DRAFT

Citations and renewal:

A window into public and private patent value

Alan C. Marco

Deputy Chief Economist

US Patent and Trademark Office[1]

24 February 2012

Abstract

Using historical information on patent citations and patent maintenance in the US, I investigate the informational content provided to technology markets by the renewal decision. Using parametric and non-parametric survival time estimation I confirm the finding by other researchers that patents that are renewed longer are also more highly cited. The distinction between highly renewed patents and non-renewed patents exists prior to any observable renewal decision by the patent holder. I also identify several new findings:

- During each renewal period, there exists a significant drop in patent citations, followed by a rebound once the renewal decision is observed.
- The rebound effect exists for both renewed and non-renewed patents.
- While renewed patents are more highly cited over their entire lifetimes, the size of the rebound effect is larger for non-renewed patents.

The results are precisely measured, and robust to the specification, the sample period, and the definition of when the citations occur (e.g., the application date or the issue date). While preliminary, the findings suggest that citers delay citation to observe renewal decisions. Further, the rebound effect suggests that there are two types of citers: those who benefit from expiration and those who benefit from continued exclusivity by the patent holder. This research takes an important step towards distinguishing between different *types* of patent citations, and also towards how economists should interpret the information contained in patent citations.

Keywords: Patent citations, patent maintenance, patent renewal, patent valuation, survival analysis, hazard rate.

[1] The views expressed represent those of the author and not those of the USPTO.

1 Introduction[2]

Economists have long studied both patent renewal (using maintenance fees) and patent citations (citations by later granted patents) as correlates with (or proxies of) patent value. Patents that are maintained for a longer period of time must be more valuable to patent holders than those that are not maintained. For instance, if a patent holder is unwilling to pay a maintenance fee of $1500 to renew its patent, then one can infer that the remaining patent life has an expected NPV less than $1500. So-called "renewal models" have become one of the primary ways in which economists estimate the overall distribution of patent value (Pakes, Pakes and Schankerman, Lanjouw, Pakes, and Putnam). In a recent paper, Bessen found that patent citations are a predictor of patent renewal, indicating that the two are correlated prior to the renewal decision. The causality of the correlation is still in doubt: are citations and renewal both simply correlated with value? Or, do citations themselves indicate more reliance on that technology by the market, thus making renewal more advantageous for the patent holder? In this paper, I turn the question on its head to ask whether observed renewal events influence the hazard rate of later citations.

Before turning to the analysis it is useful to discuss more generally the role of patent citations in the economics literature. A "stylized fact" in economics is that highly cited patents represent inventions that are more valuable—on average—than those that are less frequently cited, all other things equal. The citations may reflect knowledge flows from the cited patent holder to the citing patent holder (the citer). They may reflect "follow-on innovation" where other firms build on the innovation of the pioneer inventor. Self-citations may reflect cumulative innovation by the incumbent patent holder, building on its own innovation. Or, citations may reflect the attempt by competitors to enter the product space with substitute products. In all of these cases it is important to distinguish between the social value of the patented innovation, and the private value to the patent holder.

The social value of a patented innovation includes the value to the patent holder and also the value to other users of the invention, including end users, consumers, and beneficiaries of knowledge spillovers (Jaffe, Trajtenberg, and Henderson). Much of the social value of the innovation may not be appropriated by the patent holder, leaving open the question of whether the

[2] References to the extant literature are needed throughout the paper. Please excuse the dearth of them in this draft.

patent system provides the “right” incentives to innovators. In contrast, some patents may provide opportunities for rent-seeking; in the extreme case the private value of the patent right may exceed the public value of the innovation. In particular, in the case of an inappropriately granted patent, the “innovation” may have no social value.[3] Broad claims on that patent may enable transfers from licensees or litigants, which make the patent right valuable to the patent holder. However, transfer payments do not increase social welfare and the transaction costs ensure that the net impact on welfare is negative.

Most empirical research has validated the basic precept that patent citations are correlated with patent value. However, accounting separately for the private value to patent holders (appropriation, including appropriated knowledge flows) and the public value of the innovation (knowledge spillovers)—especially using patent citations—has largely eluded researchers [but see Michael Ward. Event studies on citations help to distinguish competitive effects from private appropriation.] This paper takes a first step towards distinguishing private from public value by exploiting the timing of patent citations that occur around the renewal decision of patent holders.[4]

I use parametric and non-parametric survival methods to investigate the hazard of citation before, during, and after renewal events. How should we expect patent citations and renewal to be correlated? There are reasons why citation rates could be either positively or negatively correlated with renewal rates. If both renewal rates and citation rates are correlated with value, they should be positively correlated. Or, the relationship could be directional: citations could “follow” renewals, or indeed renewals could “follow” citations. Alternatively, if citations indicate areas where other patent holders are doing research, we could imagine citation rates being higher for those inventions that are no longer under patent protection, corresponding to the notion that citations would “rush in” when follow-on innovators’ use of the knowledge becomes cheaper (e.g., no payment of royalties). In this way we might expect citation rates to increase after a patent has expired for failure to pay maintenance fees.

The non-parametric results show that there is a positive correlation between citation rates and renewal. However, a more in depth analysis shows that the dynamics are more nuanced. The data

[3] Such patents are the alleged weapon of patent “trolls.” However, might I recommend that commentators have focused too much on patent trolls, instead of the root culprit of “trollable patents.”
[4] Throughout I refer equivalently to maintenance events and renewal events. In the US, there is no distinction. Alternatively, at the EPO patent applicants pay maintenance fees to maintain their pending patent *applications*. Renewal fees are paid to maintain the *patent*. In the US, post-grant renewal fees are called maintenance fees, although the patent is said to have been renewed.

reveal that patent citations vary significantly from quarter to quarter. In particular, citation rates drop prior to maintenance events. Following the observation of the renewal decision, citation rates rebound *both for renewed and non-renewed patents*. However, the timing and magnitude of the rebound is different between the two cases. The results are preliminary, but provocative. In particular, they suggest that citers play a wait-and-see game. Further, the rebound effect suggests that there may be two types of citers: one group of citers waits to exploit public domain knowledge, the other group benefits from the exclusivity of the patent holder (perhaps through exclusive licensing). Lastly, I show a strong and robust increase in citation rates following the end of the statutory term of the patent for all patents, *even those that had previously expired due to non-payment of maintenance fees.*

2 Literature review

TBD

- Patent renewal studies (Pakes; Schankerman; Bessen; Lanjouw, Pakes, and Putnam)
- Spillovers and knowledge flows (Jaffe, Trajtenberg, and Henderson)
- NBER data description (Hall, Jaffe, and Trajtenberg)

3 Model

TBD

- Something that shows two types citers: one cites more after renewal and one after expiration.
- Perhaps an option model that could demonstrate the citation dip.
- Reference model by Ward on event studies of patent citations.

