Evaluating The Forward Citations-Patent Value Relationship: The Role Of Competition

Neelesh T. Moorthy

Primary Advisor: Wesley Cohen, Fuqua School of Business
Secondary Advisor: Michelle Connolly, Department of Economics

Department of Economics, Duke University
Durham, North Carolina
April 18, 2018

Honors thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University

Acknowledgements:

I would like to thank my thesis advisors, Dr. Wesley Cohen and Dr. Michelle Connolly, for guiding me through this process. This final product would not have been possible without their advice and revision. I also want to thank the students in my thesis workshop, who suggested ways to improve my thesis and proposed new avenues for research.

Abstract:

I assess whether forward citations—how often patents are cited by subsequent patents—reliably capture patent quality. A high-quality invention might lack forward citations if there are no competing, patenting firms. This introduces measurement error in using citations to measure patent value. I test whether greater competition makes forward citations better measures of patent quality, with eight and twelve-year patent renewal rates serving as my benchmark measures of patent quality. Patent data come from the manufacturing survey in Cohen, Nelson, and Walsh (2000). I conduct logit regressions of patent renewal on forward citations and the number of competitors faced by surveyed manufacturing labs. While the regression results do not support the competition hypothesis, they confirm that forward citations positively predict renewal. They also lend insight into firms' strategic renewal decisions.

JEL Codes: O31, O34

Keywords: Innovation, Patents, Patent Quality, Forward Citations, Renewal, Competition

Introduction:

Economists often use patents and research and development expenditures to estimate innovation in a nation's economy. Research and development expenditures represent how much effort a nation puts into technological advance, with patents ideally capturing the output. The simplest measure of innovation through patents is a count of the number of patents a country issues, but this is a flawed measure. It labels countries with looser patent-issuance procedures "more innovative" by default than those with stricter patent prosecution procedures. Due to these concerns, the most common proxy for patent quality is the number of forward citations a patent receives, or the number of times a given patent is cited in subsequent patent applications. This contrasts with backward citations, which are how many citations a patent itself includes as prior art in the technological field. Researchers use forward citations to measure a patent's importance—if other inventors build off the initial invention, the patent is supposed to be valuable and contribute to scientific knowledge. However, there are remaining concerns with using forward citations as a measure of patent quality.

Some methodological problems with forward citations limit their validity as a proxy for patent quality. Forward citations do not always indicate that a subsequent inventor used the prior patent for inspiration. Citations are frequently added by the patent examiner. Alcacer, Gittelman and Sampat (2009) find that about 63 percent of all citations from 2001-2003 were added by patent examiners. The forward citations-patent value relationship also differs by whether another firm cites the patent (external citations), the firm cites its own patent (self-citations), or the firm indirectly cites its own patent (internal citations[1]). Forward citations also raise the issue of truncation bias if the patent's forward citations occur after the study's time frame. The reliability of forward citations as a measure of patent value differs across industry types. Forward citations might reflect patent value in what Merges and Nelson (1990) call cumulative industries, or industries where inventions build on prior inventions. But they might not reliably capture patent value in discrete industries, where inventions do not build upon each other. Similarly, the forward citations-value relationship changes in what Cohen, Nelson and Walsh (2000) call complex industries, where products consist of multiple patents such that each individual patent is less relevant.

[1] For example, Intel citing a patent from Microsoft when Microsoft's patent was itself based on a prior Intel patent.

My thesis builds on studies differentiating the citation-value relationship across firms and industries. I differentiate firms based on how much competition they face, which has not been done so far in the citations-value literature. I hypothesize that when competition is introduced into the market, forward citations become a better proxy for patent quality. I suggest this since high-value patents generated by monopolistic firms would likely have fewer forward citations if there is little market entry by other firms. This creates a measurement error in using forward citations. I use a 1994 Carnegie Mellon survey of 539 manufacturing labs across 180 industries from Cohen, Nelson and Walsh (2000).[2] The dataset, as supplemented by Cohen and Roach (2013), contains 14,192 patents from those labs, as well as each lab's self-identified number of U.S. competitors that can effectively diminish its profits from innovation. I assess through a logit model whether forward citations are a better proxy for patent renewal—my benchmark measure of patent quality—when the number of competitors is factored into the model.

