.0
2	72.8	224.5	148.4	97.9	59.5	53.5
3	213.0	956.2	509.4	338.8	181.4	138.8
4	531.8	3237.6	1330.1	986.2	482.9	301.5
5	1245.6	8623.1	2656.7	2405.0	1126.1	557.9
6	3030.5	17980.5	4102.7	4915.6	2293.5	893.3
7	7714.8	29360.5	4946.4	8443.5	4071.9	1254.5
8	18839.4	37784.5	4680.6	12246.7	6345.7	1563.7
9	40436.1	38795.0	3478.8	15114.9	8845.9	1739.6
10	72452.0	32514.3	2027.2	16046.4	11294.3	1709.9
11	105482.9	23309.9	927.6	14859.0	13316.7	1443.2
12	122958.5	15470.9	341.6	12145.8	14139.1	1011.9
13	113886.9	10191.6	110.0	8759.5	12847.9	577.3
14	83686.7	6549.2	36.5	5465.1	9463.2	269.6
15	49179.8	3776.1	14.0	2854.8	5395.9	110.2
16	23967.0	1838.4	5.5	1211.3	2299.6	48.2
17	10922.3	756.3	1.9	413.2	720.5	28.2
18	5857.8	281.8	0.6	116.9	174.0	19.7
19	3980.6	106.8	0.1	30.0	43.3	12.6
20	2779.2	43.9	0.0	7.6	17.1	6.4
21	1669.5	19.0	0.0	1.8	8.2	2.4
22	802.2	7.9	0.0	0.4	3.2	0.6
23	298.8	2.8	0.0	0.0	0.9	0.1
24	84.3	0.8	0.0	0.0	0.2	0.0
25	17.5	0.1	0.0	0.0	0.0	0.0
26	2.6	0.0	0.0	0.0	0.0	0.0
27	0.2	0.0	0.0	0.0	0.0	0.0
28	0.0	0.0	0.0	0.0	0.0	0.0
29	0.0	0.0	0.0	0.0	0.0	0.0

Table 5 - Average Qi by year

Order

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
1975	6.54	42.68	239.78	1122.33	4271.17	13085.31	32466.45	66840.83	119201.45	193392.55	292038.09	400645.58	474625.88	463911.04
1976	6.73	43.60	241.86	1110.59	4133.99	12385.83	30084.74	60701.65	106217.10	168245.76	243302.77	310376.85	332605.19	288491.54
1977	6.92	44.86	247.45	1111.19	3976.75	11240.92	25253.63	46254.91	72046.73	99250.47	122077.00	130639.42	116875.69	84531.90
1978	6.91	43.78	232.62	994.18	3325.76	8632.37	17494.20	28314.15	38086.27	44381.00	45646.58	40832.92	30658.45	18638.01
1979	6.92	43.85	227.92	928.05	2912.94	7030.46	13176.28	19685.69	24405.34	26105.17	24401.76	19557.25	12963.38	6886.52
1980	6.81	42.80	215.83	838.37	2476.46	5546.00	9520.56	12884.24	14242.29	13241.82	10477.06	6964.64	3782.93	1656.32
1981	6.90	42.56	205.79	749.32	2042.45	4176.41	6489.32	7868.53	7700.37	6290.29	4405.96	2650.34	1337.66	558.78
1982	7.05	43.50	205.43	722.03	1880.92	3641.98	5303.97	5936.68	5249.56	3773.98	2265.83	1156.64	511.32	203.36
1983	7.10	43.36	197.27	654.66	1579.87	2779.23	3610.80	3533.96	2674.16	1627.32	830.42	366.00	144.17	56.57
1984	7.08	42.46	188.65	599.46	1335.12	2093.60	2358.27	1960.52	1245.75	638.90	285.74	121.37	50.45	19.15
1985	7.11	40.96	169.72	493.85	992.24	1380.76	1347.13	950.00	511.47	228.94	95.67	41.35	17.55	6.97
1986	7.17	39.65	151.72	392.63	685.34	810.68	663.25	392.19	182.07	78.34	36.83	17.65	7.09	2.62
1987	7.33	39.47	138.96	315.22	466.49	465.82	328.14	173.35	76.01	31.41	13.64	5.65	1.85	0.59
1988	7.09	34.53	107.07	210.05	266.19	219.17	122.58	53.77	21.72	8.59	3.58	1.48	0.51	0.14
1989	6.67	28.77	74.06	116.28	116.38	77.73	38.02	15.78	6.24	2.70	1.14	0.38	0.10	0.02
1990	6.34	24.53	53.62	69.27	55.92	29.15	10.64	4.01	2.41	1.68	0.90	0.31	0.07	0.02
1991	5.87	19.43	35.07	38.29	24.87	8.88	2.40	1.02	0.67	0.39	0.17	0.07	0.03	0.01
1992	5.48	15.20	21.32	16.65	7.60	2.31	0.69	0.39	0.38	0.29	0.16	0.06	0.01	0.00
1993	4.90	10.73	11.19	6.17	1.88	0.43	0.11	0.03	0.01	0.00	0.00	0.00	0.00	0.00
1994	4.22	6.88	4.96	1.74	0.36	0.11	0.10	0.11	0.09	0.05	0.01	0.00	0.00	0.00
1995	3.30	3.50	1.59	0.35	0.07	0.04	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00
1996	2.34	1.42	0.33	0.06	0.06	0.11	0.16	0.17	0.13	0.06	0.02	0.00	0.00	0.00
1997	1.28	0.37	0.04	0.03	0.03	0.03	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
1998	0.48	0.04	0.01	0.01	0.01	0.02	0.03	0.05	0.04	0.02	0.01	0.00	0.00	0.00
1999	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
All	4.94	23.24	94.21	323.49	932.11	2248.44	4566.95	7962.64	12358.66	17780.68	24043.39	29696.42	31871.83	28468.53

Table 5 - Average Qi by year (cont.)