4 Data and Method

The data comprise two datasets. Patent maintenance records are obtained from the PTO's bulk downloads available from Google.[5] The file—updated daily—contains all maintenance fee events from September 1, 1981 to the present. Bibliographic patent data and patent citation data through 2010 are obtained from the USPTO's Technology Assessment and Forecast (TAF) Database.[6]

These data were used to track citation information (both self-citations – those citations made by the inventor or assignee – and citations by others) and renewal decisions for each utility patent issued from 1981-2009. Survival analysis was used to estimate the hazard rate of citations,[7] based

[5] Patent Maintenance Fees available at www.google.com/googlebooks/uspto-patents-maintenance-fees.html.
[6] TAF data description available at www.uspto.gov/web/offices/ac/ido/oeip/taf/data/.
[7] The hazard rate of citations can be thought of as measuring the citation rate in terms of patent citations per year. The only distinction is that the hazard rate is a continuous measure, and can be measured instantaneously, i.e., at a particular instance in time. Nonetheless thinking about this rate as "citations per year" is accurate at an intuitive level.

on whether or not the patent was renewed. For patents more than 12 years old, we can observe up to three renewal decisions.

A one percent sample was taken consisting of approximately 31,000 patents. The subject patents experienced 306,000 citations during the sample period, of which 43,000 are estimated to be self-citations.

US patents face maintenance fee payments at three points after issuance. The payments are due at 3.5, 7.5, and 11.5 years after issuance. Payments cannot be made more than 6 months early, and can be paid up to 6 months late with a penalty. Thus, there are three one-year windows during which one can observe renewal at any point within that window. Expiration due to non-payment formally takes place at 4, 8, or 12 years after issuance; however, petitions can be made for reinstatement in the case of unintentionally or unavoidably delayed payments, subject to certain restrictions and penalties [cite: USC XX section XX].[8] My sample contains 58,000 maintenance events, including first, second, and third renewals. Figures 1-3 show the frequency of renewal by various cohorts. For patents issued prior to 1999, we can observe all three renewal decisions. Figure 1 shows that 41.5% of patents are renewed to the full term, and around 16.5% are never renewed. For later cohorts (Figures 2-3), we can only observe two or one renewal decision. Renewal at the first maintenance payment has increased since 2000 (13% non-renewal).

The one-year window for payment—and the possibility of reinstatement—creates uncertainty over whether the patent will be renewed. Figure 4 shows the quarterly distribution of payments within the renewal windows for the entire sample. The vast majority of payments are made in the three months prior to the due date.

In this paper, I use survival analysis to estimate the hazard rate of receiving a citation, conditional on the renewal status of the patent. In the preliminary results, I do not exploit many observable characteristics of the subject or citing patents such as foreign ownership, family size, or technology classes. Nonetheless, the results are remarkably precise and robust.

---

[8] Reinstatements are rare; my data contain only 215 such events.

5 Non-parametric estimation

Non-parametric Kaplan-Meier estimation is used for the exploratory analysis. Figures 5-8 show the smoothed hazard estimates for the hazard rate being cited by a new patent. I classify patents into four groups based on whether they were renewed three times (at years 3.5, 7.5, and 11.5), two times (years 3.5 and 7.5), one time (year 3.5), or not at all.[9] The groups are labeled R3, R2, R1, R0, respectively.

5.1 Overall citation rates

The results for all the non-parametric estimates demonstrate that renewed patents have higher citation rates than those that are not renewed. In each figure, the vertical bars show the renewal windows, with an additional indicator at 17 years from date of issuance.[10] Figures 5 and 6 show the hazard estimates of being cited, excluding self-citations. The date of the citation is defined alternatively as the *grant date* of the citing patent (Figure 5) or the *application date* of the citing patent (Figure 6).[11] Throughout the analysis, the results are robust to the definition of the citation event date.

5.2 Renewal versus non-renewal

Several striking features present themselves in Figure 5. Even from birth (before renewal is observed), patents that *will be* fully renewed receive more citations than R2, R1, or R0. This rank ordering persists through about year ten. This suggests that even from day one, the "market" (in this case other patentees) understand the value of the patented invention even before observing the actual maintenance events. Because renewal rates and citation rates are positively correlated, the primary inference is that—on average–total social value is highest for those patents whose private value is also high.

[9] In some ways the payment of the issue fee is a de facto "first renewal."

[10] All patents issued prior to June 8, 1995 were subject to an expiration date of 17 years from the grant date. After that point, the US harmonized with other members of the World Trade Organization under the TRIPs agreement, and adopted a patent term of 20 years from the date of application. We have yet to observe the first expiration of these patents, so the 17 year date is the relevant expiration date for all observed expirations in the sample.

[11] References to prior art can be made throughout patent prosecution. Some references are included by the applicant at the time of application, some are added by the examiner at the time of examination, and some are added later by the applicant through an Information Disclosure Statement (IDS). An IDS can even be made after an allowance, just prior to issuance. Thus, the exact timing of the patent citation cannot be observed. The application date and grant date serve as bounds on the event date, so I perform each estimation using each of the two dates to mitigate measurement errors. In unreported results, I also use the median date between application and grant.

Interestingly, between the first and second renewal events R1 and R0 begin to converge, at which point both are expired. More strikingly, R2 declines significantly between the second and third renewal events, at which point it joins its expired brethren. [12] Because each decline occurs prior to the renewal decision, patent citations can be thought of as a leading indicator of renewal.

Following the third renewal, R3 maintains its higher, albeit declining, citation rate through the end of its statutory patent term. This suggests that, at least later in life that the primary distinction in patents is between those that are fully renewed, and those that are not. Note that for patents that are fully renewed, there is a very skew distribution of value: it is impossible to distinguish between those that were "barely" worth the third maintenance fee, and those that would have been worth 100 maintenance fees. It may be that those patents at the upper tail of this distribution are driving the separation between fully-renewed and less-than-fully-renewed patents, at the mean.

5.3 The off-patent effect

A somewhat surprising result is that the citation rates for all patents increase substantially after the end of the statutory patent term, and the statistical distinction between R3, R2, R1, and R0 essentially disappears. Moreover, the hazard rate is as high as, or higher than, the on-patent peak hazard rate. This result calls into question the stylized fact in the literature that most citation activity occurs within the first five years after grant [cite]. For R3, the off-patent jump makes sense: these are the most valuable patents, and they are now in the public domain. Competitors are free to use the inventions, and that activity may generate a burst of citations. However, the off-patent effect is quite surprising for R0-R2, given that they had expired prior to the end of the statutory patent term and experienced no discernible expiration bump after non-renewal.

The off-patent jump may have two explanations. First, information about non-renewal may not be readily available or reliable. The possibility of reinstatement may also create a chilling effect on using non-renewed inventions. Second, the off-patent activity may be generated by specialists (such as generic manufacturers). If products can be described as systems—the component parts of which are covered by patents—generic manufacturers may need access to all the related patents. In this way, a non-renewed patent is not useable until the entire system is useable. Thus the timing of the off-patent effect depends on the last renewed patent. Note that this trend can only be observed for patents issued prior to 1995. Thus, we do not want to draw too strong a conclusion from this observation with respect to newer patents.