My thesis evaluates the citation-value relationship, but it also lends insight into the overall validity of patents as a measure of innovation. Cohen, Nelson and Walsh (2000) find that most surveyed manufacturing labs focus on secrecy, not patents, to protect product innovation. They also find that blocking firms from entering the market is a strong reason for patenting, indicating strategic concerns. These results suggest that patents both understate total innovation by ignoring trade secrets and overstate innovation by capturing so-called "blocking patents." Understanding how citations relate to quality informs policymakers as to how much weight patents should be given in innovation policy. For example, Abrams, Akcigit and Popadak (2013) compare patent licensing revenues to forward citations, and find that high-revenue patents have fewer citations than low-revenue patents. My thesis hypothesizes a similar inverse relationship: high-value patents generated by monopolistic firms have fewer citations due to high entry barriers that prevent other firms from competing. If this is the case, the results suggest that non-patent measures of innovation are reliable measures in some markets but not in others. Further research into the narrower citation-value relationship informs the broader question about the role of patents in the modern economy.

[2] As defined by the Standard Industrial Classification (SIC) codes at the four-digit level of aggregation.

The next section provides a literature review on the relationship between patent citations and patent value. Afterwards, I discuss my empirical specifications before explaining my data sources. I continue onto my results and discussion before concluding.

Literature Review:

Several studies suggest that forward citations are good proxies for patent quality. Trajtenberg (1990) is one of the first studies to tackle this question. He compares forward citations of patents in the Computed Tomography scanner—a form of x-ray used in hospitals—field to patent value estimates from a social welfare function. He finds a positive, non-linear relationship, indicating that the usefulness of forward citations as a proxy for quality rises on the margin. Building on this work, Harhoff (1999) uses survey data from inventors in the U.S. and Germany and compares their economic value estimates to forward citations. He too finds a positive relationship across all eight classes of the International Patent Classification System.[3] Looking at firms in the aggregate, Hall, Jaffe and Trajtenberg (2005) compare firms' stock market value to average forward citations. Once again, they find a positive relationship across six sectors—drugs, chemicals, computers, electronics, metals and miscellaneous low-tech industries. Fischer and Leidinger (2013) and Abrams and Sampat (2017) find a positive relationship between revenue and citations. Fischer and Leidinger (2013) use 2006-2009 patent auction sales data from auctions conducted by the firm Ocean Tomo, and Abrams and Sampat (2017) use per-drug sales data from the Optum ClinFormatics database. Taken together, the aforementioned studies demonstrate a positive citation-value relationship using several different estimates of patent value.

However, other studies find either no citation-value relationship or an inverse relationship. Sampat and Ziedonis (2002) use forward citation data for patents developed and licensed by Columbia University and the University of California, but find no relationship between forward citations and licensing revenue. It is possible that their findings stem from the uniqueness of university research, which is less applied than research at commercial firms. But Nikulanien, Hermans and Kulvik (2008) reach the same conclusion when studying the biotechnology sector in Finland. They compare citation data from the European Patent Office with present value estimates for biotechnology firms' revenue flows, and find no relationship between forward citations and revenue. These studies do not necessarily contradict those in the prior paragraph; the lack of

[3] Human Necessities, Transporting, Chemistry, Textiles, Fixed Constructions, Mechanical Engineering, Physics, and Electricity.

relationship could stem from noise in the data. Abrams, Akcigit and Popadak (2013) and Chen and Chang (2010) conclude that higher-value patents receive fewer forward citations than lower-value patents in the technology and pharmaceutical sectors, respectively. The former study quantifies patent value through licensing revenue, while the latter uses pharmaceutical firms' market value. They also arrive at different explanations for the inverse-U. Abrams, Akcigit and Popadak (2013) propose that strategic patents block other firms from patenting, while Chen and Chang (2010) propose that R&D spillovers lead to a decline in market value for firms holding frequently-cited patents. The difference in explanation could stem from the different level of analysis. Spillovers can decrease a firm's market value through competition but not necessarily impact a specific patent's revenue. Taken together, the studies push back on the notion that citations always share a positive relationship with patent quality.

Recent research is starting to carefully note the difference between examiner and applicant citations. The Patent and Trademark Office (PTO) began distinguishing applicant and examiner citations from 2001 onwards, but most of the patent valuation literature does not differentiate between the two citation types. There are exceptions. Hedge and Sampat (2009) define patent value as the likelihood of renewal and find that examiner citations predict value better than applicant citations. Alcacer, Gittelman and Sampat (2009) find that, between 2001-2003, examiners inserted 63 percent of all citations. Given the large number of examiner citations and differences between examiners and applicants in citing behavior, forward citation studies need to take this distinction into account.

