

NBER WORKING PAPER SERIES

FLOWS OF KNOWLEDGE FROM
UNIVERSITIES AND FEDERAL LABS:
MODELING THE FLOW OF PATENT
CITATIONS OVER TIME AND ACROSS
INSTITUTIONAL AND GEOGRAPHIC
BOUNDARIES

Adam B. Jaffe
Manuel Trajtenberg

Working Paper 5712

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1996

This paper was prepared for the National Academy of Sciences Colloquium on Science, Technology and the Economy, Irvine, California, October 20, 1995. This research was supported by the National Science Foundation under grants SBR-9320973 and SBR-9413099. This paper is part of NBER's research program in Productivity. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

© 1996 by Adam B. Jaffe and Manuel Trajtenberg. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

FLOWS OF KNOWLEDGE FROM
UNIVERSITIES AND FEDERAL LABS:
MODELING THE FLOW OF PATENT
CITATIONS OVER TIME AND ACROSS
INSTITUTIONAL AND GEOGRAPHIC
BOUNDARIES

ABSTRACT

The extent to which new technological knowledge flows across institutional and national boundaries is a question of great importance for public policy and the modeling of economic growth. This paper develops a model of the process generating subsequent citations to patents as a lens for viewing knowledge diffusion. We find that the probability of patent citation over time after a patent is granted fits well to a double-exponential function that can be interpreted as the mixture of diffusion and obsolescence functions. The results indicate that diffusion is geographically localized. Controlling for other factors, within-country citations are more numerous and come more quickly than those that cross country boundaries.

Adam B. Jaffe
Department of Economics
Brandeis University
Waltham, MA 02254-9110
and NBER
jaffe@binah.cc.brandeis.edu

Manuel Trajtenberg
Department of Economics
Tel Aviv University
Tel Aviv 69978
ISRAEL
and NBER
manuel@ccsg.tau.ac.il

I. INTRODUCTION

The rate at which knowledge diffuses outward from the institutional setting and geographic location in which it is created has important implications for the modelling of technological change and economic growth, and for science and technology policy. Models of endogenous economic growth, such as Romer (1990) or Grossman and Helpman (1991), typically treat knowledge as completely diffused within an economy, but implicitly or explicitly assume that knowledge does not diffuse across economies. In the policy arena, ultimate economic benefits are increasingly seen as the primary policy motivation for public support of scientific research. Obviously, the economic benefits to the U.S. economy of domestic research depend on the fruits of that research being more easily or more quickly harvested by domestic firms than by foreign firms. Thus for both modeling and policy-making purposes it is crucial to understand the institutional, geographic and temporal dimensions of the spread of newly created knowledge.

In a previous paper with Rebecca Henderson (Jaffe, Henderson and Trajtenberg, 1993) we explored the extent to which citations by patents to previous patents are geographically localized, relative to a baseline likelihood of localization based on the predetermined pattern of technological activity. This paper extends that work in several important dimensions. First, we use a much larger number of patents over a much longer period of time. This allows us to explicitly introduce time, and hence diffusion, into the citation process. Second, we enrich the institutional comparisons we can make by looking at 3 distinct sources of potentially cited patents: U.S. corporations, U.S. universities and the U.S. government. Third, the larger number of patents allows us to enrich the geographic portrait by examining separately the diffusion of knowledge from U.S. institutions to inventors in Canada, Europe, Japan and the rest of the world. Finally, our earlier work took the act of citation as exogenous, and simply measured how often that citation came from nearby. In this paper we develop a modelling framework that allows the generation of citations from multiple distinct locations to be generated by a random process whose parameters we estimate.

II. THE DATA

We are in the process of collecting from commercial sources a complete database on all

U.S. patents¹ granted since 1963 (approximately 2.5 million patents), including data for each indicating the nature of the organization if any, to which the patent property right was assigned;² the names of the inventors and the organization, if any to which the patent right was assigned; the residence of each inventor;³ the date of the patent application and the patent grant;⁴ a detailed technological classification for the patent; and miscellaneous other information. The data on individual patents are complemented by a file indicating all of the citations made by U.S. patents since 1977 to previous U.S. patents (approximately 9 million citations). Using the citation information in conjunction with the detailed information about each patent itself, we have an extremely rich mine of information about individual inventive acts *and* the links among them as indicated by citations made by a given patent to a previous one.

We have discussed elsewhere at great length the advantages and disadvantages of using patents and patent citations to indicate inventions and knowledge links among inventions (Jaffe, Henderson and Trajtenberg, 1993; Trajtenberg, Henderson and Jaffe, 1992; See also Griliches, 1990). Patent citations perform the legal function of delimiting the patent right by identifying previous patents whose technological scope is explicitly placed outside the bounds of the citing patent. Hence the appearance of a citation indicates that the cited patent is, in some sense, a technological antecedent of the citing patent. Patent applicants bear a legal obligation to disclose any knowledge that they might have of relevant prior inventions, and the patent examiner may also add citations not identified by the applicant.

Our basic goal in this paper is to explore the process by which citations to a given patent arrive over time, how this process is affected by characteristics of the cited patent, and how different potentially citing locations differ in the speed and extent to which they "pick up"

1. By "U.S. patents," we mean in this context patents granted by the U.S. patent office. All of our research relies on U.S. patents in this sense. Currently about half of U.S. patents are granted to foreigners. Hence, later in the paper, we will use the phrase "U.S. patents" to mean patents granted to residents of the U.S., as opposed to those granted to foreigners.

2. Patents are reported as unassigned, or else assigned to one of: U.S. corporation, non-U.S. corporation, U.S. individual, non-U.S. individual, U.S. government, non-U.S. government.

3. city and state for U.S. inventors, country for non-U.S. inventors

4. We have data only on successful patent applications. Unsuccessful applications are not public information.

existing knowledge, as evidenced by their acknowledgement of such existing knowledge through citation. Because of the policy context mentioned above, we are particularly interested in citations to university and government patents. We recognize that much of the research that goes on at both universities and government laboratories never results in patents, and presumably has impacts that cannot be traced via our patent-citations-based research. We believe, however, that at least with respect to relatively near-term economic impacts, patents and their citations are at least a useful window into the otherwise "black box" of the spread of scientific and technical knowledge.

The analysis in this paper is based on the citations made to 3 distinct sets of "potentially cited" patents. The first set is a 1-in-10 random sample of all patents granted between 1963 and 1990 and assigned to U.S. corporations (88,257 patents). The second set is the *universe* of all patents granted between 1965 and 1990 to U.S. universities, based on a set of assignees identified by the Patent Office as being universities or related entities such as teaching hospitals (10,761 patents).⁵ The third set is the universe of patents granted between 1963 and 1990 to the U.S. government (38,254 patents). Based on comparisons with numbers published by NSF, these are overwhelmingly coming from federal labs, and the bulk come from the large federal labs. The U.S. government set also includes, however, small numbers of patents from diverse parts of the federal government. We have identified all patents granted between 1977 and 1993 that cite any of the patents in these three sets (479,861 citing patents). Thus we are using temporal, institutional, geographic and technological information on over 600,000 patents over about 30 years.

Some simple statistics from these data are presented in Table One. On average, university patents are more highly cited, despite the fact that more of them are recent.⁶ Federal patents are *less* highly cited than corporate patents. But it is difficult to know how to interpret these averages, because many different effects all contribute to these means. First, the differences in timing are important, because we know from other work that the overall rate of citation has been

5. There are, presumably university patents before 1965, but we do not have the ability to identify them as such.

6. In previous work (Henderson, Jaffe and Trajtenberg, 1995), we showed that university patents applied for up until about 1982 were more highly cited than corporate patents, but that the difference has since disappeared.

rising over time (Caballero and Jaffe, 1993), so more recent patents will tend to be more highly cited than older ones. Second, there are significant differences in the composition of the different groups by technical field. Most dramatically, university patents are much more highly concentrated in Drugs and Medical Technology, and less concentrated in Mechanical Technology, than the other groups. Conversely, the federal patents are much more concentrated in Electronics, Optics and Nuclear Technology than either of the other groups, with less focus on Chemicals. To the extent that citation practices vary across fields, differences in citation intensities by type of institution could be due to field effects. Finally, as shown in Table Two, different potentially citing locations have different field focuses of their own, with Japan more likely to cite Electronics patents and less likely to cite Drug and Medical patents. The main contribution of this paper is the exploration of an empirical framework in which all of these different effects can be sorted out, at least in principle.