Order

14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
363692.88	226025.89	112715.83	48784.21	22839.70	14165.28	9972.66	6298.25	3212.01	1270.62	380.41	83.88	13.10	1.35	0.08	0.00
198617.24	108430.33	48850.73	21243.00	11758.37	8332.95	5759.96	3289.42	1459.96	489.03	120.45	20.99	2.43	0.16	0.00	0
48502.57	22343.50	9184.53	4446.37	2941.48	2062.56	1227.72	571.74	201.44	52.28	9.62	1.18	0.08	0.00	0	0
9010.12	3615.55	1491.73	852.40	590.01	365.94	178.44	65.33	17.48	3.29	0.41	0.03	0.00	0	0	0
2935.62	1118.65	514.46	329.97	219.50	120.21	49.87	15.00	3.15	0.44	0.04	0.00	0	0	0	0
624.06	261.30	158.11	110.06	64.72	28.27	8.72	1.85	0.25	0.02	0.00	0	0	0	0	0
205.87	85.67	49.20	30.13	15.14	5.54	1.35	0.21	0.02	0.00	0	0	0	0	0	0
79.98	37.53	21.52	11.33	4.37	1.12	0.18	0.02	0.00	0	0	0	0	0	0	0
27.74	17.47	10.38	4.66	1.46	0.30	0.04	0.00	0	0	0	0	0	0	0	0
7.48	3.56	1.57	0.50	0.10	0.01	0.00	0	0	0	0	0	0	0	0	0
3.03	1.26	0.37	0.07	0.01	0.00	0	0	0	0	0	0	0	0	0	0
0.96	0.29	0.06	0.01	0.00	0	0	0	0	0	0	0	0	0	0	0
0.20	0.04	0.01	0.00	0	0	0	0	0	0	0	0	0	0	0	0
0.03	0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	0	0	0
0.00	0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	0	0	0
0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0
0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0
0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0
0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0
0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0
0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0
0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0
0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0
0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0
0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0
20613.07	11997.20	5742.18	2513.28	1270.15	828.29	569.12	339.82	162.79	60.52	17.06	3.55	0.52	0.05	0.00	0.00

Table 6. Estimations by Year

Variable	1975				1976				1977			
	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE
Constant	1.841	0.009	12.363	0.336	1.882	0.009	11.928	0.253	1.912	0.009	12.616	0.415
Category 1	-0.014	0.012	3.052	0.464	-0.015	0.013	3.600	0.345	-0.022	0.013	1.648	0.574
Category 2	0.279	0.019	3.754	0.748	0.287	0.020	2.891	0.554	0.338	0.020	1.654	0.882
Category 3	0.434	0.022	0.701	0.831	0.315	0.022	0.723	0.603	0.237	0.021	-0.649	0.934
Category 4	0.066	0.014	1.713	0.537	0.059	0.015	1.969	0.397	0.078	0.015	0.587	0.643
Category 5	-0.047	0.012	2.515	0.460	-0.077	0.013	1.189	0.350	-0.106	0.013	0.731	0.572
LR Index (Pseudo-R^2)	0.002		0.035		0.002		0.052		0.002		0.015	

Table 6. Estimations by Year

Variable	1978				1979			
	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE
Constant	1.878	0.009	10.778	0.222	1.854	0.011	10.292	0.303
Category 1	0.035	0.013	2.132	0.309	0.064	0.016	1.853	0.425
Category 2	0.435	0.020	2.932	0.482	0.401	0.023	2.879	0.649
Category 3	0.214	0.022	0.879	0.529	0.285	0.025	1.050	0.684
Category 4	0.112	0.015	1.983	0.353	0.156	0.017	2.196	0.479
Category 5	-0.076	0.013	1.689	0.310	-0.041	0.016	1.443	0.424
LR Index (Pseudo-R^2)	0.002		0.023		0.002		0.024	

Table 6. Estimations by Year

Variable	1980				1981				1982			
	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE
Constant	1.835	0.01	9.937	0.218	1.838	0.009	9.253	0.167	1.838	0.011	8.974	0.154
Category 1	0.043	0.014	1.468	0.309	0.073	0.013	1.247	0.231	0.073	0.015	1.060	0.216
Category 2	0.472	0.021	2.832	0.464	0.521	0.021	2.999	0.364	0.521	0.021	2.821	0.307
Category 3	0.323	0.022	0.682	0.474	0.401	0.020	1.300	0.362	0.401	0.023	1.125	0.326
Category 4	0.151	0.016	1.573	0.350	0.150	0.015	1.867	0.265	0.150	0.016	1.676	0.236
Category 5	-0.064	0.015	1.128	0.308	-0.036	0.014	1.129	0.235	-0.035	0.015	1.029	0.216
LR Index (Pseudo-R^2)	0.003		0.024		0.004		0.028		0.004		0.027	