Differentiating between type of industry is also important to understanding the citation-value relationship. Merges and Nelson (1990) propose a typology of innovation: discrete versus cumulative. Cumulative inventions are built upon by future inventions, whereas discrete inventions are not. Patent monopolies in industries with more cumulative innovations are more likely to stunt future innovation, they argue. Belenzon (2012) argues that internal citations—when a company indirectly cites one of its own prior patents—represents cumulative invention via technical advance.[4] He finds a stronger positive citation-stock market value relationship for this type of citation. Complex innovation occurs when products are made up of several patents, making each individual patent less relevant (Cohen, Nelson, and Walsh 2000). Hall, Jaffe and Trajtenberg

[4] For example, Intel citing a patent from Microsoft when Microsoft's patent was itself based on a prior Intel patent.

(2005) find that the citation-stock market value relationship is weaker in complex-innovation industries. Overall, studies show that expecting a positive citation-value relationship does not make sense for all industries.

Empirical Framework:

Specification (1) represents the fullest version of my regression, with *i* designating the patent, *j* designating the lab, and *t* designating the industry.[5] The dependent variable represents the binary variable renewed after eight or twelve years, which is my benchmark for patent value.[6] *ForwardCite* represents forward citations, *Industry* represents the four-digit SIC code for the industry to which the patenting firm belongs, and *Year* represents the patent's grant year. *USCompete* is a measure of technological competition that captures the number of U.S. competitors a firm faces in its invention space. *HHI* measures industry concentration for the industry to which the patenting firm belongs for the year when the patent was granted, increasing to one as competition decreases. These variables capture different forms of competition, with the former capturing the number of competitor firms innovating in a narrow invention space and HHI capturing market power in a broader market.[7] *Employees* proxies for the size of the firm via the number of employees, *BackwardCite* represents how much prior art a patent cited, and *PatentEffectiveness* reflects the firm's historical success in protecting its innovations with patents. *MeasurePerformance, LicensingRevenue, PreventLitigation, CrossLicensing, PreventCopying, PreventOtherPatents,* and *EnhanceReputation* all represent labs' reasons for patenting to assess whether firms patenting for different reasons perceive renewal value differently. *InventionOverlap* measures what percentage of competitor firms' innovations overlap with the surveyed lab's innovations and controls for whether firms facing overlap place more weight on renewing patents. *GraduateDegree* controls for researchers' human capital as the percentage of research and development workers with doctorates.

[5] This section briefly defines my variables. For a further discussion of how they were derived from the survey, see the data section below. A table of data definitions is also included in Appendix I.

[6] There is not enough variation in four-year renewal rates to consider empirically.

[7] The correlation between the two is slightly negative, -.0503.

$$(1)\ Logit(Renewal)_{ijt} = \beta_1 ForwardCite_i + \beta_2 Industry_{jt} + \beta_3 Year_i + \beta_4 USCompete_j + \beta_5 HHI_{jt} + \beta_6 Employees_j + \beta_7 BackwardCite_i + \beta_8 PatentEffective_j + \beta_9 MeasurePerformance_j + \beta_{10} LicensingRevenue_j + \beta_{11} PreventLitigation_j + \beta_{12} CrossLicensing_j + \beta_{13} PreventCopying_j + \beta_{14} PreventOtherPatents_j + \beta_{15} EnhanceReputation_j + \beta_{16} InventionOverlap_j + \beta_{17} GraduateDegree_j + u_{ijt}$$

In order to detect the measurement error posited by my hypothesis, I run several iterations of this regression, which are shown in Tables 2 and 3 of my results section. In Model 1, I run the regression with only forward citations, backward citations, patent effectiveness, employees, and industry and year dummies as my dependent variables. I subsequently run that regression with the competitors variable for Model 2. If the reliability of forward citations as a value proxy is understated due to measurement error, the coefficient on forward citations should increase from Model 1 to Model 2. In Model 3, I again omit competitors but add the HHI variable, and, in Model 4, I reinsert competitors. If my hypothesis is correct, the forward citations coefficient should increase from Model 3 to Model 4. In Model 5, I omit competitors but add the rest of the dependent variables, and, in Model 6, I reintroduce competitors. If my hypothesis is correct, the forward citations coefficient should increase from Model 5 to Model 6.