III. THE MODEL

We seek a flexible descriptive model of the random processes underlying the generation of citations, which will allow us to estimate parameters of the diffusion process while controlling for variations over time and technological fields in the "propensity to cite." For this purpose we adapt the formulation of Caballero and Jaffe (1993), in which the likelihood that any particular patent K granted in year T will cite some particular patent k granted in year t is assumed to be determined by the combination of an exponential process by which knowledge diffuses and a second exponential process by which knowledge becomes obsolete. That is:

$$p(k,K) = \alpha(k,K) \exp[-\beta_1(k,K)(T-t)][1 - \exp(-\beta_2(T-t))] \quad (1)$$

where β_1 determines the rate of obsolescence and β_2 determines the rate of diffusion. We refer to the likelihood determined by Equation (1) as the "citation frequency," and the citation frequency as a function of the citation lag $(T-t)$ as a citation function. The dependence of the parameters α and β_1 on k and K is meant to indicate that these could be functions of certain attributes of both the cited and citing patents. In this paper, we consider the following as attributes of the cited patent k that might affect its citation frequency: t , the grant year of the potentially cited patent; $i=1..3$, the institutional nature of the assignee of the potentially cited

patent (corporate, university or government); and $g=1..5$, the technological field of the potentially cited patent. As attributes of the potentially citing patent K that might affect the citation likelihood we consider: T , the grant year of the potentially citing patent, and $L=1..5$, the location of the potentially citing patent.

To illustrate the plausibility of this formulation, we plot the average citation functions (citation frequency as a function of time elapsed from the potentially cited patent), for each of the 5 geographic regions (Figure One) and each of the 3 institutional groups (Figure Two).⁷ These figures show that citations show a pattern of gradual diffusion and ultimate obsolescence, with maximal citation frequency occurring after about 5 years. The contrasts across groups and countries in these raw averages are striking: U.S. patents are much more likely to cite our 3 groups of U.S. patents than are any other locations, with an apparent ranking among other regions of Canada, "Rest of World," E.E.C., and then Japan. Figure Two shows that universities are much more highly cited than corporations, at least up until lags of about 20 years, and federal labs are less highly cited throughout. While many of these contrasts will survive more careful scrutiny, it is important at this point to note that these comparisons do not control for time or technical field effects. In particular, the country comparisons are distorted by differences in technical field focus, and the university/corporate comparison is distorted by differences in timing. The relatively high frequencies associated with the longest lags are also distorted by the fact that these long lags can only be observed for the most recent citing years, and the intensity of citations made is higher in these years than earlier in the data.

Additional insight into this parameterization of the diffusion process can be gained by determining the lag at which the citation function is maximized ("the modal lag"), and the maximum value of the citation frequency achieved. A little calculus shows that the modal lag is approximately equal to $1/\beta_1$; increases in β_1 shift the citation function to the left. The maximum value of the citation frequency is approximately determined by β_2/β_1 ; increases in β_2 holding β_1 constant increase the overall citation intensity.⁸ Indeed, increases in β_2 , holding

7. For convenience, we will refer to the citation frequency as a function of citation lag as a "citation function."

8. The approximation involved is that $\log(1 + \beta_2/\beta_1) \approx \beta_2/\beta_1$. Our estimations all lead to β_2/β_1 on the order of 10^{-6} , and indeed the approximation holds to 5 significant figures for lags up to 30 years.

β_1 constant are very close to equivalent to increasing the citation frequency *proportionately* at every value of (T-t). That is, variations in β_2 , holding β_1 constant are not separately identified from variations in α . Hence because the model is somewhat easier to estimate and interpret with variations in α , we do not allow variations in β_2 .

Consider now a potentially cited patent with particular i,t,g attributes, e.g. a university patent in the Drug and Medical area granted in 1985. The expected number of citations that this patent will receive from a particular T,L combination (e.g. Japanese patents granted in 1993) is just the above likelihood, as a function of i,t,g,T and L, times the *number* of patents in the particular T,L group that are thereby potential citing patents. Even aggregating in this way over T and L, this is still a very small expected value, and so it is not efficient to carry out estimation at the level of the individual potentially cited patent. Instead we aggregate across all patents in a particular i,t,g cell, counting all of the citations *received* by, e.g., university drug patents granted in 1985, *given by*, e.g., Japanese patents in 1993. The expected value of this total is just the expected value for any one potentially cited patent, times the number of potentially cited patents in the i,t,g cell. In symbols:

$$E[c_{igTL}] = (n_{TL})(n_{ig})\alpha_{igTL}\exp[-(\beta_1)_{igTL}(k,K)(T-t)][1 - \exp(-\beta_2(T-t))] \quad (2)$$

or

$$\frac{E[c_{igTL}]}{(n_{TL})(n_{ig})} = \alpha_{igTL}\exp[-(\beta_1)_{igTL}(k,K)(T-t)][1 - \exp(-\beta_2(T-t))] \quad (3)$$

implying that the equation

$$p_{igTL} \equiv \frac{c_{igTL}}{(n_{TL})(n_{ig})} = \alpha_{igTL}\exp[-(\beta_1)_{igTL}(k,K)(T-t)][1 - \exp(-\beta_2(T-t))] + \epsilon_{igTL} \quad (4)$$

can be estimated by non-linear least squares if the error ϵ_{igTL} is well-behaved. The data set consists of one observation for each feasible combination of values of i,t,g,T and L. The corporate and federal data each contribute 9275 observations (5 values of g times 5 values of L times 28 values of t times either 17 (for years before 1977) or 1993-t (for years beginning in

1977) values of T .⁹ Because the university patents start only in 1965, there are only 8425 university cells, for a total number of observations of 26,975. Of these, about 25% have zero citations;¹⁰ the mean number of citations is about 18 and the maximum is 737. The mean value of p_{igtL} is 3.3×10^{-6} .

IV. MODEL SPECIFICATION AND INTERPRETATION

The first specification issue to consider is the difficulty of estimating effects associated with cited year, citing year and lag. This is analogous to estimating “vintage,” time, and age effects in a wage model or a hedonic price model. If lag (our “age” effect) entered the model linearly, then it would be impossible to estimate all three effects. Given that lag enters our model non-linearly, all three effects are identified, in principle. In practice, we found that we could not get the model to converge with the double-exponential lag function and separate α parameters for each cited year and each citing year. We were, however, able to estimate a model in which cited years are grouped into 5 year-intervals. Hence we assume that $\alpha(t)$ is constant over t for these intervals, but allow the intervals to differ from each other.

All of the estimation is carried out including a “base” value for β_1 and β_2 , with all other effects estimated relative to a base value of unity.¹¹ The various different effects are included by entering multiplicative parameters, so that the estimating equation looks like:

$$p_{igtL} = \alpha_i \alpha_p \alpha_g \alpha_T \alpha_L \exp[-(\beta_1) \beta_{it} \beta_{lg} \beta_{lL} (T-t)] [1 - \exp(-\beta_2 (T-t))] + \epsilon_{igtL} \quad (5)$$

where $i=c,u,f$ (cited institution type); $t=1963-1990$ (cited year) $tp=1\dots 6$ (five year intervals for cited year, except first interval is 1963-65); $g=1\dots 5$ (technological field of cited patent); $T=1977\dots 1993$ (citing year); and $L=1\dots 5$ (citing region). In this model, unlike the linear case,

9. We exclude cells for which $t=T$, where the model predicts that the number of citations is identically zero. In fact, the number of citations in such cells is almost always zero.