[12] These results are attenuated if one defines the citation date at the date of application (Figure7).

5.4 Self-citations

Figures 7 and 8 show the non-parametric estimates for the hazard of self-citations. Whether measured by the date of the application or the date of the grant, the hazard rates follow similar patterns. Self-citation rates peak earlier than citations from other patent holders, suggesting that the patent holder's cumulative innovation occurs early on in the patents life relative. In comparison to citations from other patent holders, self-citations do not exhibit as clear a pattern with respect to renewal decisions. However, it is clear that in the first several years R2 and R3 patents are more highly cited than R1 or R0. By about year ten, most of the difference between groups disappears.

Interestingly there is an imprecisely measured off-patent effect, something that should not be relevant to the patent holder in terms of cumulative innovation. In fact, there is some evidence that those patents that are not renewed at all have **higher** self-citation rates than other patents.

There are two methodological problems inherent in using non-parametric hazard estimation. First, the results are sensitive to the particular kernel used for smoothing the hazard functions. In particular, the smoothing can hide some important variation in the hazard rates over analysis time. Second, it is difficult to control for observable characteristics. For that I turn to parametric estimation.

6 Parametric estimation

For the parametric estimation, I use maximum likelihood estimation assuming an exponential distribution for the duration between citations. The exponential distribution implies a constant hazard rate over analysis time. Based on the non-parametric results, this is clearly inappropriate. Thus, I develop a very flexible form for the hazard function over time by using quarterly fixed effects based on the time since issuance. The result is something akin to a Cox model, but without the proportional hazards assumption.

For the preliminary results, I control only for the application year of the subject patent (using annual fixed effects), and the calendar year of the analysis time (again, using annual fixed effects). The calendar year fixed effects control in part for the overall trends in citation rates over time. Other observable patent characteristics can be included, such as technology class, number of claims, location, etc. Additionally, one can control for time-varying covariates such as cumulative counts of previous citations received.[13]

The primary results of the parametric estimation are summarized below and in Figures 9-18.

6.1 Specification

- Specification
  - Fixed effects for application year and calendar year of the citation event
  - Quarterly fixed effects for time since issuance
- Estimation repeated for:
  - Citation date = grant date of citing patent versus citation date = application date of citing patent
  - Self-citation versus citation by others
  - Estimation is done separately for R0-R3
- Results shown graphically in terms of predicted hazard rates
  - Mean hazard rate (and confidence interval)
  - Median hazard rate (and interquartile range)

6.2 Results

- Predicted citation hazards across all specifications in Figures 9-12 show a robust and surprisingly precise effect.
  - Prior to each renewal decision **the citation rate drops almost to zero**.
  - This drop occurs at the **beginning** of the renewal window for patents that will be renewed, and at the **end** of the renewal period for non-renewed patents.
  - There is a **rebound effect** for both renewed and non-renewed patents
  - The **rebound is higher for non-renewed patents**. After the rebound there is a steady decline in citation rates.
  - After non-renewal **future renewal events are not relevant**.
  - There is a significant **off-patent effect** for all patents.
- Self-citation rates in Figures 13-18 show the same drop and rebound surrounding renewal events.

7 Conclusion

TBD

[13] Unreported regressions show that the qualitative and quantitative results below are robust across many specifications using several covariates.

Figure 1.

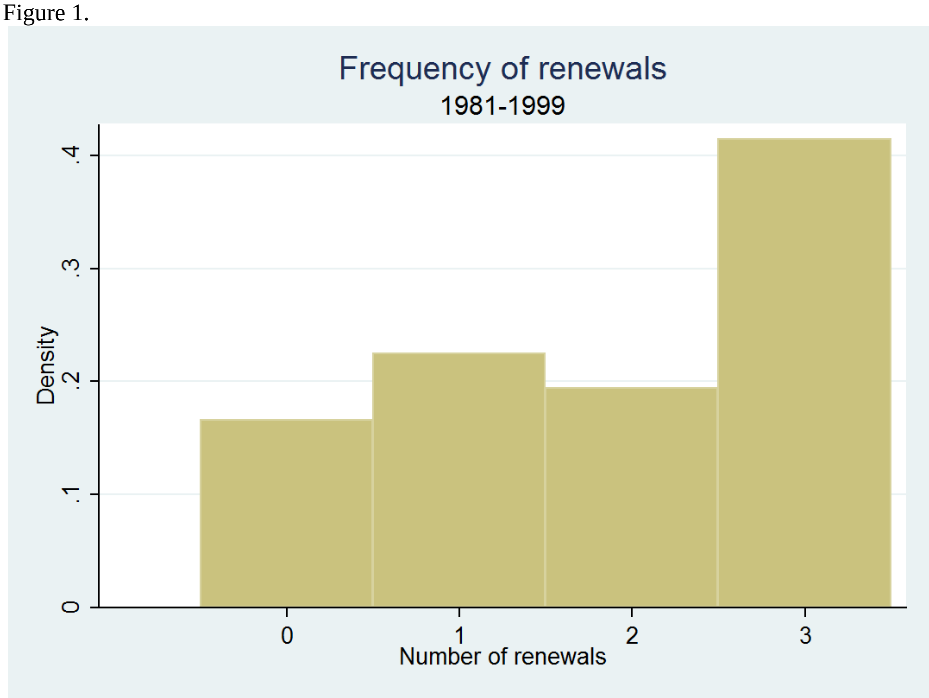

Figure 2.

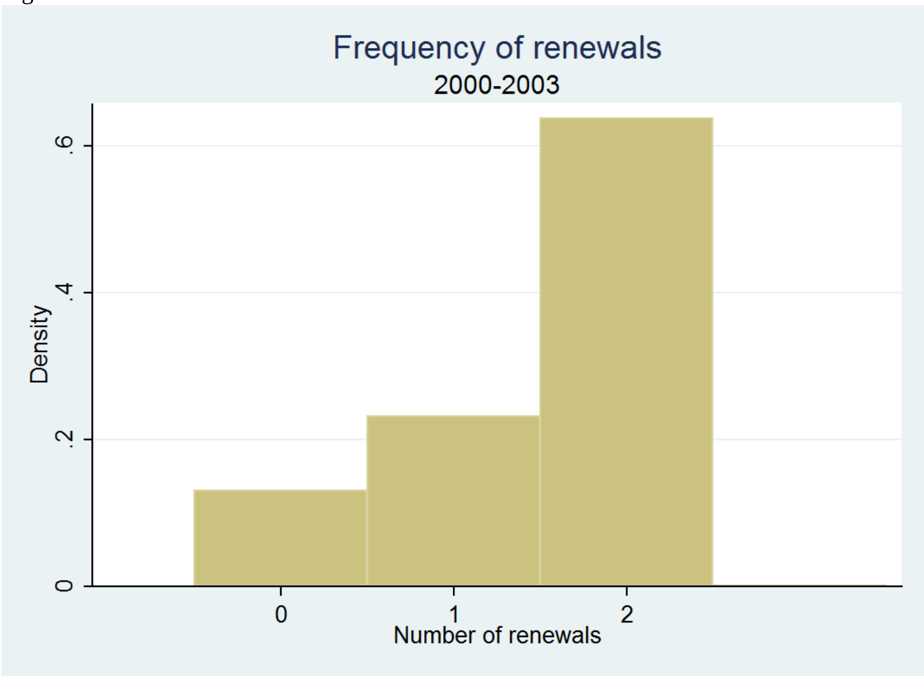

Figure 3.