Table 6. Estimations by Year

Variable	1983				1984			
	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE
Constant	1.826	0.011	8.286	0.126	1.817	0.010	8.005	0.128
Category 1	0.088	0.015	1.152	0.175	0.096	0.014	0.345	0.184
Category 2	0.558	0.022	3.151	0.249	0.664	0.020	3.024	0.263
Category 3	0.416	0.024	1.064	0.271	0.427	0.021	1.418	0.278
Category 4	0.182	0.017	1.796	0.193	0.178	0.015	1.518	0.198
Category 5	-0.001	0.015	1.015	0.175	-0.034	0.014	0.859	0.181
LR Index (Pseudo-R^2)	0.004		0.037		0.006		0.043	

Table 6. Estimations by Year

Variable	1985				1986				1987			
	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE
Constant	1.821	0.009	7.312	0.084	1.802	0.009	6.782	0.083	1.804	0.009	6.403	0.093
Category 1	0.091	0.013	0.383	0.121	0.111	0.014	0.310	0.124	0.091	0.014	0.216	0.139
Category 2	0.602	0.018	3.108	0.164	0.367	0.017	2.911	0.159	0.645	0.016	2.647	0.168
Category 3	0.436	0.020	1.818	0.180	0.535	0.019	1.999	0.173	0.597	0.018	1.706	0.188
Category 4	0.192	0.014	1.644	0.129	0.187	0.014	1.558	0.128	0.212	0.014	1.324	0.137
Category 5	-0.039	0.013	0.826	0.117	-0.017	0.013	0.661	0.117	-0.016	0.013	0.686	0.130
LR Index (Pseudo-R^2)	0.005		0.051		0.007		0.052		0.008		0.045	

Table 6. Estimations by Year

Variable	1988				1989			
	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE
Constant	1.792	0.010	5.647	0.124	1.684	0.008	4.967	0.144
Category 1	0.06	0.014	0.496	0.181	0.097	0.013	0.697	0.211
Category 2	0.639	0.017	2.656	0.218	0.704	0.014	2.436	0.247
Category 3	0.509	0.019	1.812	0.244	0.548	0.016	1.532	0.273
Category 4	0.190	0.014	1.415	0.184	0.249	0.013	1.198	0.215
Category 5	-0.045	0.014	0.815	0.171	-0.023	0.012	0.869	0.202
LR Index (Pseudo-R^2)	0.008		0.045		0.009		0.038	

Table 6. Estimations by Year

Variable	1990				1991				1992			
	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE
Constant	1.606	0.009	4.488	0.091	1.541	0.009	3.861	0.104	1.449	0.009	3.375	0.067
Category 1	0.081	0.013	0.167	0.133	0.040	0.013	0.428	0.150	0.045	0.014	0.092	0.094
Category 2	0.772	0.015	2.273	0.160	0.752	0.015	2.176	0.179	0.826	0.016	2.083	0.112
Category 3	0.584	0.016	1.640	0.172	0.590	0.016	1.908	0.191	0.591	0.017	1.355	0.119
Category 4	0.299	0.014	1.277	0.137	0.298	0.013	1.120	0.153	0.308	0.014	0.947	0.097
Category 5	0.020	0.013	0.533	0.128	-0.002	0.012	0.361	0.144	0.012	0.014	0.365	0.092
LR Index (Pseudo-R^2)	0.011		0.048		0.011		0.050		0.013		0.050	

Table 6. Estimations by Year

Variable	1993				1994			
	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE
Constant	1.312	0.010	2.731	0.043	1.128	0.009	2.165	0.0755
Category 1	0.019	0.014	0.089	0.059	0.038	0.014	0.289	0.106
Category 2	0.875	0.015	1.967	0.068	0.923	0.015	1.813	0.115
Category 3	0.547	0.017	1.165	0.073	0.556	0.017	0.838	0.129
Category 4	0.355	0.014	0.913	0.062	0.358	0.014	0.763	0.107
Category 5	0.013	0.014	0.298	0.058	0.020	0.014	0.216	0.105
LR Index (Pseudo-R^2)	0.016		0.052		0.018		0.047	

Table 6. Estimations by Year

Variable	1995				1996				1997			
	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE
Constant	0.882	0.010	1.553	0.023	0.489	0.010	0.858	0.191	-0.092	0.012	0.318	0.098
Category 1	-0.012	0.014	0.021	0.033	0.004	0.015	0.997	0.263	-0.056	0.017	-0.263	0.146
Category 2	0.902	0.014	1.575	0.034	0.988	0.014	1.408	0.262	0.977	0.016	0.974	0.132
Category 3	0.454	0.016	0.592	0.038	0.371	0.017	0.381	0.307	0.164	0.018	-0.068	0.154
Category 4	0.382	0.014	0.725	0.032	0.413	0.014	0.618	0.263	0.429	0.016	0.407	0.135
Category 5	0.037	0.014	0.173	0.033	0.083	0.015	0.255	0.264	0.092	0.016	0.019	0.138
LR Index (Pseudo-R^2)	0.019		0.047		0.023		0.037		0.026		0.033	