10. About two-thirds of the zero citation observations are for cells associated with either Canada or Rest of World.

11. As noted above, α is not separately identified from β_1 and β_2 . Hence we do not estimate a “base” value for the parameter α ; it is implicitly unity.

the null hypothesis of no effect corresponds to parameter values of unity rather than zero. For each effect, one group is omitted from estimation, i.e. its multiplicative parameter is constrained to unity. Thus the parameter values are interpreted as relative to that base group. The base group for each effect is as follows:

Cited time period (tp)	1963-65
Cited field (g)	"All Other"
Type of Cited Institution (i)	Corporate
Citing year (T)	1977
Citing region (L)	U.S.

The estimate of any particular $\alpha(k)$, say $\alpha(g=\text{Drugs and Medical})$, is a proportionality factor measuring the extent to which the patents in the field "Drugs and Medical" are more or less likely to be cited over time vis à vis patents in the base category "All Other." Thus, an estimate of $\alpha(k=\text{Drugs}) = 1.4$ means that the likelihood that a patent in the field of Drugs and Medical will receive a citation is 40% higher than the likelihood of a patent in the base category, controlling of course for a wide range of factors. Notice that this is true across all lags; we can think of an α greater than unity as meaning that the citation function is shifted upward proportionately, relative to the base group. Hence the integral over time (i.e. the total number of citations per patent) will also be 40% larger.

We can think of the overall rate of citation intensity measured by variations in α to be composed of two parts. Citation intensity is the product of the "fertility" (Caballero and Jaffe, 1993) of the underlying ideas in spawning future technological developments,¹² and the average "size" of a patent, i.e. how much of the unobservable advance of knowledge is packaged in a typical patent. Within the formulation of this paper, it is not possible to decompose the α -effects

12. The "fertility" label for differences in the likelihood of citations is just evocative language suggesting that cited patents can be thought as "giving birth" to further technological developments, or as providing the grounds on which such subsequent applications grow. "Importance" is a term that we have used in this context in previous work, and refers to the fact that patents that collect more citations over time are deemed on that basis to be more "important". Since differences in α are time-independent, the integral over time would also be higher or lower by the same factor, and hence the equivalence with this notion of "importance."

into these two components.¹³

In the case of $\alpha(K)$, that is, when the multiplicative factor varies with attributes of the *citing* patents, variations in it should be interpreted as differences in the "propensity to cite" (or in the probability of *making* a citation) of patents in a particular category vis à vis the base category of the citing patents. If, for example, $\alpha(K=Europe)$ is 0.5, this means that the average patent granted to European inventors is half as likely as a patent granted to inventors residing in the US to cite any given US patent.

Variations in β_1 (again, by attributes of either the cited or the citing patents) imply differences in the rate of decay or "obsolescence" across categories of patents. Higher values of β_1 mean higher rates of decay, which pull the citations function downwards and leftward. In other words, the likelihood of citations would be lower everywhere for higher β_1 's, and would peak earlier on. Thus a higher α means more citations at all lags; a lower β_1 means more citations at *later* lags.

When both $\alpha(k, K)$ and $\beta_1(k, K)$ vary, the citation function can shift upward at some lags while shifting downward at others. For example, if $\alpha(g=Electronics) = 2.00$, but $\beta_1(g=Electronics) = 1.29$, then patents in electronics have a very high likelihood of citations relative to the base category, but they also become obsolete faster. Because obsolescence is compounded over time, differences in β_1 eventually result in large differences in the citation frequency. If we compute the ratio of the likelihood of citations for patents in electronics relative to those in "all other" using these parameters, we find that 1 year after being granted patents in electronics are 89% more likely to be cited, but 12 years down the road the frequencies for the two groups are about the same, and at a lag of 20 years Electronics patents are actually 36% *less* likely to be cited than patents in the base category.

V. RESULTS

5.1 Benchmark results

Table 3 shows results for a "benchmark" specification where we allow α to vary but hold

13. Caballero and Jaffe (1993) attempt to identify the size of patents by allowing exponential obsolescence to be a function of accumulated patents rather than elapsed calendar time. We intend to explore this possibility in future work.

β_1 constant. We carry out all estimation weighting the observations by $nn=(n_{gi} * n_{TL})^{**0.5}$, the square root of the product of the number of potentially cited and potentially citing patents corresponding to the cell. This weighting scheme should take care of possible heteroskedasticity, since the observations correspond essentially to "grouped data", that is, each observation is an average (in the corresponding cell), computed by dividing the number of citations by $(n_{gi} * n_{ij})$.¹⁴

The parameters for the citing years (α_t), show a clear upward trend, reaching a peak in 1989, with a slight decline thereafter. The estimates imply that citations made per patent, controlling for other factors, was twice as high in 1989 as in 1977.¹⁵ This reflects a well known institutional phenomenon, namely, the increasing propensity to make citations at the patent office, apparently associated with the computerization of the patent file and of the operations of patent examiners. Thus, these coefficients serve strictly as controls, and as very important ones for that matter, given the magnitude of the changes over time. Without these controls, obsolescence would appear to be much slower (since our only observations on long lags come from recent citing years), and foreign countries would appear to make more citations (relative to U.S.), since the foreign share of patents has increased dramatically over the time period.

Turning to the effects of the *cited* time period, because we include interaction effects between the cited period and the source institution (discussed below), the "pure" cited time period estimates in Table 3 correspond to variations in α over cited year for the base group of corporate patents. The citations received by this group, as implied by the parameter estimates, declined steadily, relative to the normalized value of unity in the base period 1963-65, to 0.65 in 1981-85, somewhat recovering in 1986-90 to 0.73. As discussed above, this means either that ideas underlying corporate patents declined in fertility, that the average "package size" of corporate patents fell, or both.

The number of patents granted per dollar of R&D fell from the 1960s through the mid-1980s, recovering slightly since then. Because of this, it is typically assumed that patents have,

14. The qualitative results are essentially the same if we run an unweighted model, or if we weight just by n_{gi} . Weighting increases the fit of the regression dramatically, suggesting that the model does not fit very well for those cells with few potentially citing patents, which frequently produce zero citations.

15. This difference is even greater than the raw change in average citations per patent, which is an increase by about 60%.

in some sense, been getting "bigger" rather than smaller, at least until the mid-1980s. In and of itself, this would tend to cause our α_p to rise up until the mid-1980s, and perhaps fall after that. Since what we found was the exact opposite, it seems unlikely that our results are due to changes in the "size" of patents. Hence they suggest that "fertility" fell from the 1960s until the mid 1980s, with a mild recovery thereafter. While we would not push this very hard, this interpretation of our results would be consistent with a technology-driven slowdown in productivity in the 1970s and early 1980s.¹⁶

The variations in α by technological fields indicate that patents in Drugs and Medical and in Electronics are 36-40% more likely to be cited than in other fields. In addition to the fertility and patent-size effects discussed above, it may also be the case that differences across fields stem in part from variations in *citation practices* (of the patentees and/or of the patent examiners) across fields. As we shall see in the next section, the results of the model where β_1 is also allowed to vary by field are more telling, and correspond quite closely to what is known about these fields. Thus, there is reason to believe that the findings here do reflect real differences across fields, but that the fixity of β_1 does not allow for the full picture to emerge.

We turn now to the parameters associated with institutional variations, starting with universities. Ignoring 1965,¹⁷ we see that university patents became increasingly more "fertile" than corporate ones in the 1970's and early 1980's, but their relative citation intensity declined in the late 1980's. We have previously gotten similar results using a model that did not explicitly deal with the time path of citations (Henderson, Jaffe and Trajtenberg, 1994).¹⁸ Government patents, on the other hand, are significantly *less* fertile than corporate patents at each of the

16. If the base period reached further back (say into the 1950's), and if we could observe a longer period of "recovery" of fertility post-1985, then one could argue perhaps that the drop in "fertility" during the 1970's and 1980's was associated with the transition between different "general purpose technologies." As it is this remains only a conjecture that cannot be substantiated with these data.