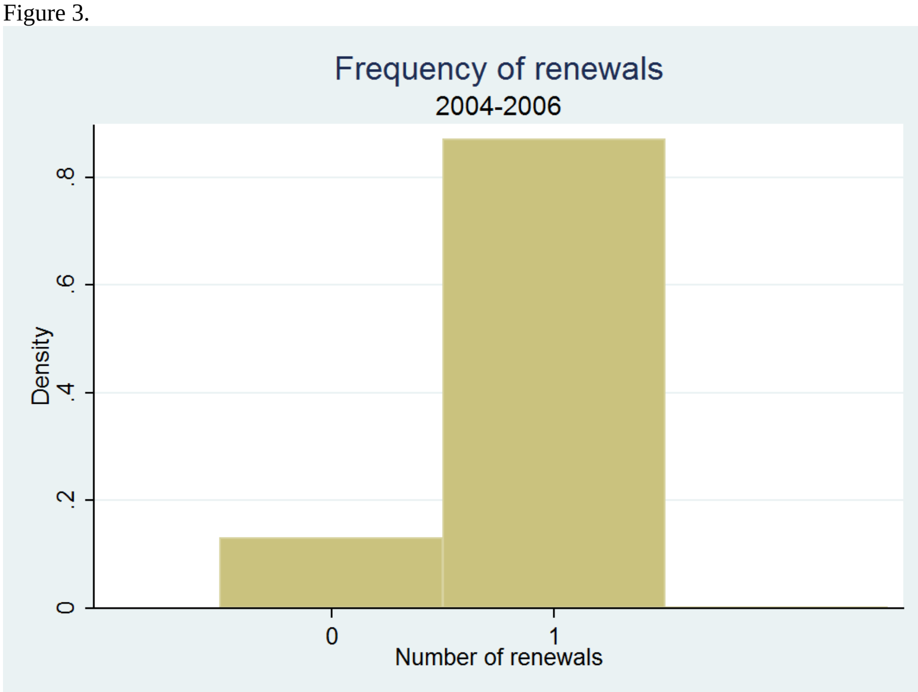

Figure 4.

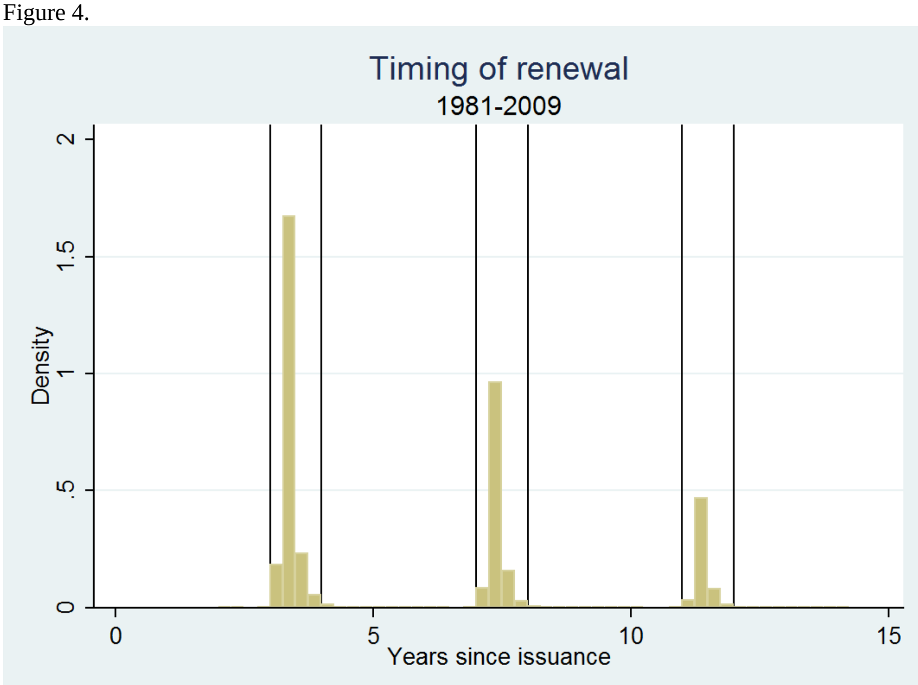

Figure 5. Hazard of being cited, from date of issuance, by **grant date of citing patent**, 1981-2009. R indicates number of lifetime renewals.

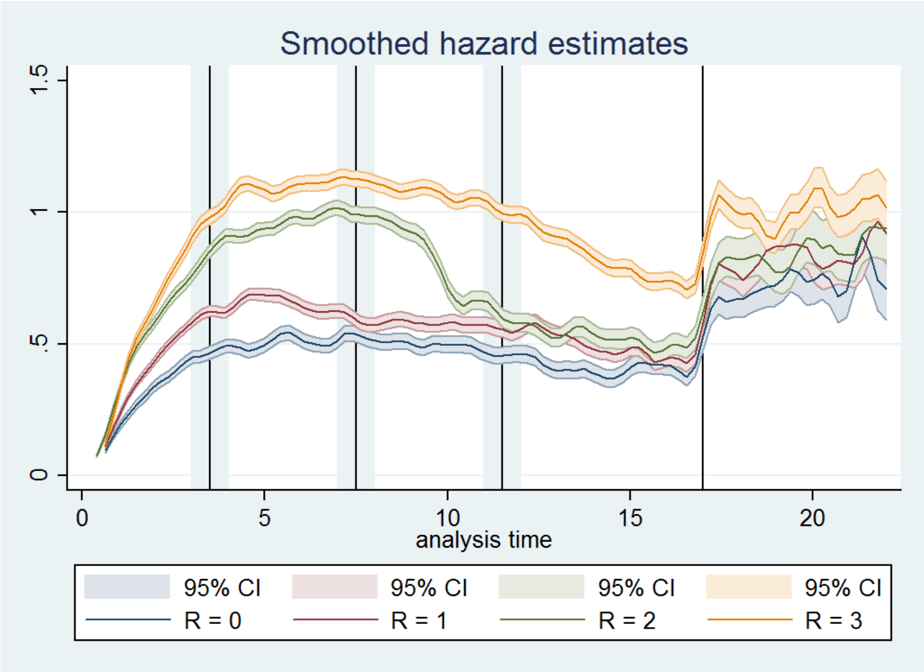

Figure 6. Hazard of being cited, from date of issuance, by **application date** of citing patent, 1981-2009. R indicates number of lifetime renewals.