Data:

My survey data come from the Cohen, Nelson and Walsh (2000) 1994 Carnegie Mellon survey. The dataset provides the data for lab competitors through each lab's response to question 34, "How many firms are able to introduce competing innovations in time to effectively diminish your firm's profits from your innovations?" The competitors variable takes the form of a dummy, with dummies for the ranges of 0, 1-2, 3-5, 6-10, 11-20, and more than 20 competitors. The 0-competitor dummy is dropped during my regressions. The data for patent effectiveness come from each lab's answer to question 32b, "During the last three years, for what percent of your product innovations was [patent protection] effective in protecting your firm's competitive advantage from those innovations?" Labs responded with a number from one to five, one representing "below 10%" and five "over 90%." The number of employees variable comes from survey question 62, "For your focus industry, please estimate how many employees your firm has working in that industry in the U.S." *MeasurePerformance, LicensingRevenue, PreventLitigation, CrossLicensing, PreventCopying, PreventOtherPatents,* and *EnhanceReputation* are all dummy variables with a value of one if the firm said it was a reason for patenting in response to question 38, "For your

R&D unit's most recent application for a product or process patent, which of the following reasons motivated your decision to apply for a patent?"[8] *InventionOverlap* is the answer to question 28, "What percent of the projects started by your R&D unit in the last three years have the same technical goals as an R&D project conducted by at least one of your competitors?" It is a categorical variable from one to five, with one representing zero percent and five 76-100 percent. *GraduateDegree* answers question 54, "of the [number of professional and technical R&D employees] how many are Ph.D. or M.D. scientists?"

Roach and Cohen (2013) later supplemented the data by matching labs to their respective patents. I separately match the patent numbers to forward citations collected by the National Bureau of Research's Patent Data File until 2006. NBER also provides the issuance year dummy variable. I match the patents to backward citation data compiled in an online database by Sampat (2011). The initial Roach and Cohen (2013) data include 21,659 patents for 1,476 different manufacturing labs across 313 industries as defined by the four-digit SIC code, but after removing labs that did not answer the question about competitors, my data include 14,192 patents for 539 different manufacturing labs across 180 industries. I create dummies for eight and twelve-year renewal by matching the patents to the Reed Tech Patent Maintenance Events database. 73.34 percent of the patents were renewed after eight years, and 63.54 percent after twelve years. The HHI indices are computed by Keil (2017) and represent the squared market share of all firms in the industry, normalized between zero and one. She estimates the HHI using an equation that weighs the four, eight, and twenty-firm concentration ratios in the industries.[9] The concentration ratios inputted into Keil's equation are generated every five years by the Census Bureau.

To ensure that any relationship between forward citations and renewal is not being driven by backward citations, I check correlation between the two citation types. The correlation is positive but small (.0804). I also check correlations between competitor and industry dummies. In 54 percent of industries, all labs reported the same number of competitors. In other industries,

[8] *MeasurePerformance* is "to measure the performance of your researchers," *LicensingRevenue* is "to obtain revenue through licensing the invention," *CrossLicensing* is "to improve your position in negotiations with other firms, for example, in cross-licensing agreements," *PreventCopying* is "to prevent other firms from copying your invention," *PreventLitigation* is "to prevent patent infringement suits against your firm," *PreventOtherPatents* is "to prevent other firms from patenting a related invention," and *EnhanceReputation* is "to enhance the reputation of the firm or its R&D employees."

[9] A concentration ratio measures what percentage of industry sales is controlled by a certain number of firms in the market.

firms report different numbers of competitors. For example, in the "Pharmaceutical Preparations[10]" industry (SIC 2834), one lab reports zero competitors, two report 1-2 competitors, seven report 3-5 competitors, and eight report 6-10 competitors. The different number of competitors likely stems from the nature of the question—the survey asks not how many competitors the lab faces, but rather how many can introduce competing innovations and diminish the firm's profit. If labs in the same industry assess their rivals differently or work in different subfields, their answers will vary. This variable reflects true differences in firms' innovation potential, although there is some subjectivity in firms' survey responses. Similarly, in 53 percent of industries, all labs have the same recorded patent effectiveness, whereas in other industries labs differed. In the Pharmaceutical Preparations industry, one firm answered "1," three answered "3," five answered "4," and nine answered "5." Table 1 provides a correlation matrix between all variables in my dataset.