17. We have university patents starting only in 1965, not in 1963, and there were a very small number of such patents that year. Not surprisingly, the university effect for the first cited time period is estimated with relatively little precision.

18. On the other hand, the earlier paper utilized a much finer set of technology field controls, at the level of about 400 distinct patent classes.

sub-periods, with a moderate upward trend over time in their relative citation intensity (from 72% of contemporaneous corporate in 1963-66 to 75% in 1981-85), except for a decline in the last period. Their overall lower fertility level may be due to the fact that these labs had been traditionally quite isolated from mainstream commercial innovations, and thus those innovations that they did choose to patent were in some sense marginal. By the same token the upward trend in the fertility ratio may be due to the increasing "openness" of federal labs, and their efforts to reach out and make their innovations more commercially oriented. However, for the time being, and until we gather additional evidence, this interpretation must be regarded as conjectural.

The region coefficients show very significant "localization" effects. That is, US patents are much more likely to cite previous US patents than are patents granted to inventors of other countries. Furthermore, the extent to which patents granted to foreign residents are likely to cite US patents seems to depend upon a metric of geographic and cultural proximity: Canada is highest with a coefficient of 0.65, followed by Europe with 0.5, and Japan with 0.44. (The fifth category, ROW, contains a small number of patents from a large number of other countries, and hence it is difficult to interpret.)

Thus these data do not support the claim that Japan is particularly apt at exploiting technological advances made in the US (as has been often suggested). In fact, patents granted to Japanese inventors are less than *half* as likely to make citations to US patents than patents granted to U.S. residents. Once again, if we interpret a citation as an indication that the citing patent "builds upon" the cited patent (or as evidence of a spillover from the cited to the citing patent), then our results indicate that, at least in the realm of patented inventions, Japanese inventors draw from US inventions much less than do U.S. residents.¹⁹

Finally, the estimate of $\beta_1 = 0.2$ means that the citation functions reaches its maximum at about 5 years. This estimate is not surprising, given the empirical citation distributions shown in Figures One and Two. The R^2 of 0.52 is fairly high for models of this kind, suggesting that the postulated double exponential combined with the effects that we have identified fit the data

19. From a purely statistical perspective, one cannot rule out the interpretation that Japanese inventors rely on U.S. inventions as much as their U.S. counterparts, but somehow manage to avoid citing them. Since the examiners are the ultimate arbiters of what citations appear on the patents, this would require that U.S. patent examiners are differentially lax in their duties vis à vis Japanese inventors. This seems unlikely.

reasonably well.²⁰

5.2 The Full Model: Allowing β_1 to vary

We present in Table 4 the results from the estimation of the full model, including variations in β_1 . We discuss below each set of results separately.²¹

Technological Fields. In addition to the multiplicative factor, α_g , we allow now for variations in β_1 according to the technological field of the cited patent. Thus, for example, field 3, Electronics, Optics and Nuclear technology has a coefficient $\alpha(g=3) = 2.00$, meaning that patents in this field are likely to get twice as many citations as in the base field "All Other", and 70-88% more than in the four other fields. However, its β_1 is 30% higher than that of the base, which means that the rate of citations to these patents decays faster than for others, and thus the initial "citation advantage" fades quickly over time.

The effect of allowing these joint variations in α and β_1 is shown graphically in Figure Three. This Figure shows the predicted citation function for potentially cited patents in the Electronics, Optics and Nuclear Technology field. The Figure shows that both models yield parameters that predict a greater overall rate of citation of these patents than the base group. Allowing β_1 to vary shows, however, that this greater overall citation intensity is not uniform over time. These patents are much more highly cited in the first few years, but also show faster obsolescence, so that in later years they are actually less highly cited than the base group.

Figure Four shows the citation functions for each of the Technological Fields. Because the relative frequencies are difficult to see on the graph at higher lags, Part (A) of Table Five also gives the value of the ratio of the citation probability of each of the technological fields to the citation probability of the base field, at different lags, i.e. 1, 5, 10, 20 and 30 years after the granting of the cited patents. Looking once again at Electronics, we see that the ratio starts indeed very high, 1.89, but after 12 years it is the same as the base field, after 20 years it

20. It is interesting that the fit of the *unweighted* model is considerably poorer, with an R^2 of about 15 percent. Recall that there are a significant number of observations with zero observed citation frequency, for which the model cannot fit for plausible parameter values. These observations correspond, however, to cells with few potentially cited and/or citing patents, and hence they get reduced weight in the weighted regression, thus improving the overall fit.

21. The results for cited time period and citing year are virtually the same as for the benchmark model and hence are excluded from the Table and not discussed here.

declines to 0.64, and further down to 0.36 after 30 years. This implies that this field is extremely dynamic, with a great deal of "action" in the form of follow-up developments taking place during the first few years after an innovation, but also with a very high obsolescence rate: a decade down the road the wave of further advances that built upon the original innovation subsides and is not very different from what happens in most other fields (it is even lower than in Drugs and Medical). Thirty years later citations have virtually ceased, relative to the rate in some other fields. Commonly held perceptions about the technological dynamism of this field are thus amply confirmed by these results, and given a precise quantitative expression.

For other fields the results are perhaps less dramatic but still interesting. Drugs and Medical starts off at 133%, of the base citation frequency, but due to the *low* obsolescence rate it grows over time (at a slow pace), so that 10 years later it stands at 150%, and 20 years down the road it has grown to over 170% relative to the base field, and more than that relative to *all* other fields. The conjecture here is that due to the long lead times in pharmaceutical research, including the process of getting approval from the FDA, follow up developments are slow in coming. In other words, whereas in Electronics a given innovation has very little impact 10-20 years later because the field is evolving so fast, in pharmaceuticals a new drug may prompt follow-up innovations much later, after its medical and commercial viability have been established.

As to the Chemical field, we see that it starts off at 127% of the base field, but due to a high obsolescence rate the advantage fades over time (though not as fast as in Electronics), falling behind the base field in less than a decade. The Mechanical field is similar to the base field, slowly losing ground over time. Note that after 20 years the ranking of fields changes dramatically compared to the ranking at the beginning. In all, we see that these results present a much more complex picture than having just multiplicative factors for fields, and that there is a great deal to learn from them, beyond serving just as controls.

Location. The most interesting result in this context, beyond those already discussed in the benchmark model, is that all foreign countries except Japan have significantly *lower* β_1 's than the US. The interpretation of a lower β_1 for a category of the *citing* patents is that their propensity to cite (i.e. to "absorb spillovers") *increases* over time relative to patents in the base

category. In the present context this means that the localization effects that we found, as reflected in the lower α 's for foreign countries, fade over time.²² As shown in Figure Five, the citation rates for European and Canadian patents eventually catch up to those for U.S. patents. That is, at first the spillovers from US patents flow mostly to inventors residing in the US: the probability that a foreign inventor would cite a patent of a U.S. inventor is 42-56% lower than that of a US-resident inventor 1 year after grant. However, the relative citation probabilities for foreign inventors grow over time (due to the lower β 's for foreign countries), so that 20 years down the line the difference has shrunk to 20-36%. This is not true of Japan: the "receptiveness" of Japanese inventors to US inventions remains low, 56% lower than that of U.S. inventors, and after a while significantly lower even than that of other foreign inventors.

This "fading" effect in the geographic dimension corresponds to the intuitive notion that knowledge eventually diffuses evenly across geographic (and other, such as institutional) boundaries, and that any initial "local" advantage in that sense will eventually dissipate. Once again, these results offer a quantitative idea of the extent of the initial localization, and the speed of fading.