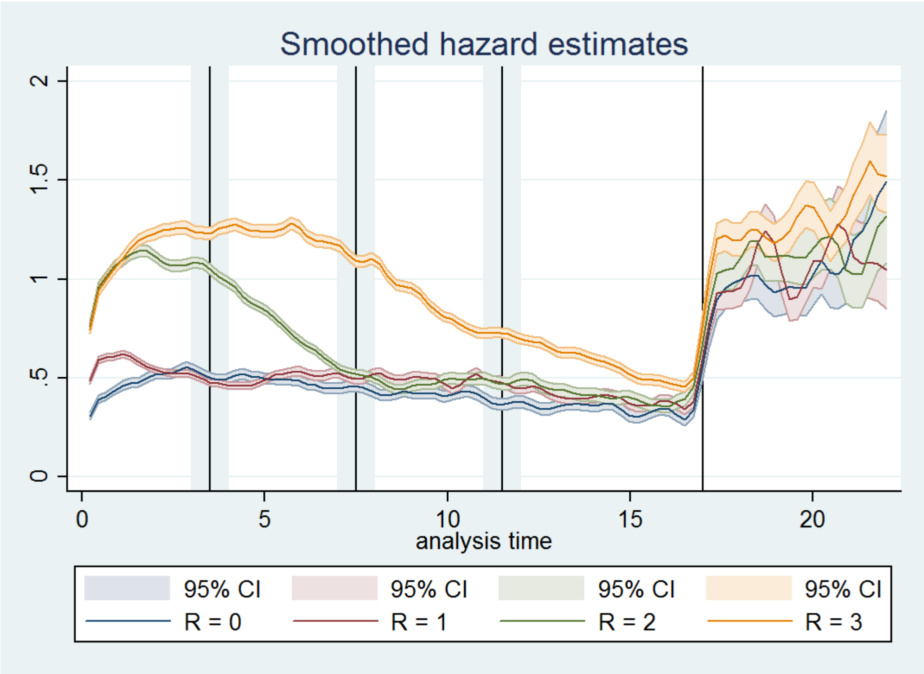

Figure 7. Hazard of **self-citation**, from date of issuance, by grant date of citing patent, 1981-2009. R indicates number of lifetime renewals.

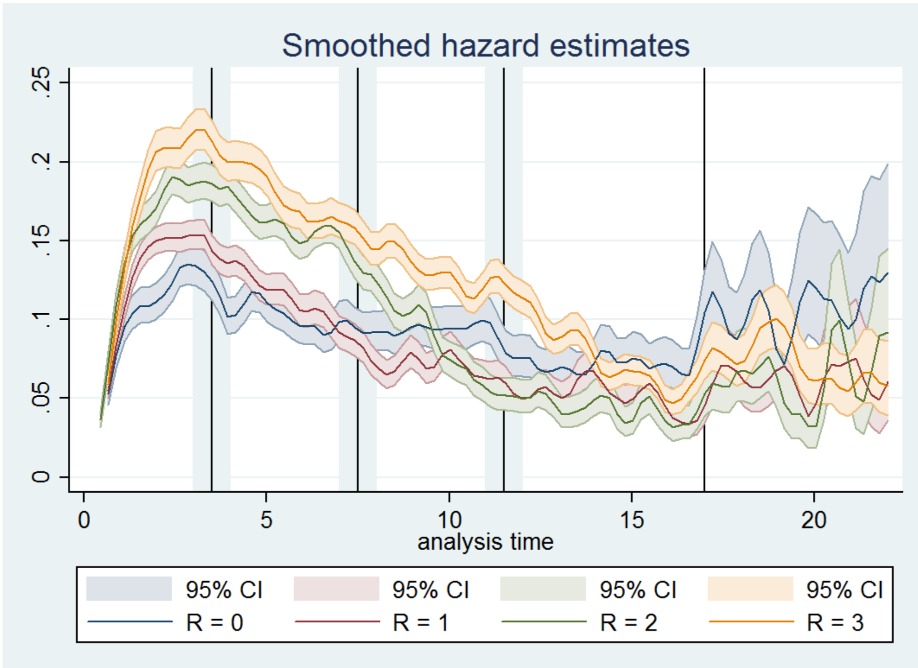

Figure 8. Hazard of self-citation, from date of issuance, by **application date** of citing patent, 1981-2009. R indicates number of lifetime renewals.

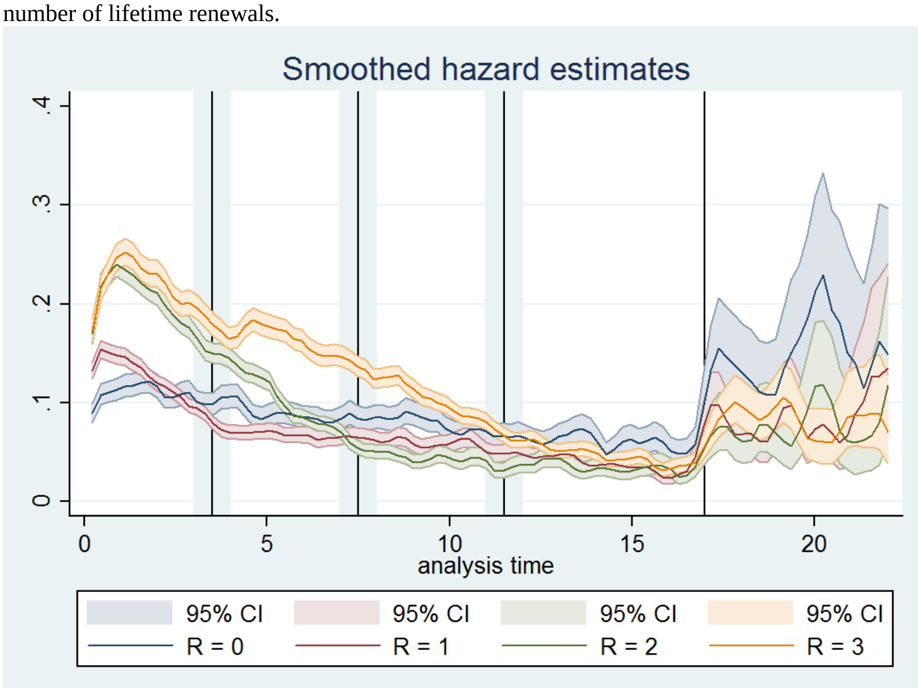

Figure 9. Mean predicted hazard of being cited from parametric estimates.
Fixed effects for application year, calendar year of event, and quarter since issuance.
From date of issuance, by grant date of citing patent. R indicates number of lifetime renewals.