Results and Discussion:

Table 2 provides the results for the Renewed8 regressions and Table 3 provides the results for the Renewed12 regressions. For both eight-year and twelve-year renewals, the coefficient on forward citations never changes in a statistically significant manner when competitors are introduced. This suggests that competition-based measurement error does not impair forward citations' utility as a proxy for value. To ensure that collinearity between the competition variables and industry dummies is not skewing results toward statistical insignificance, I conduct correlation checks of my regression coefficients. The highest correlation between any industry coefficient and a competitors coefficient is .32. Most correlations are less than .1. The highest correlation between the HHI coefficient and any industry coefficient is .29. Most correlations are in the .1 range.

[10] This industry includes finished drugs such as tablets, capsules, powders and ointments.

Table 1: Correlation Matrix

	ForwardCitations	BackwardCitations	USCompetitorsQ34A	PatentEffectiveness	EmployeesQ62	HHI	MeasurePerformance	LicensingRevenue	CrossLicensing	PreventLitigation	PreventCopying	PreventOtherPatents	EnhanceReputation	InventionOverlap	GraduateDegree
ForwardCitations	1.0000														
BackwardCitations	0.0540	1.0000													
USCompetitorsQ34A	-0.0946	-0.0550	1.0000												
PatentEffectiveness	0.0771	-0.0411	-0.0734	1.0000											
EmployeesQ62	0.0067	0.1132	0.0616	-0.3374	1.0000										
HHI	-0.0629	-0.0845	0.0794	-0.0433	0.7625	1.0000									
MeasurePerformance	-0.0271	-0.0171	0.0285	0.0869	-0.0606	-0.0748	1.0000								
LicensingRevenue	0.0724	-0.1218	0.0461	0.0494	0.0986	-0.0764	0.0983	1.0000							
CrossLicensing	0.0909	-0.1883	-0.1283	-0.0154	0.3532	0.1990	0.0446	0.0675	1.0000						
PreventLitigation	0.0589	-0.0014	0.0292	-0.0102	0.2253	0.0615	0.0362	0.1589	0.4387	1.0000					
PreventCopying	0.0276	0.0175	-0.0282	0.1696	0.0838	0.0095	0.0123	0.0487	0.0943	0.2581	1.0000				
PreventOtherPatents	0.1203	0.0206	-0.1846	0.2309	-0.3994	-0.5605	-0.1292	0.1878	0.1131	0.0927	0.1549	1.0000			
EnhanceReputation	-0.0262	-0.0571	0.0372	-0.0389	0.0878	0.2204	0.1109	-0.1276	0.2366	0.1593	0.0481	-0.3612	1.0000		
InventionOverlap	0.0935	0.0280	-0.2196	-0.0771	0.1931	0.0496	-0.0178	0.1568	0.1163	-0.0086	-0.0137	0.0302	-0.0496	1.0000	
GraduateDegree	0.0076	-0.0400	0.1553	0.2103	-0.3729	-0.2807	0.0067	0.3073	-0.0893	-0.0544	0.0442	0.0633	0.1619	0.0935	1.0000

Table 2: Renewed8 Regressions

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4	(5) Model 5	(6) Model 6
ForwardCitations	0.0241***	0.0241***	0.0248***	0.0246***	0.0243***	0.0242***
	(0.00276)	(0.00277)	(0.00316)	(0.00318)	(0.00320)	(0.00322)
BackwardCitations	0.00419*	0.00372*	0.00308	0.00196	0.00253	0.00142
	(0.00217)	(0.00217)	(0.00251)	(0.00250)	(0.00260)	(0.00261)
PatentEffectiveness	0.0470	0.0470	0.0779*	0.0846*	0.141***	0.119**
	(0.0373)	(0.0370)	(0.0440)	(0.0447)	(0.0507)	(0.0534)
EmployeesQ62	1.80e-06*	1.47e-06	1.79e-06	1.11e-06	1.92e-06*	1.15e-06
	(1.05e-06)	(1.04e-06)	(1.15e-06)	(1.12e-06)	(1.05e-06)	(1.05e-06)
1-2 USCompetitorsQ34A		-0.0493		-0.344		-0.167
		(0.196)		(0.282)		(0.305)
3-5 USCompetitorsQ34A		-0.231		-0.552**		-0.315
		(0.199)		(0.268)		(0.310)
6-10 USCompetitorsQ34A		-0.443**		-0.916***		-0.862***
		(0.220)		(0.299)		(0.317)
11-20 USCompetitorsQ34A		1.599*		1.078		1.280
		(0.837)		(0.886)		(0.906)
>20 USCompetitorsQ34A		-0.147		0.271		0.195
		(0.672)		(0.834)		(0.819)
HHI			9.231	8.993	8.111	7.988
			(7.209)	(7.270)	(7.079)	(7.115)
MeasurePerformance					-0.0817	-0.0321
					(0.298)	(0.328)
LicensingRevenue					-0.0363	0.0555
					(0.130)	(0.151)
CrossLicensing					-0.288**	-0.352***
					(0.124)	(0.127)
PreventLitigation					0.117	0.358*
					(0.173)	(0.188)
PreventCopying					0.310	0.209
					(0.333)	(0.334)
PreventOtherPatents					0.389***	0.325**
					(0.149)	(0.152)
EnhanceReputation					-0.281**	-0.371***
					(0.121)	(0.127)
InventionOverlap					0.216***	0.186***
					(0.0680)	(0.0705)
GraduateDegree					1.049**	1.473***
					(0.508)	(0.532)
Constant	-1.656	-1.373	-0.346	0.271	-1.143	-0.727
	(1.104)	(1.122)	(1.267)	(1.307)	(1.333)	(1.395)
Industry Dummies	Y	Y	Y	Y	Y	Y
Year Granted Dummies	Y	Y	Y	Y	Y	Y
Pseudo R^2	.1115	.1134	.1187	.1219	.1247	.1275
Observations	10,767	10,767	7,965	7,965	7,490	7,490