Institutional Type. The results in this context do not change much vis à vis the benchmark model: the α 's for universities are slightly lower for most periods relative to the model with a fixed β , while a slightly smaller $\beta(U)$ compensates for that to some extent. The same holds for government patents, but the change for them is more significant. In fact, the upward trend is now significantly steeper, with the citation intensity relative to contemporaneous corporate going from 0.59 in 1963-66 to 0.68 in 1981-85, then declining then slightly to 0.66 in the last period. Figure Six shows the citation functions for selected university and federal cohorts; the picture shows that the government inferiority gradually fades away. This can also be seen in Table 5.C, but it takes thirty years for the government citation rate to completely catch up.

22. We have anticipated this result, in a cruder way and only for a small sample of patents, in Jaffe, Trajtenberg, and Henderson, (1993). Notice that here we discuss only localization effects at the national level. However, the data would permit analogous analysis at lower levels of geographic aggregation.

VI. CONCLUSION

The computerization of patent citations data provides an exciting opportunity to examine the links among inventions and inventors, over time, space, technology and institutions. The ability to look at very large numbers of patents and citations allows us to begin to decompose overall citation flows in ways that better reflect reality. This paper represents an initial exploration of these data. Many variations that we have not explored are possible, but this initial foray provides some intriguing results. First, we confirm our earlier results on the geographic localization of citations, but now provide a much more compelling picture of the process of diffusion of citations around the world over time. Second, we find that federal government patents are significantly less cited than corporate patents, although they do have somewhat greater "staying power" over time. Third, we confirm our earlier findings regarding the importance or fertility of university patents. Interestingly, we do not find that university patents are, to any significant extent, more likely to be cited after long periods of time. Finally, we show that citation patterns across technological fields conform to prior beliefs about the pace of innovation and the significance of "gestation" lags in different areas, with Electronics, Optics and Nuclear Technology showing very high early citation but rapid obsolescence, while Drugs and Medical Technology generates significant citations for a very long time.

The list of additional questions that could be examined with these data and this kind of model is even longer. First, it would be interesting to examine if the geographic localization differs across the corporate, university and federal cited samples. Second, the interpretation that we give to the geographic results could be strengthened by examining patents granted in the U.S. to foreign corporations. Our interpretation suggests that the lower citation rate for foreign inventors should not hold for this group of cited patents. Third, we could apply a similar model to geographic regions within the U.S., although some experimentation will be necessary to determine how small such regions can be and still yield reasonably large numbers of citations in each cell while controlling for other effects. Fourth, it would be useful to confirm the robustness of these results to finer technological distinctions, although our previous work with citations data lead us to believe that this will not make a big difference. Finally, we would like to investigate the feasibility of modelling obsolescence as a function of accumulated patents. Caballero and

Jaffe (1993) implemented this approach, but in that analysis patents were not distinguished by location or technological field.

REFERENCES

- Caballero, Ricardo J. and Adam B. Jaffe (1993), "How High are the Giants' Shoulders: An Empirical Assessment of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth," in Olivier J. Blanchard and Stanley Fischer, eds, NBER Macroeconomics Annual 1993, Cambridge: M.I.T. Press
- Griliches, Zvi (1990), "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature* 92:630-653
- Grossman, Gene M. and Elhanan Helpman (1991), "Quality Ladders in the Theory of Growth," *Quarterly Journal of Economics*, 106:557-586
- Jaffe, Adam B., Rebecca Henderson and Manuel Trajtenberg (1993), "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics* 108: 577-598
- Krugman, Paul (1991), Geography and Trade, Cambridge: M.I.T. Press
- Romer, Paul M. (1990), "Endogenous Technological Change," *Journal of Political Economy* 98: S71-S102
- Trajtenberg, Manuel, Rebecca Henderson and Adam B. Jaffe (1992), "Ivory Tower versus Corporate Lab: An Empirical Study of Basic Research and Appropriability," *NBER Working Paper No. 4146, August 1992*

TABLE ONE
Summary Statistics by Institutional Subsample

	U.S. Corporations	U.S. Universities	U.S. Government
Range of Cited Patents	1963-1990	1965-1990	1963-1990
Range of Citing Patents	1977-1993	1977-1993	1977-1993
Total Potentially Cited Patents	88,257 (1 in 10)	10,761 (Universe)	38,254 (Universe)
Total Citations	321,326	48,806	109,729
Mean Citations	3.6	4.5	2.9
Mean Cited Year	1973	1979	1973
Mean Citing Year	1986	1987	1986
<u>Percent Cited Patents by Field</u>			
Drugs & Medical	4.89	29.12	3.36
Chemical exc. Drugs	30.37	28.71	20.73
Electronics, Optics & Nuclear	26.16	27.39	45.40
Mechanical	28.18	9.51	17.09
Other	10.39	5.28	13.42
<u>Percent Citations by Region</u>			
U.S.	70.6	71.8	70.8
Canada	1.6	1.7	1.7
E.E.C	14.5	13.2	16.8
Japan	11.3	11.0	8.6
Rest of World	1.9	2.4	2.1

TABLE TWO
Distributions of Citations Across Cited Field, by Citing Region
(Percent)

	U.S.	Canada	E.E.C.	Japan	Rest of World	Total
Drugs & Medical	9.24	7.35	7.07	4.91	10.74	8.45
Chemical exc. Drugs	26.77	24.12	27.83	23.78	26.27	26.56
Electronics Optics & Nuclear	32.77	29.16	31.92	44.66	27.7	33.75
Mechanical	21.14	24.47	23.82	21.72	22.48	21.68
Other	10.08	14.90	9.36	4.93	12.81	9.56
Total Citations	339657	7945	71626	51168	9465	479861

TABLE THREE
Weighted Non-linear Least Squares Estimation of Base Model

	Parameter	Asymptotic Standard Error	t-statistic for H ₀ : Parameter=1
Citing Year Effects (Base=1963-65)			
1966-70	0.747	0.02871	-8.82
1971-75	0.691	0.02820	-10.97
1976-80	0.709	0.03375	-8.62
1981-85	0.647	0.03647	-9.69
1986-90	0.728	0.04752	-5.72
Cited Year Effects (Base=1977)			
1978	1.115	0.03449	3.32
1979	1.223	0.03795	5.88
1980	1.308	0.03943	7.80
1981	1.400	0.04217	9.48
1982	1.511	0.04637	11.01
1983	1.523	0.04842	10.80
1984	1.606	0.05209	11.64
1985	1.682	0.05627	12.12
1986	1.753	0.06073	12.40
1987	1.891	0.06729	13.24
1988	1.904	0.07085	12.76
1989	2.045	0.07868	13.29
1990	1.933	0.07795	11.97
1991	1.905	0.07971	11.36
1992	1.994	0.08627	11.52
1993	1.956	0.08918	10.73
Technological Field Effects (Base="All Other")			
Drugs and Medical	1.409	0.01798	22.73
Chemicals Exc. Drugs	1.049	0.01331	3.65
Electronics, Optics and Nuclear	1.360	0.01601	22.51
Mechanical	1.037	0.01370	2.69
Citing Country Effects (Base=U.S.)			
Canada	0.647	0.00938	-37.59
E.E.C.	0.506	0.00534	-92.49
Japan	0.442	0.00542	-102.99
Rest of World	0.506	0.00824	-59.93
University/Corporate Differential by Cited Time Period			
1965	1.191	0.12838	1.49
1966-70	0.930	0.04148	-1.70
1971-75	1.169	0.02419	7.00
1976-80	1.216	0.01765	12.26
1981-85	1.250	0.01718	14.55
1986-90	1.062	0.01746	3.57
Federal Government/Corporate Differential by Cited Time Period			
1963-65	0.720	0.04592	-6.11
1966-70	0.739	0.02498	-10.45
1971-75	0.744	0.01531	-16.71
1976-80	0.759	0.01235	-19.51
1981-85	0.754	0.01284	-19.15
1986-90	0.709	0.01551	-18.78
BETA1*	0.213	0.00247	86.28
BETA2*	3.86E-06	1.97E-07	19.61

26,975 observations; R-square=.5161

*t-statistic is for H₀: parameter=0

TABLE FOUR
Weighted Non-linear Least Squares Estimation of Model with Varying Beta1