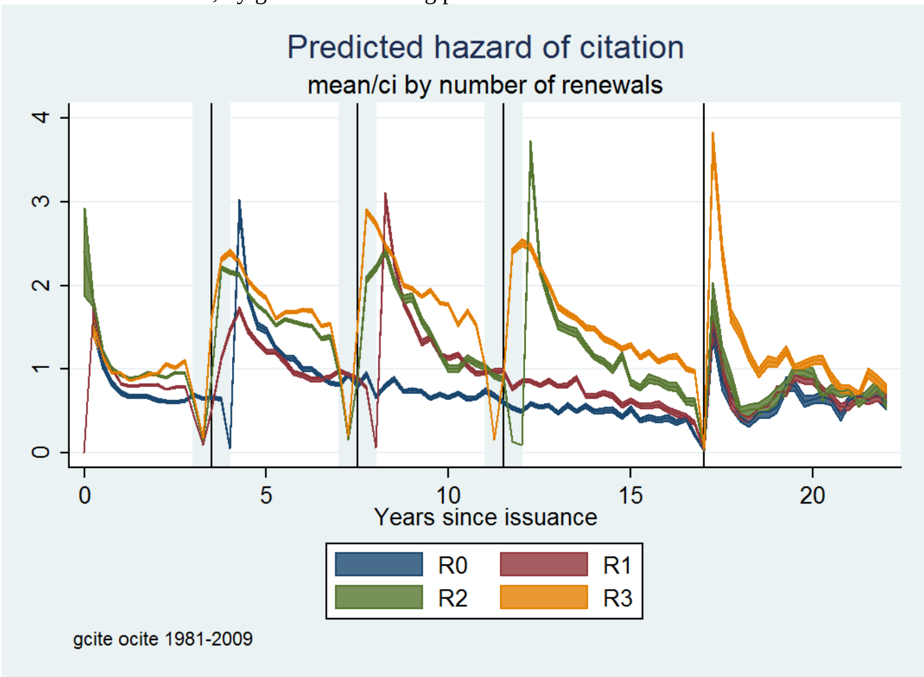

Figure 10. Median predicted hazard of being cited from parametric estimates.
Fixed effects for application year, calendar year of event, and quarter since issuance.
From date of issuance, by grant date of citing patent. R indicates number of lifetime renewals.

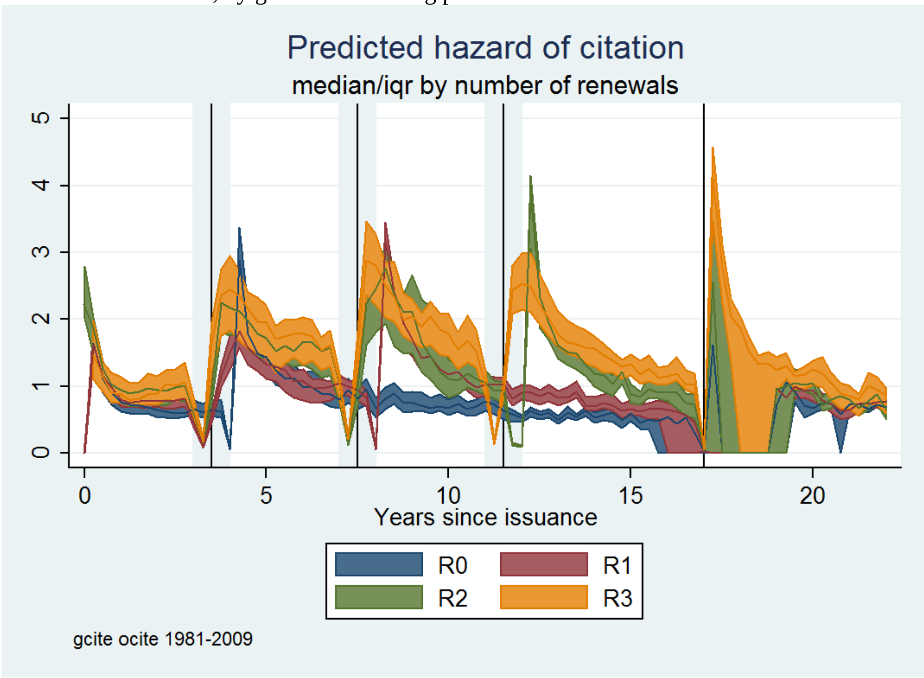

Figure 11. Mean predicted hazard of being cited from parametric estimates.
Fixed effects for application year, calendar year of event, and quarter since issuance.
From date of issuance, by grant date of citing patent. R indicates number of lifetime renewals.

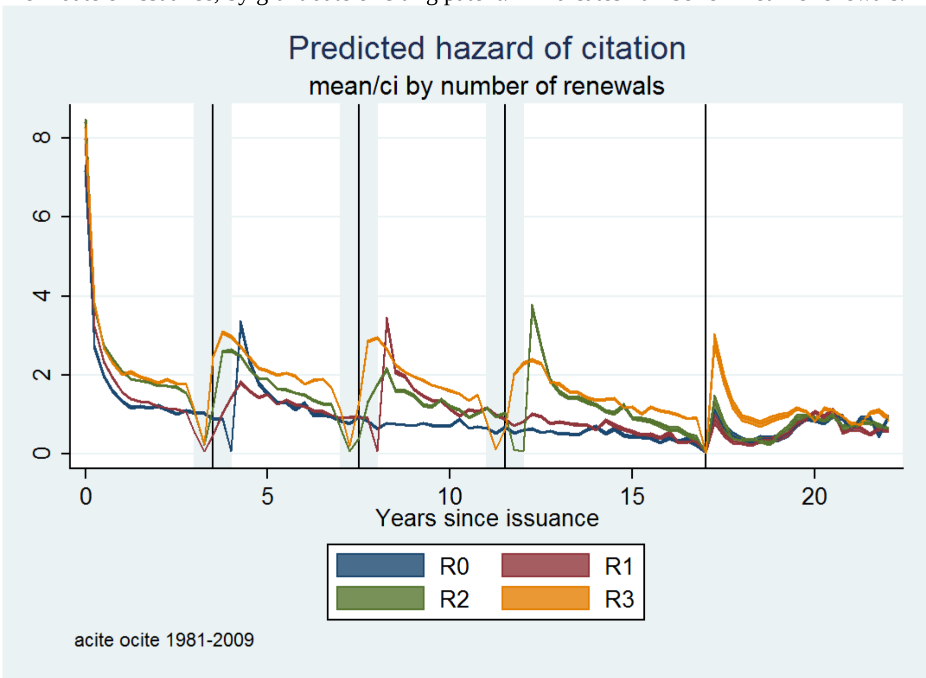

Figure 12. Median predicted hazard of being cited from parametric estimates.
Fixed effects for application year, calendar year of event, and quarter since issuance.
From date of issuance, by grant date of citing patent. R indicates number of lifetime renewals.