Table 6. Estimations by Year

Variable	1998				1999			
	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE
Constant	-1.064	0.014	-0.778	0.326	-3.567	0.040	-3.567	0.041
Category 1	-0.068	0.021	-0.200	0.495	-0.033	0.059	-0.033	0.060
Category 2	0.912	0.017	1.376	0.390	0.767	0.048	0.768	0.048
Category 3	-0.049	0.022	-0.184	0.522	-0.238	0.068	-0.234	0.068
Category 4	0.398	0.019	0.192	0.445	0.545	0.051	0.558	0.051
Category 5	0.139	0.019	-0.034	0.463	0.141	0.056	0.144	0.056
LR Index (Pseudo-R^2)	0.026		0.076		0.014		0.014	

Table 7. Estimations by Category

Variable	Category 1				Category 2				Category 3			
	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE
Constant	1.827	0.009	15.415	0.227	2.121	0.019	16.117	0.347	2.275	0.024	13.064	0.145
Year 1976	0.039	0.013	0.114	0.321	0.048	0.027	-1.296	0.492	-0.077	0.034	-0.411	0.203
Year 1977	0.062	0.014	-1.150	0.333	0.129	0.027	-1.846	0.492	-0.126	0.034	-1.098	0.200
Year 1978	0.085	0.014	-2.504	0.330	0.192	0.027	-2.406	0.492	-0.182	0.035	-1.405	0.204
Year 1979	0.090	0.015	-3.269	0.363	0.134	0.029	-2.945	0.529	-0.135	0.036	-1.722	0.214
Year 1980	0.051	0.014	-4.009	0.341	0.186	0.027	-3.346	0.492	-0.115	0.033	-2.444	0.196
Year 1981	0.086	0.014	-4.914	0.330	0.214	0.027	-3.864	0.495	-0.115	0.033	-2.510	0.194
Year 1982	0.084	0.015	-5.379	0.350	0.238	0.027	-4.320	0.485	-0.035	0.034	-2.963	0.198
Year 1983	0.087	0.015	-5.975	0.349	0.262	0.027	-4.679	0.488	-0.032	0.034	-3.713	0.202
Year 1984	0.086	0.014	-7.064	0.339	0.360	0.026	-5.087	0.473	-0.031	0.033	-3.640	0.193
Year 1985	0.085	0.014	-7.718	0.336	0.302	0.025	-5.692	0.456	-0.017	0.032	-3.933	0.190
Year 1986	0.085	0.015	-8.322	0.347	0.318	0.025	-6.423	0.449	0.062	0.031	-4.282	0.186
Year 1987	0.067	0.014	-8.794	0.338	0.328	0.024	-7.065	0.427	0.126	0.030	-4.954	0.180
Year 1988	0.025	0.014	-9.270	0.340	0.310	0.024	-7.813	0.429	0.026	0.030	-5.603	0.181
Year 1989	-0.046	0.014	-9.749	0.323	0.267	0.023	-8.713	0.410	-0.042	0.029	-6.564	0.172
Year 1990	-0.139	0.014	-10.759	0.325	0.258	0.023	-9.355	0.417	-0.084	0.029	-6.936	0.173
Year 1991	-0.245	0.014	-11.125	0.321	0.172	0.023	-10.079	0.413	-0.143	0.029	-7.295	0.171
Year 1992	-0.333	0.013	-11.946	0.319	0.154	0.023	-10.657	0.411	-0.235	0.029	-8.333	0.170
Year 1993	-0.495	0.013	-12.593	0.318	0.066	0.022	-11.417	0.404	-0.415	0.029	-9.167	0.169
Year 1994	-0.660	0.014	-12.960	0.323	-0.069	0.022	-12.138	0.398	-0.591	0.028	-10.060	0.169
Year 1995	-0.957	0.014	-13.840	0.336	-0.337	0.022	-12.989	0.395	-0.938	0.028	-10.919	0.169
Year 1996	-1.333	0.015	-13.559	0.332	-0.643	0.021	-13.850	0.389	-1.414	0.028	-11.824	0.169
Year 1997	-1.976	0.016	-15.359	0.386	-1.236	0.022	-14.825	0.394	-2.203	0.029	-12.814	0.172
Year 1998	-2.961	0.019	-16.393	0.453	-2.273	0.022	-15.519	0.384	-3.390	0.031	-14.027	0.185
Year 1999	-5.428	0.050	-19.015	1.251	-4.922	0.035	-18.916	0.644	-6.081	0.071	-16.865	0.434
LR Index (Pseudo-R^2)	0.084		0.473		0.136		0.538		0.154		0.488	