	Parameter	Asymptotic Standard Error	t-statistic for H ₀ : Parameter=1
Technological Field Effects (Base="All Other")			
Drugs and Medical	1.308	0.03348	9.18
Chemicals Exc. Drugs	1.308	0.03338	9.22
Electronics, Optics and Nuclear	2.003	0.04734	21.18
Mechanical	1.120	0.02971	4.04
Citing Country Effects (Base=U.S.)			
Canada	0.567	0.01626	-26.61
E.E.C.	0.433	0.00907	-62.54
Japan	0.444	0.01069	-52.02
Rest of World	0.434	0.01399	-40.46
University/Corporate Differential by Cited Time Period			
1965	1.199	0.12180	1.64
1966-70	0.853	0.04600	-3.20
1971-75	1.110	0.03494	3.15
1976-80	1.161	0.02800	5.74
1981-85	1.225	0.02469	9.12
1986-90	1.075	0.02013	3.74
Federal Government/Corporate Differential by Cited Time Period			
1963-65	0.593	0.04435	-9.18
1966-70	0.634	0.03030	-12.07
1971-75	0.654	0.02150	-16.11
1976-80	0.678	0.01788	-18.02
1981-85	0.684	0.01592	-19.88
1986-90	0.658	0.01578	-21.68
BETA1*	0.198	0.00345	57.50
BETA2*	3.54E-06	1.95E-07	18.13
Beta1 Multipliers by Technological Field (Base is "All Other")			
Drugs and Medical	0.932	0.01422	-4.76
Chemicals Exc. Drugs	1.158	0.01702	9.29
Electronics, Optics and Nuclear	1.288	0.01779	16.20
Mechanical	1.054	0.01588	3.38
Beta1 Multipliers by Cited Insitution Class (Base is Corporate)			
University	0.978	0.01195	-1.88
Federal Government	0.932	0.01231	-5.55
Beta1 Multipliers by Citing Country (Base is U.S.)			
Canada	0.914	0.01542	-5.59
E.E.C.	0.899	0.01113	-9.09
Japan	1.002	0.01409	0.16
Rest of World	0.900	0.01711	-5.82

26,975 observations; R-squared=.5338

Estimation also included cited time period and citing year effects as in Table Three.

*t-statistic is for H₀: parameter=0

TABLE FIVE

(A) Citation Probability Ratio by Technological Field

Technological Field	Beta1 Multiplier	Lag				
		1	5	10	20	30
DRUGS & MEDICAL	0.932	1.33	1.40	1.50	1.71	1.96
CHEMICAL	1.158	1.27	1.12	0.96	0.70	0.51
ELECTRONICS ETC.	1.288	1.89	1.50	1.13	0.64	0.36
MECHANICAL	1.054	1.11	1.06	1.01	0.91	0.81
OTHER	1.000	1.00	1.00	1.00	1.00	1.00

(B) Citation Probability Ratio by Geographic Area

Location	Beta1 Multiplier	Lag				
		1	5	10	20	30
CANADA	0.914	0.58	0.62	0.67	0.80	0.95
EUROPE	0.899	0.44	0.48	0.53	0.65	0.79
JAPAN	1.002	0.44	0.44	0.44	0.44	0.44
REST OF WORLD	0.900	0.44	0.48	0.53	0.64	0.78
US	1.000	1.00	1.00	1.00	1.00	1.00

(C) Citation Probability Ratio by Institution

Research Institution	Beta1 Multiplier	Lag				
		1	5	10	20	30
UNIVERSITIES 1981-1985	0.978	1.23	1.25	1.28	1.34	1.40
UNIVERSITIES 1986-1990	0.978	1.08	1.10	1.12	1.18	1.23
FEDERAL LABS 1981-1985	0.932	0.69	0.73	0.78	0.90	1.03
FEDERAL LABS 1986-1990	0.932	0.67	0.70	0.75	0.86	0.99
CORPORATE	1.000	1.00	1.00	1.00	1.00	1.00