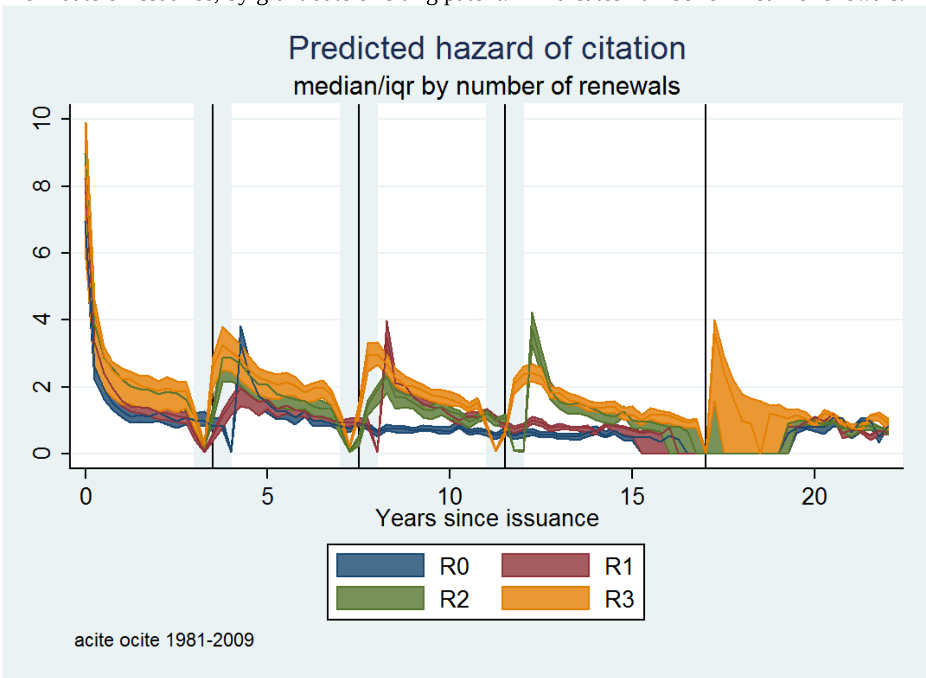

Figure 13. Mean predicted hazard of self-citation from parametric estimates.
Fixed effects for application year, calendar year of event, and quarter since issuance.
From date of issuance, by grant date of citing patent. R indicates number of lifetime renewals.

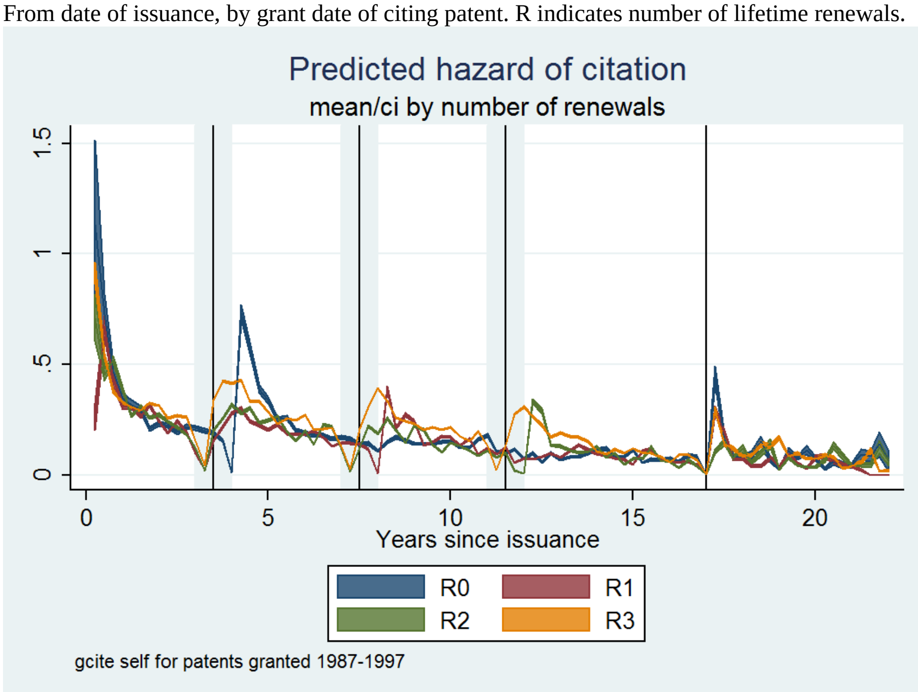

Figure 14. Median predicted hazard of self-citation from parametric estimates.
Fixed effects for application year, calendar year of event, and quarter since issuance.
From date of issuance, by grant date of citing patent. R indicates number of lifetime renewals.

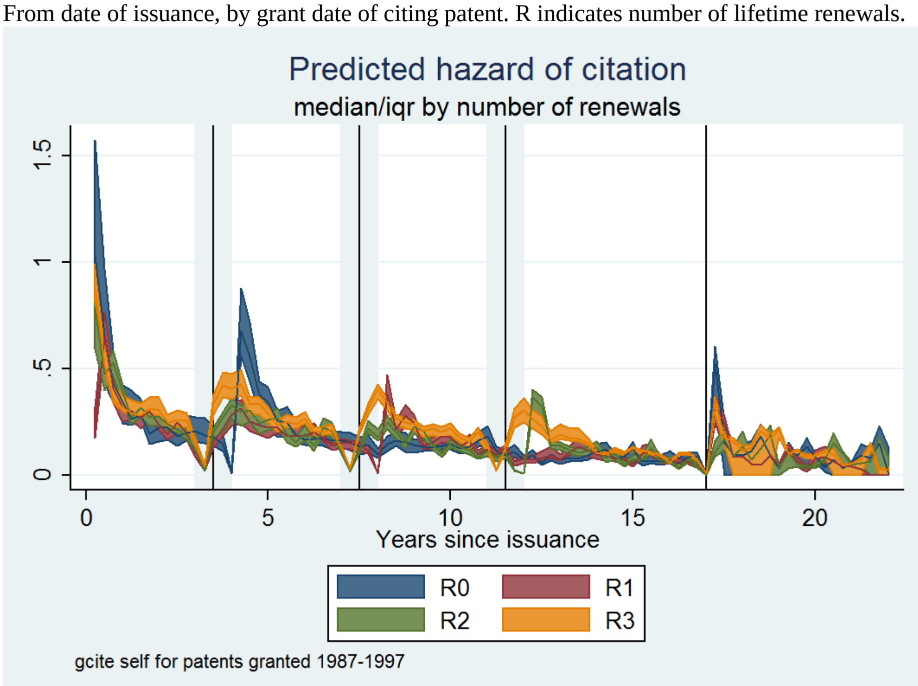

Figure 15. Mean predicted hazard of self-citation from parametric estimates.
Fixed effects for application year, calendar year of event, and quarter since issuance.
From date of issuance, by grant date of citing patent. R indicates number of lifetime renewals.