Table 7. Estimations by Category

Variable	Category 4				Category 5				Category 6			
	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE	Q_0	SE	Q_T	SE
Constant	1.907	0.011	14.076	0.074	1.793	0.008	14.878	0.122	1.841	0.009	12.363	0.179
Year 1976	0.034	0.016	-0.178	0.104	0.011	0.012	-1.759	0.177	0.041	0.013	-0.434	0.257
Year 1977	0.082	0.016	-0.873	0.105	0.011	0.012	-1.531	0.180	0.071	0.013	0.253	0.262
Year 1978	0.083	0.016	-1.313	0.106	0.007	0.012	-2.409	0.180	0.036	0.013	-1.584	0.258
Year 1979	0.102	0.018	-1.588	0.115	0.019	0.014	-3.142	0.197	0.012	0.014	-2.078	0.281
Year 1980	0.079	0.017	-2.565	0.108	-0.023	0.013	-3.811	0.185	-0.005	0.013	-2.425	0.262
Year 1981	0.093	0.016	-2.955	0.107	-0.005	0.013	-4.494	0.182	-0.007	0.013	-3.109	0.259
Year 1982	0.081	0.016	-3.425	0.107	0.008	0.013	-4.873	0.190	-0.003	0.013	-3.388	0.271
Year 1983	0.101	0.017	-3.993	0.108	0.031	0.013	-5.576	0.190	-0.015	0.013	-4.076	0.273
Year 1984	0.088	0.016	-4.552	0.103	-0.011	0.013	-6.013	0.182	-0.024	0.013	-4.357	0.258
Year 1985	0.106	0.015	-5.118	0.101	-0.011	0.012	-6.738	0.177	-0.019	0.013	-5.050	0.255
Year 1986	0.081	0.016	-5.734	0.102	-0.009	0.012	-7.433	0.177	-0.039	0.013	-5.580	0.254
Year 1987	0.109	0.015	-6.348	0.096	-0.006	0.012	-7.787	0.172	-0.037	0.012	-5.959	0.248
Year 1988	0.075	0.015	-7.013	0.099	-0.046	0.012	-8.415	0.174	-0.048	0.012	-6.715	0.254
Year 1989	0.025	0.014	-7.909	0.095	-0.133	0.012	-9.041	0.168	-0.157	0.012	-7.395	0.241
Year 1990	-0.002	0.015	-8.310	0.096	-0.167	0.012	-9.856	0.171	-0.235	0.012	-7.875	0.244
Year 1991	-0.067	0.014	-9.094	0.095	-0.254	0.012	-10.655	0.168	-0.300	0.012	-8.501	0.243
Year 1992	-0.149	0.014	-9.752	0.095	-0.332	0.012	-11.136	0.169	-0.392	0.012	-8.987	0.245
Year 1993	-0.238	0.015	-10.430	0.095	-0.468	0.012	-11.848	0.170	-0.528	0.012	-9.631	0.250
Year 1994	-0.421	0.014	-11.147	0.094	-0.645	0.012	-12.496	0.174	-0.713	0.012	-10.197	0.249
Year 1995	-0.642	0.014	-11.797	0.094	-0.873	0.012	-13.151	0.186	-0.959	0.013	-10.810	0.256
Year 1996	-1.004	0.015	-12.599	0.096	-1.221	0.012	-13.764	0.182	-1.352	0.013	-11.504	0.264
Year 1997	-1.570	0.015	-13.350	0.099	-1.794	0.014	-15.540	0.197	-1.933	0.014	-12.044	0.280
Year 1998	-2.574	0.016	-14.662	0.108	-2.719	0.015	-15.691	0.226	-2.906	0.016	-13.141	0.314
Year 1999	-4.929	0.033	-17.084	0.226	-5.220	0.039	-18.301	0.596	-5.408	0.040	-15.930	0.869
LR Index (Pseudo-R^2)	0.093		0.452		0.078		0.449		0.087		0.428	

Table 8. General Estimations for Q_T

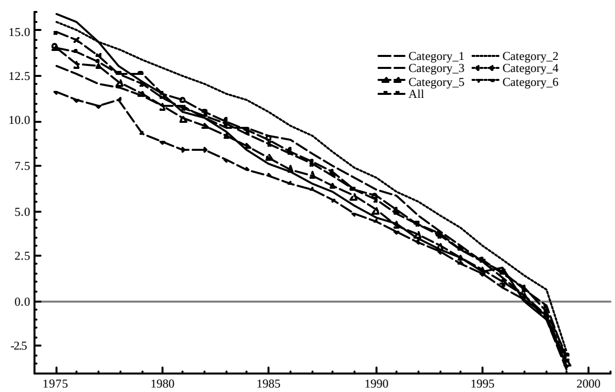
Variable	Q_T	SE
Constant	13.845	0.0042
Category 1	0.953	0.0025
Category 2	2.342	0.0028
Category 3	1.311	0.0031
Category 4	1.251	0.0024
Category 5	0.711	0.0023
Year 1976	-0.349	0.0053
Year 1977	-1.139	0.0054
Year 1978	-2.289	0.0053
Year 1979	-2.913	0.0058
Year 1980	-3.604	0.0055
Year 1981	-4.258	0.0054
Year 1982	-4.656	0.0056
Year 1983	-5.268	0.0056
Year 1984	-5.775	0.0054
Year 1985	-6.389	0.0053
Year 1986	-6.971	0.0053
Year 1987	-7.458	0.0051
Year 1988	-8.125	0.0052
Year 1989	-8.859	0.0050
Year 1990	-9.475	0.0051
Year 1991	-10.103	0.0050
Year 1992	-10.755	0.0050
Year 1993	-11.462	0.0050
Year 1994	-12.094	0.0050
Year 1995	-12.824	0.0051
Year 1996	-13.398	0.0051
Year 1997	-14.410	0.0055
Year 1998	-15.538	0.0058
Year 1999	-18.490	0.0140
LR Index (Pseudo-R^2)	0.4801	