FIGURE ONE

Raw Distribution of Citations by Citing Region and Lag

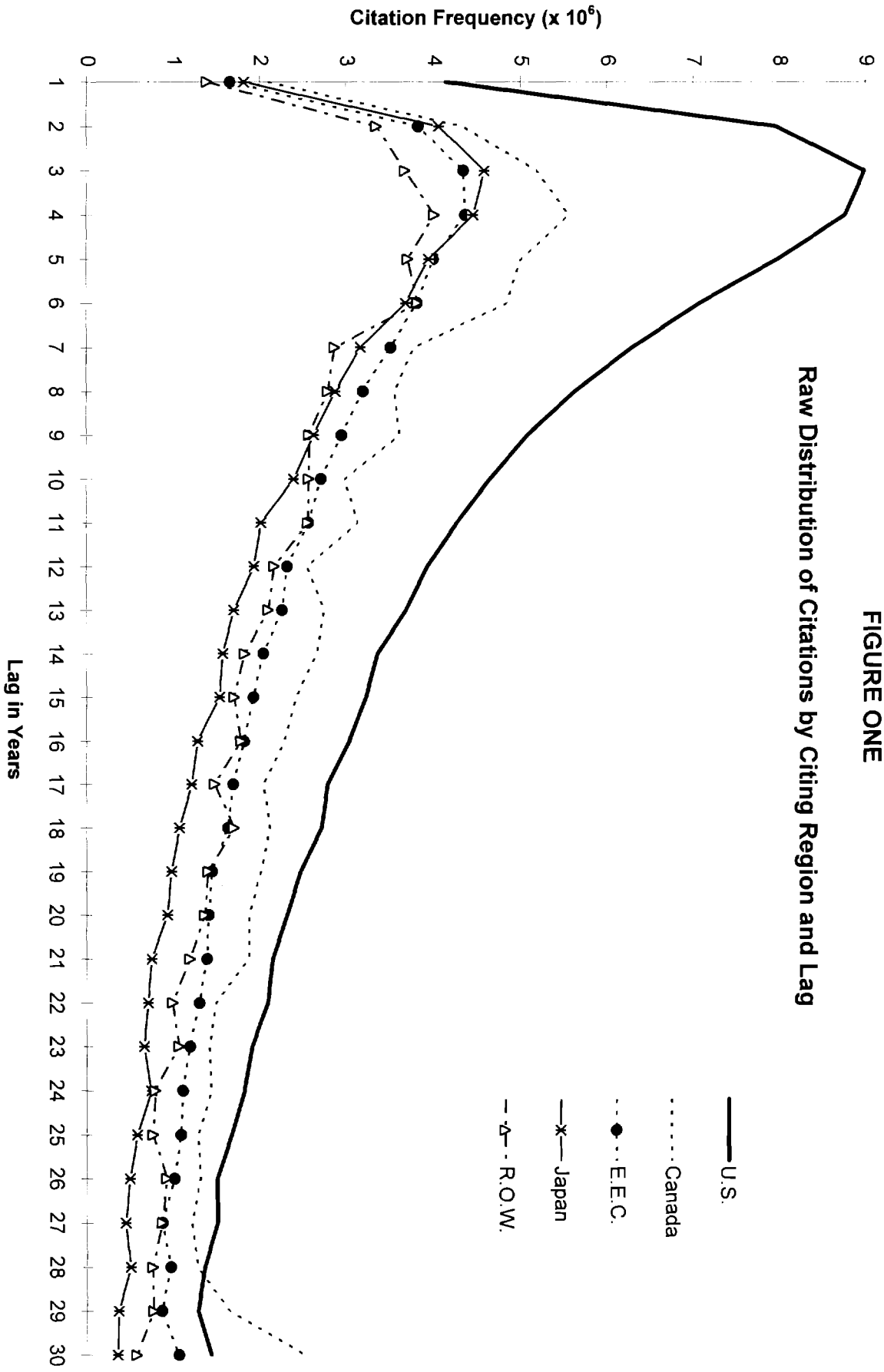


FIGURE TWO

Raw Distribution of Citations by Source Institution and Lag

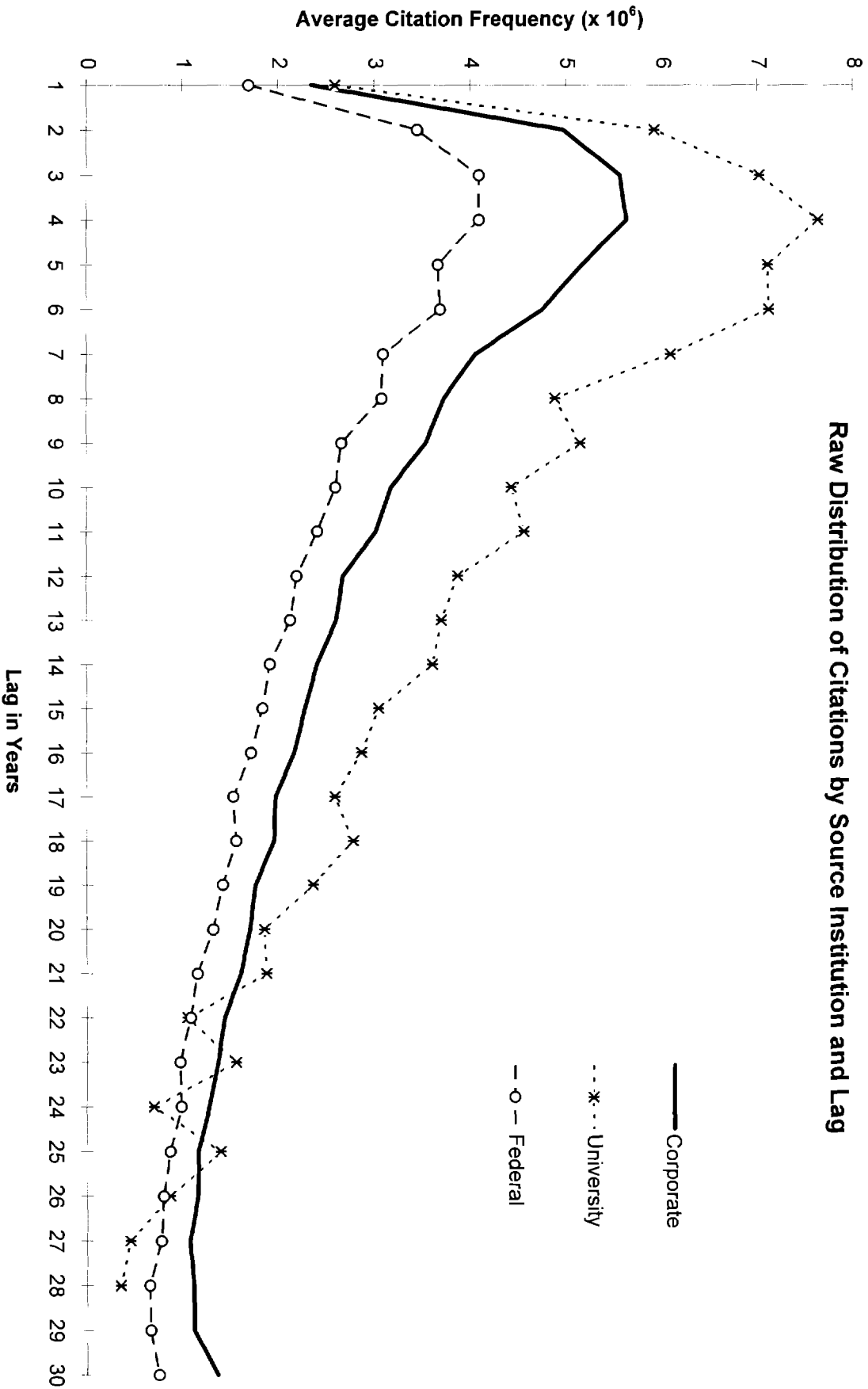


FIGURE THREE

Comparison of Constant-beta1 and Variable-beta1 Predictions for Electronics, etc.