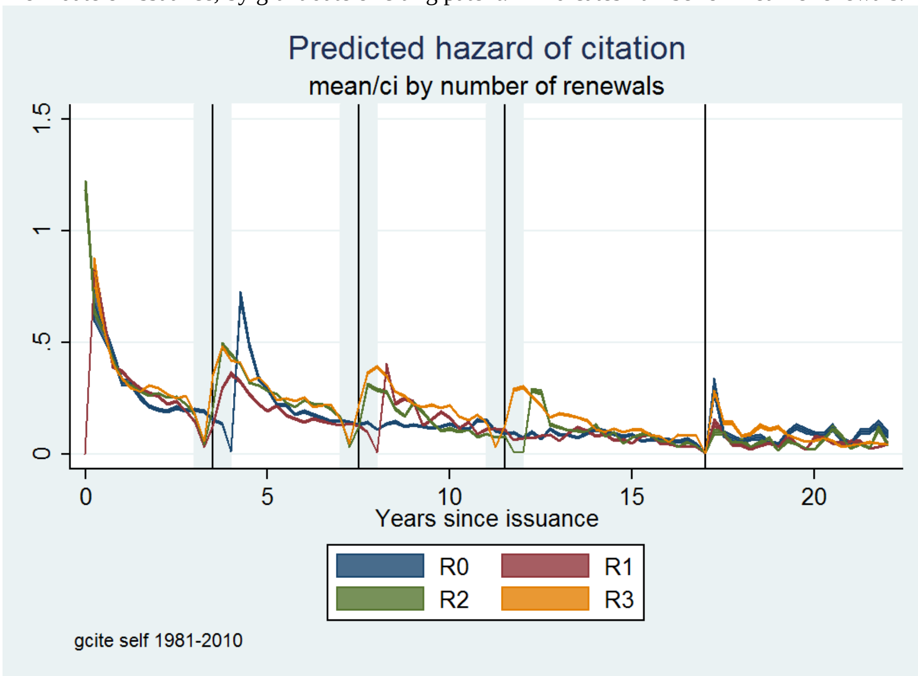

Figure 16. Median predicted hazard of self-citation from parametric estimates.
Fixed effects for application year, calendar year of event, and quarter since issuance.
From date of issuance, by grant date of citing patent. R indicates number of lifetime renewals.

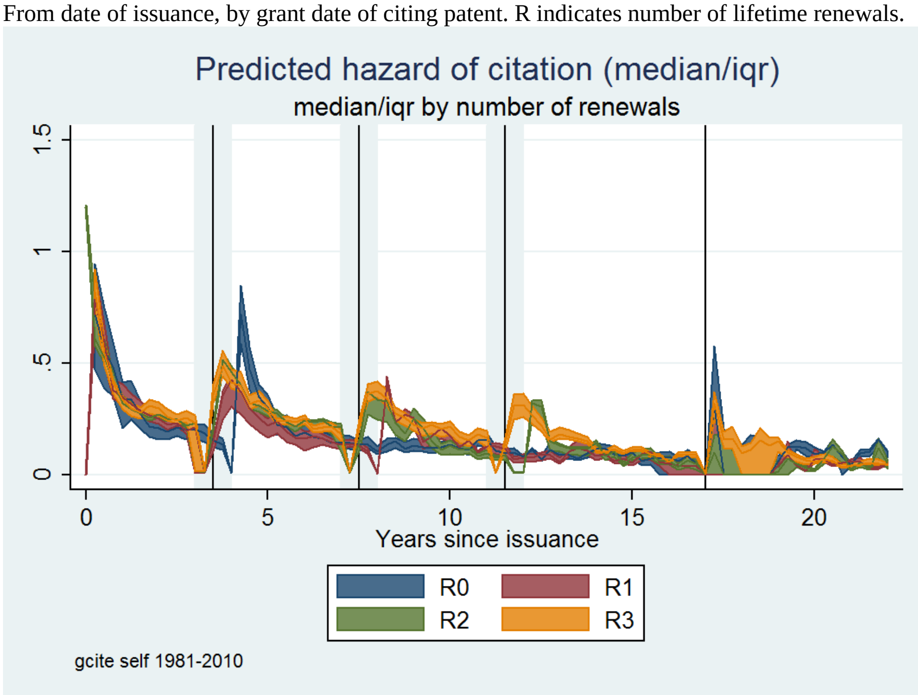

Figure 17. Mean predicted hazard of self-citation from parametric estimates.
Fixed effects for application year, calendar year of event, and quarter since issuance.
From date of issuance, by grant date of citing patent. R indicates number of lifetime renewals.

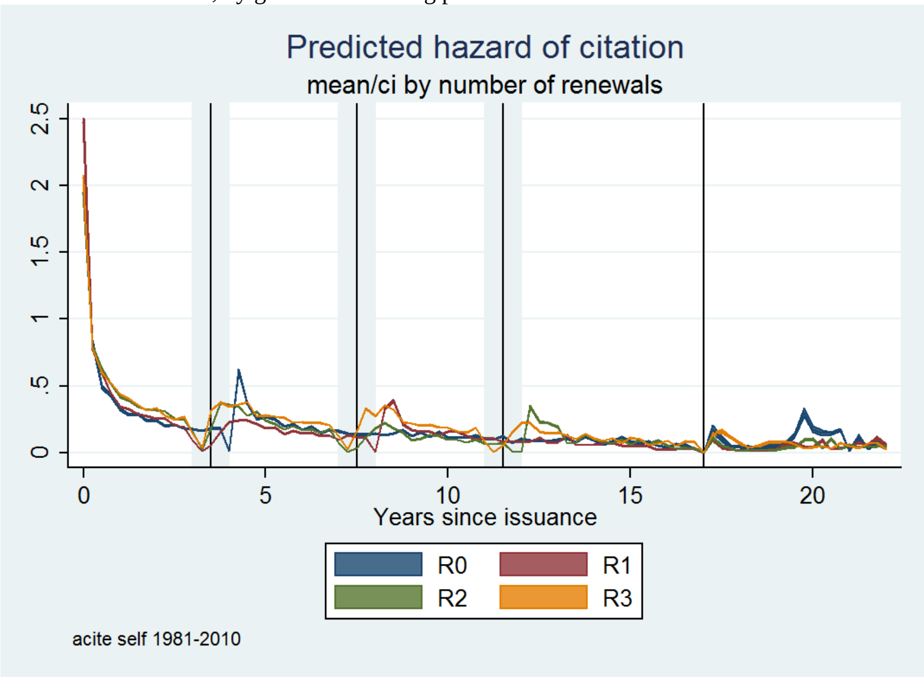

Figure 18. Median predicted hazard of self-citation from parametric estimates.
Fixed effects for application year, calendar year of event, and quarter since issuance.
From date of issuance, by grant date of citing patent. R indicates number of lifetime renewals.