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Renewed12 Regressions

VARIABLES	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4	(5) Model 5	(6) Model 6
ForwardCitations	0.0265***	0.0265***	0.0270***	0.0269***	0.0262***	0.0261***
	(0.00245)	(0.00245)	(0.00274)	(0.00274)	(0.00274)	(0.00274)
BackwardCitations	0.00561***	0.00508**	0.00566**	0.00455*	0.00634**	0.00527**
	(0.00203)	(0.00204)	(0.00238)	(0.00238)	(0.00254)	(0.00255)
PatentEffectiveness	0.0511	0.0457	0.0701*	0.0638	0.185***	0.159***
	(0.0337)	(0.0334)	(0.0392)	(0.0394)	(0.0461)	(0.0479)
EmployeesQ62	9.46e-07	6.46e-07	9.78e-07	4.38e-07	1.56e-06*	9.12e-07
	(9.77e-07)	(9.32e-07)	(1.03e-06)	(9.72e-07)	(8.84e-07)	(8.93e-07)
1-2 USCompetitorsQ34A		0.128		-0.0581		-0.0110
		(0.184)		(0.254)		(0.294)
3-5 USCompetitorsQ34A		-0.155		-0.357		-0.198
		(0.187)		(0.245)		(0.302)
6-10 USCompetitorsQ34A		-0.229		-0.539**		-0.600**
		(0.205)		(0.272)		(0.304)
11-20 USCompetitorsQ34A		1.366*		0.945		1.332
		(0.747)		(0.796)		(0.834)
>20 USCompetitorsQ34A		-0.263		0.309		-0.0597
		(0.683)		(0.786)		(0.775)
HHI			6.427	6.334	5.574	5.511
			(5.224)	(5.229)	(5.384)	(5.376)
MeasurePerformance					1.120***	1.151***
					(0.295)	(0.328)
LicensingRevenue					0.0707	0.138
					(0.121)	(0.139)
CrossLicensing					-0.280**	-0.329***
					(0.112)	(0.116)
PreventLitigation					-0.0878	0.116
					(0.158)	(0.170)
PreventCopying					0.150	0.0566
					(0.315)	(0.310)
PreventOtherPatents					0.655***	0.604***
					(0.142)	(0.146)
EnhanceReputation					-0.579***	-0.643***
					(0.114)	(0.119)
InventionOverlap					0.299***	0.267***
					(0.0611)	(0.0641)
GraduateDegree					1.400***	1.737***
					(0.459)	(0.482)
Constant	-0.590	-0.682	-0.843	-0.426	-1.752	-1.438
	(0.480)	(0.513)	(1.314)	(1.339)	(1.394)	(1.439)
Industry Dummies	Y	Y	Y	Y	Y	Y
Year Granted Dummies	Y	Y	Y	Y	Y	Y
Pseudo R^2	.1261	.1280	.1422	.1445	.1592	.1612
Observations	10,789	10,789	7,997	7,997	7,522	7,522