Table 9 - Count of non nil Qis by year

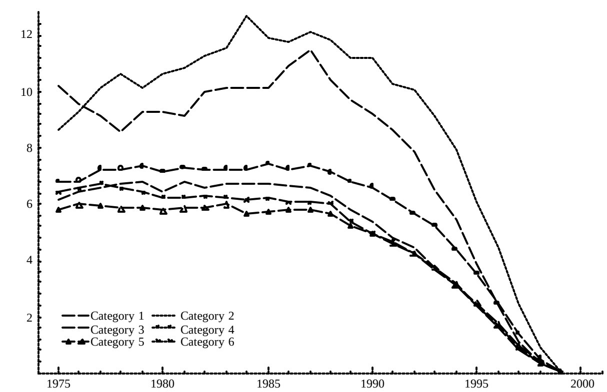
		Order													
	# of patents	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1975	72000	65126	62111	59527	57047	54186	51023	47383	43269	38654	33784	28637	23605	18932	14745
1976	70226	63699	60647	58100	55467	52655	49480	45742	41517	36829	31712	26458	21414	16866	12852
1977	65269	59452	56636	54198	51690	48809	45532	41759	37541	32820	27835	22819	18007	13818	10266
1978	66102	60101	57237	54636	51925	48848	45178	40890	36236	31077	25541	20157	15396	11427	8111
1979	48854	44365	42176	40165	38012	35570	32630	29305	25471	21457	17287	13440	10107	7280	5065
1980	61819	55913	52929	50131	47129	43635	39580	34905	29601	24104	18733	13912	10014	6856	4602
1981	65771	59526	56317	53250	49887	45925	41169	35878	30155	24127	18287	13147	9074	5907	3778
1982	57888	52572	49597	46771	43576	39808	35353	30408	25089	19735	14550	10255	6906	4436	2714
1983	56860	51517	48486	45422	42081	38066	33338	28146	22481	17074	12212	8370	5389	3305	1911
1984	67200	60884	57100	53382	48966	43649	37448	30572	23669	17106	11552	7421	4519	2624	1390
1985	71661	65100	60780	56335	51037	44822	37440	29507	21788	14928	9463	5835	3231	1632	803
1986	70860	64298	59861	55040	49161	41969	33875	25451	17698	11195	6583	3551	1767	858	380
1987	82952	75530	69885	63755	55897	46654	36087	25824	16822	9933	5417	2565	1083	452	189
1988	77924	70482	64582	57813	49068	38709	27838	18156	10680	5547	2606	1057	432	153	65
1989	95537	85531	77084	66749	53847	39217	25446	14760	7357	3070	1125	397	161	68	32
1990	90364	79854	70688	59056	44702	29791	17336	8533	3426	1217	397	126	52	24	13
1991	96513	84378	72394	57024	39154	22768	11319	4473	1394	399	93	39	29	11	6
1992	97444	83383	68752	50078	30409	14930	5996	1784	497	118	43	19	11	9	5
1993	98342	81841	62977	40401	20413	8181	2346	541	96	22	10	4	1	1	0
1994	101676	82032	57389	30966	11991	3315	656	119	20	14	7	2	2	1	0
1995	101419	76925	45815	18430	4755	849	134	35	17	8	1	1	1	0	0
1996	109645	74428	33218	7917	1080	115	36	16	10	5	2	2	2	1	0
1997	111983	58822	15158	1534	115	30	19	8	7	3	2	1	0	0	0
1998	147519	43157	3035	121	32	24	12	7	4	2	2	1	1	1	1
1999	153486	5088	25	0	0	0	0	0	0	0	0	0	0	0	0
All	2139314	1604004	1304879	1080801	897441	742525	609271	494202	394845	309444	237244	178216	131204	94662	66928

Table 9 - Count of non nil Qis by year (cont.)

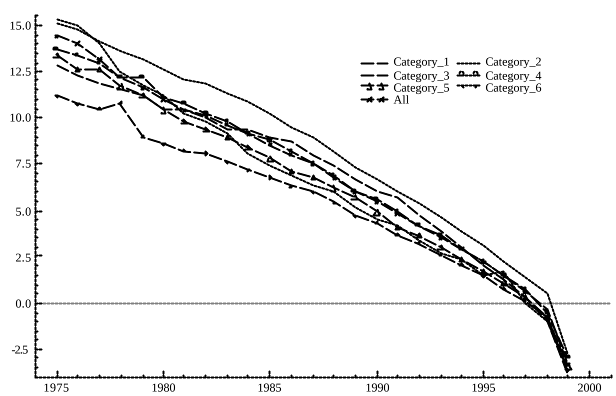
							Order								
14	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**
11122	8172	5863	4113	2768	1821	1090	557	212	99	52	27	17	16	13	3
9528	6829	4700	3178	2049	1236	686	312	141	83	43	25	23	17	7	0
7438	5250	3576	2338	1458	764	361	142	74	37	27	20	14	9	0	0
5669	3819	2501	1514	826	391	135	75	34	18	13	10	4	0	0	0
3433	2240	1413	845	437	177	72	38	15	9	7	2	0	0	0	0
2881	1747	992	544	230	87	51	24	14	8	3	0	0	0	0	0
2315	1320	743	355	128	65	40	13	7	2	0	0	0	0	0	0
1598	896	466	205	90	46	17	6	2	0	0	0	0	0	0	0
1088	557	222	82	42	24	14	3	0	0	0	0	0	0	0	0
673	284	113	35	20	12	2	0	0	0	0	0	0	0	0	0
352	162	56	23	10	1	0	0	0	0	0	0	0	0	0	0
169	69	20	5	2	0	0	0	0	0	0	0	0	0	0	0
92	18	6	2	0	0	0	0	0	0	0	0	0	0	0	0
22	12	4	1	0	0	0	0	0	0	0	0	0	0	0	0
8	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
8	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46401	31382	20678	13241	8060	4624	2468	1170	499	256	145	84	58	42	20	3