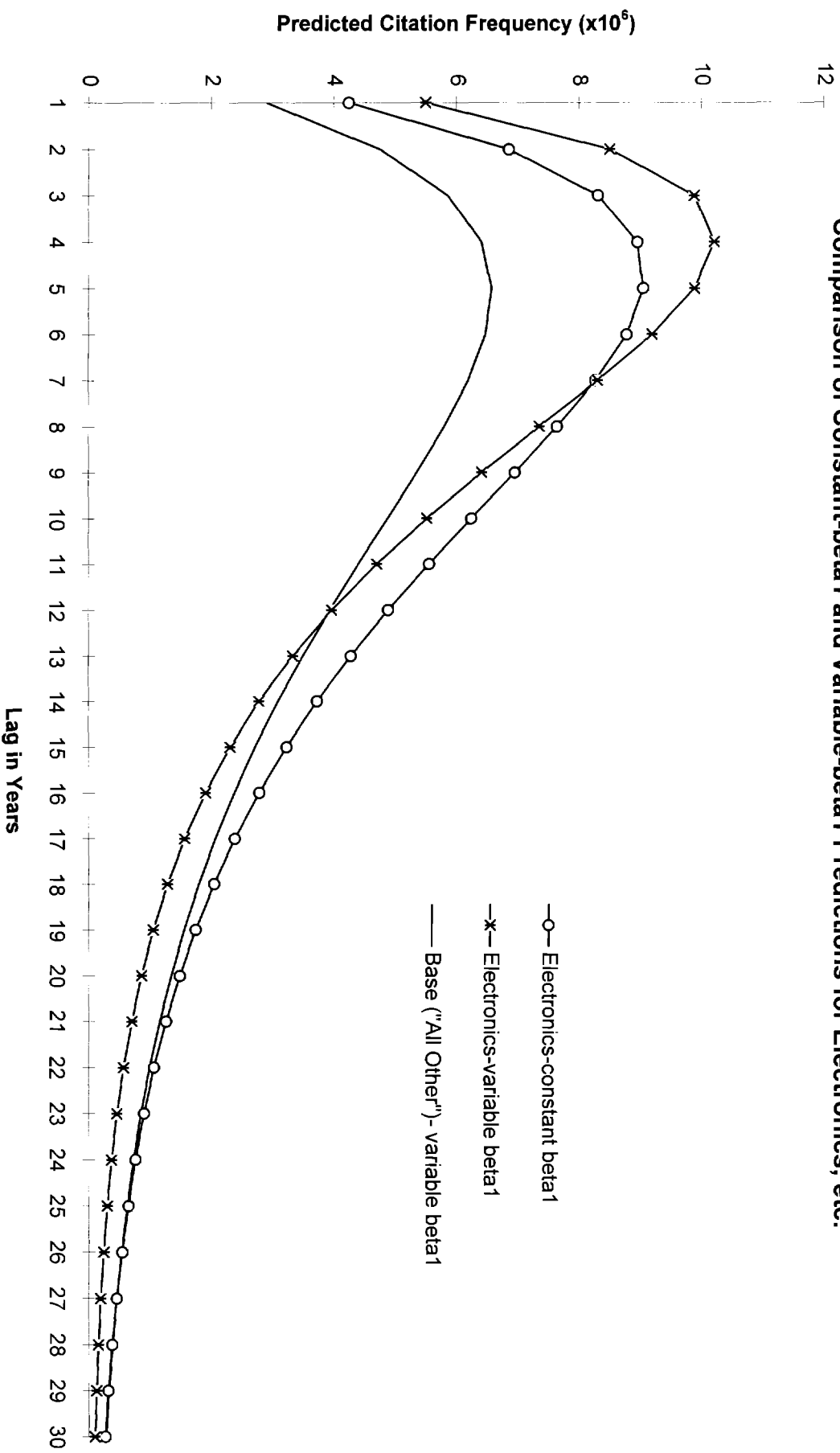


FIGURE FOUR

Fitted Citation Frequency by Technological Field

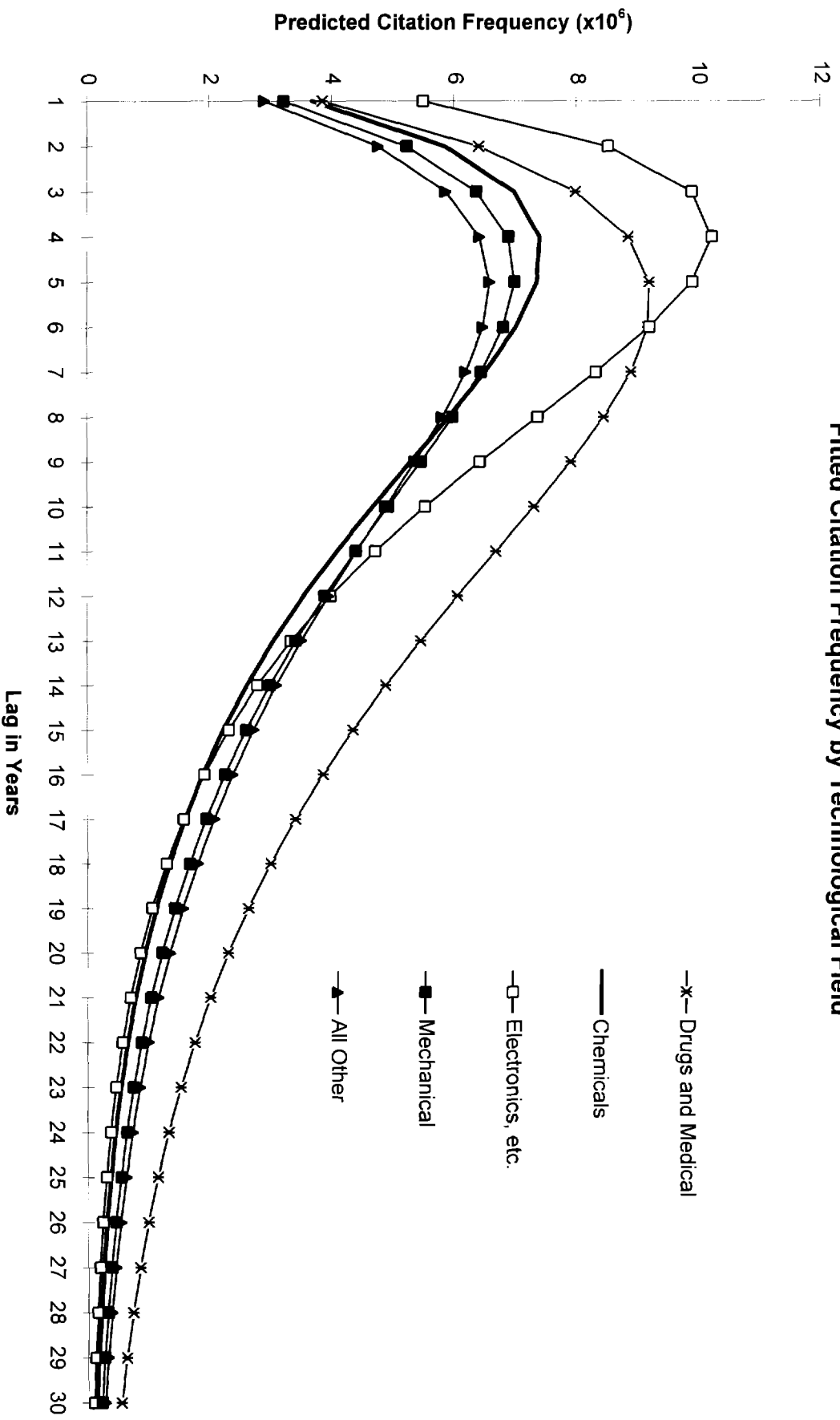
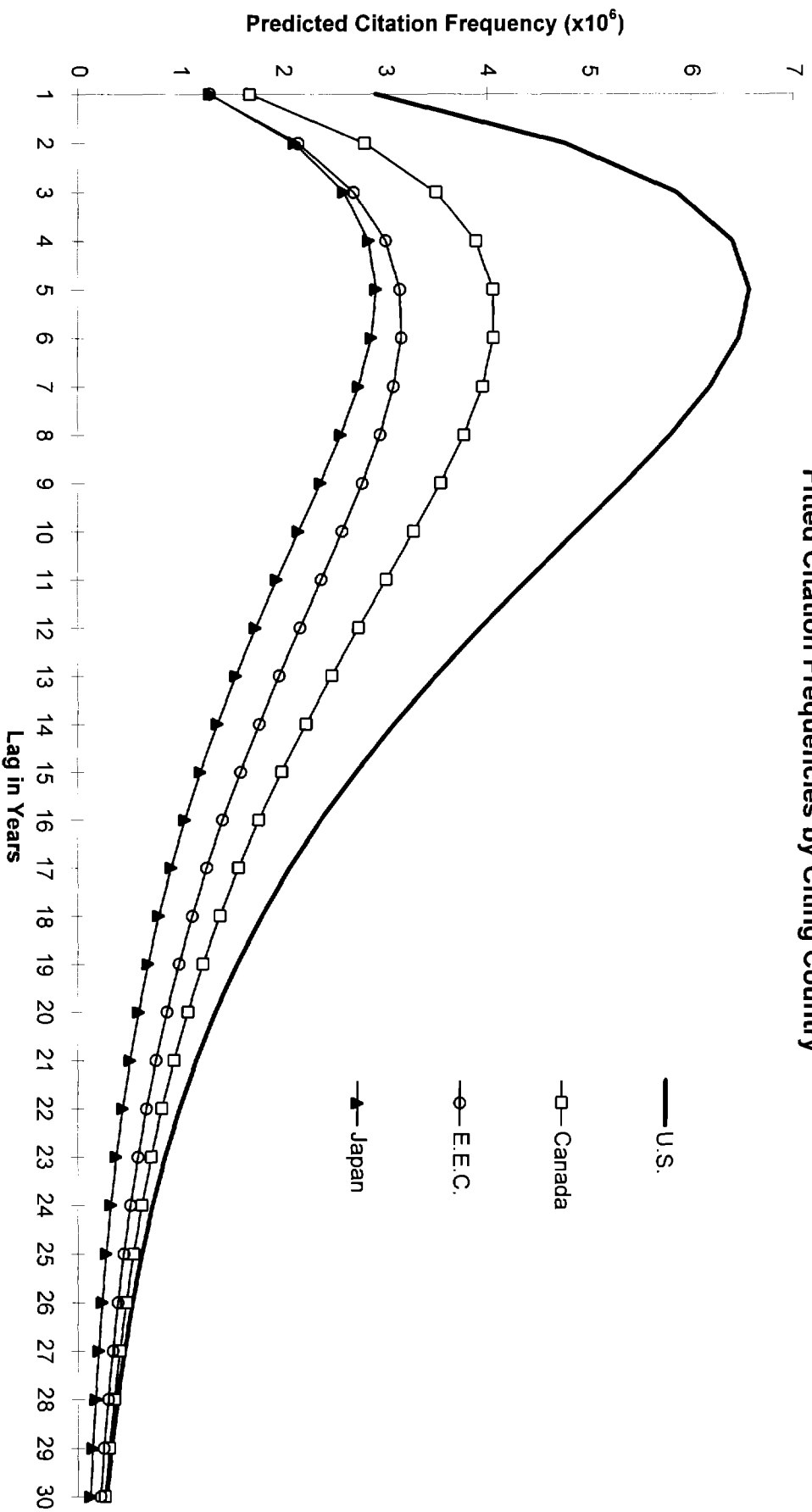


FIGURE FIVE

Fitted Citation Frequencies by Citing Country



Note: R.O.W. is omitted; it is indistinguishable from E.E.C. line.

FIGURE SIX

Fitted Citation Frequency by Cited Institution Type