(a) Average of Log(Q_T) by Category



(b) Average Q_0 by Category



(c) Average of Log(Q_w) by Category

Figure 1. Average of Log(Q_T), Q_0 and Log(Q_w) by Category

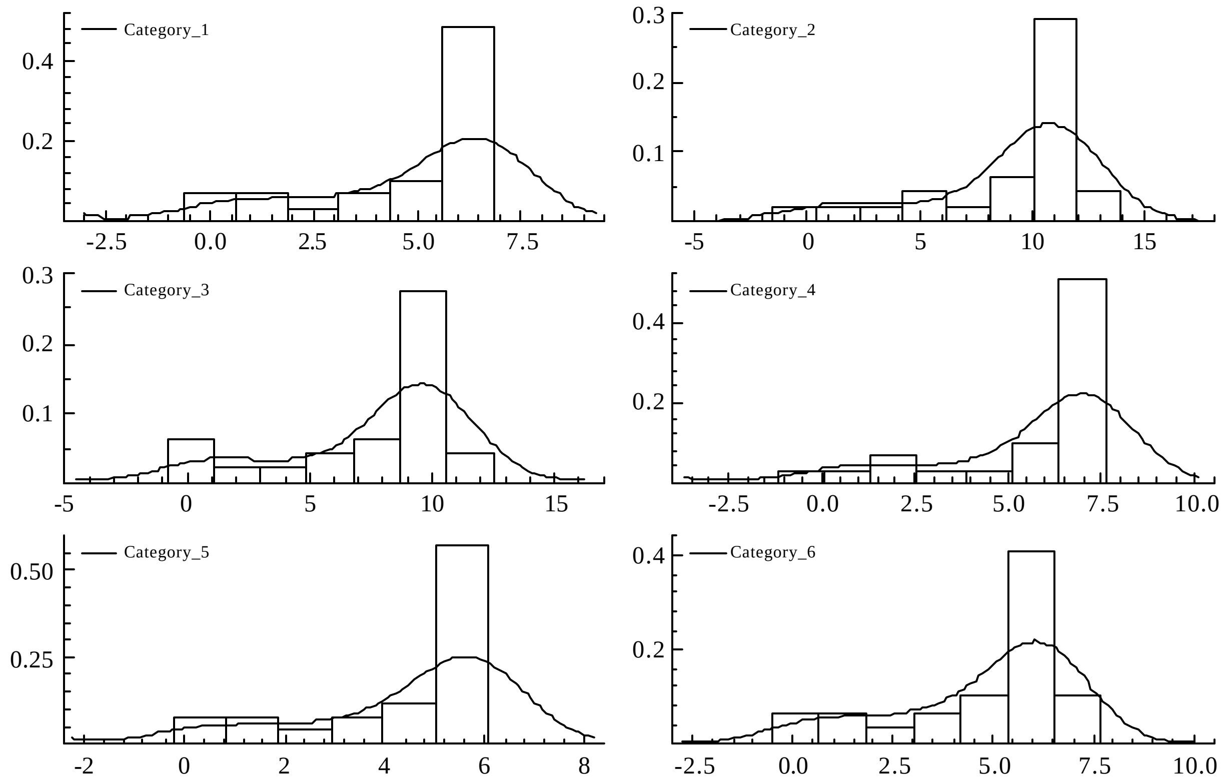


Figure 2. Histograms and Densities of the Average of Q_0 by Category

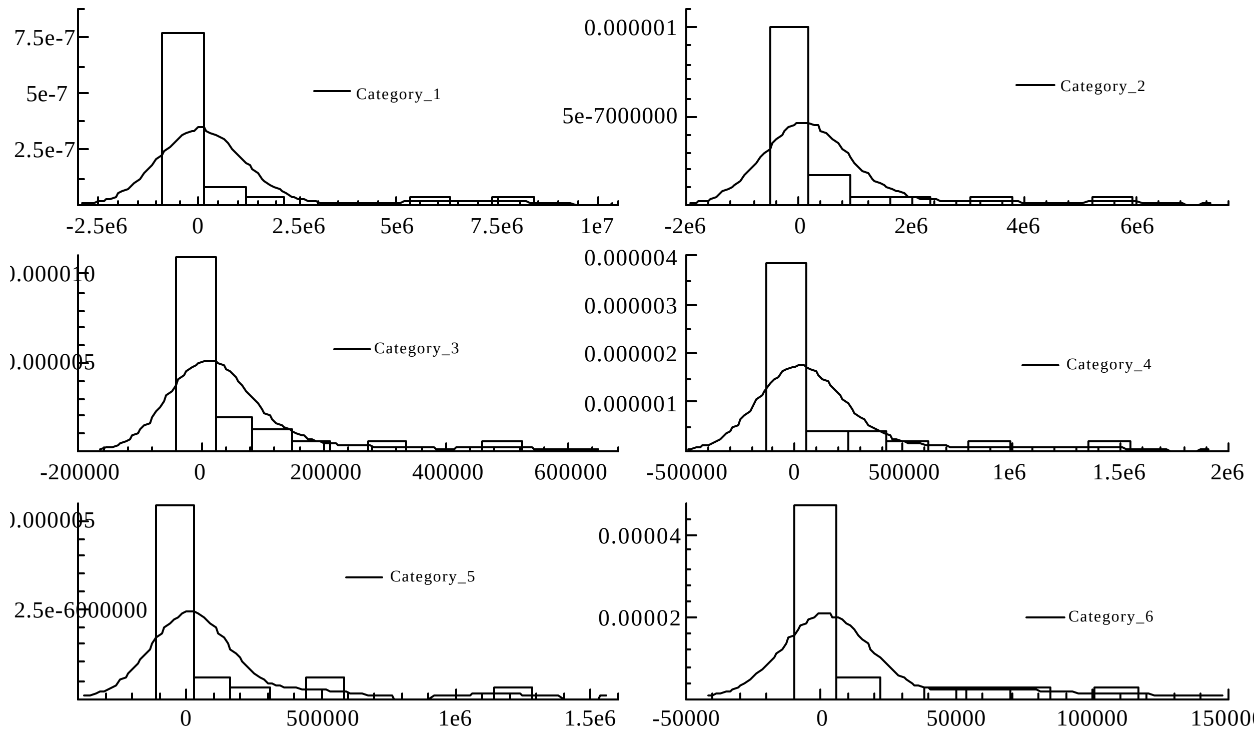


Figure 3. Histograms and Densities of the Average of Q_T by Category

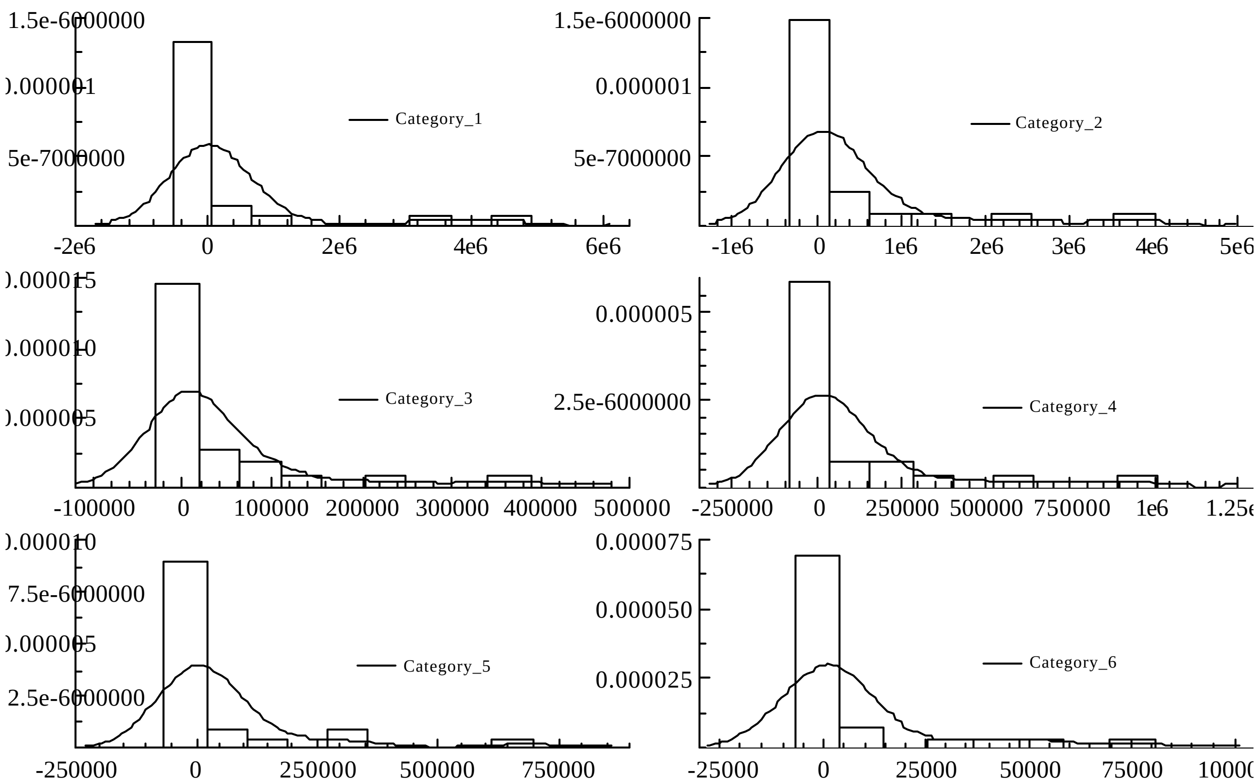


Figure 4. Histograms and Densities of the Average of Q_W by Category