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

There are several possible reasons why these data do not support the hypothesis. The hypothesis assumes firms with few competitors would have few forward citations despite patenting valuable inventions. But this might not be the case if a firm self-cites its own patent, which the NBER forward citation data do not capture. Firms might build off their prior inventions and cite their former patents to distinguish their newer inventions. If there are self-citations in my dataset, then the proposed measurement error would be harder to observe. This is a plausible reason for my findings; on average, each firm in my dataset holds 24 patents. Examiner citations could also affect my results. If examiners include a high percentage of citations, forward citations may not suffer from measurement error. As previously mentioned, Alcacer, Gittelman and Sampat (2009) find that about 63 percent of all citations from 2001-2003 were added by patent examiners. If patent examiners thoroughly search for prior art, then a firm with few competitors could still have highly-cited patents because examiners, rather than competitors, are citing patents.

There are also limitations in the competitors variable that could contribute to the results not supporting this hypothesis. Although the competitors variable measures how many competitors can diminish a firm's profits from innovation, rather than the number of other firms in the industry, it is still a lab-specific rather than patent-specific variable. The survey does not ask for every single patent how many firms could diminish profits from that patent. This could prevent my models from picking up measurement error. Alternative specifications—using interaction terms—are included in Appendix II, which also fail to pick up any statistically significant measurement error that would confirm my hypothesis.

But the results do contribute other insights to the literature. First, they confirm a positive relationship between forward citations and patent renewal. Second, the results suggest strategic rationales behind firms' patent renewal decisions. Firms patenting to cross-license their patents or to enhance their reputation are less likely to renew. The opposite is true for firms looking to prevent other firms from patenting. This suggests that "blocking patents"—patents designed solely to deter other firms—might drive renewal more so than patents designed to derive revenue. However, the data do not allow classification of individual patents as blocking or non-blocking patents, so further research is necessary to test this explanation. Lastly, firms with a higher percentage of R&D researchers holding Ph.Ds. and those facing greater invention overlap with other firms are more likely to patent. The former finding suggests human capital leads to more patent renewal, and the latter that firms renew when rival firms are stronger.

Having established that forward citations are positively related to patent quality, I run a third regression of forward citations on the remaining dependent variables. This regression is designed to observe how firm characteristics affect patent value without the strategic considerations affecting renewal. The results are depicted in Table 4. Most firm-level factors are statistically insignificant, but counterintuitively firms aiming to prevent other firms' patents have higher-cited patents. Assuming firms aiming to prevent other firms from patenting apply for more blocking patents, this finding could be an artifact of how examiners and patenting firms cite blocking patents. A patenting firm might cite a blocking patent not due to its value but because they want to explain how their invention is different. Firms patenting to enhance their reputation also have more forward citations. This makes sense—a firm's reputation should only improve if the patent is high-quality. Lastly, having a greater degree of invention overlap decreases forward citations. If there are two similar patents on the market, it is plausible that a subsequent patenting firm or examiner would only cite one of the two.

Conclusion and Future Directions:

This thesis aimed to see if forward citations suffer from measurement error when a firm's number of competitors is not considered. However, regressions using patent data from the Carnegie Mellon survey are unable to confirm this hypothesis. Future studies should better differentiate self-citations from external citations, and examiner from applicant citations. Understanding biases in forward citations as a measure of patent value is important for policymakers who have to decide whether patents issued today continue to be of high-quality or are artifacts of lenient procedures in the PTO.

The results do confirm that forward citations positively predict patent renewal. They also suggest firms have strategic motivations—such as preventing other firms from patenting—behind patent renewals. Firms aiming to block other firms from patenting surprisingly have more forward citations on their patents themselves. This is plausible if other firms cite the blocking patents not due to their value but because they want to distinguish their patents from the blocking inventions. If this is the case, then forward citations might overstate the patent's scientific value. This thesis does not classify each individual patent as a blocking or non-blocking patent, so further research should explore whether this explanation is correct.

Table 4: Forward Citations Regression

VARIABLES	(1) Model 1
BackwardCitations	0.168***
	(0.0227)
PatentEffectiveness	0.551
	(0.342)
1-2 USCompetitorsQ34A	-0.475
	(2.119)
3-5 USCompetitorsQ34A	-3.307
	(2.236)
6-10 USCompetitorsQ34A	-2.420
	(2.271)
11-20 USCompetitorsQ34A	-1.573
	(3.535)
>20 USCompetitorsQ34A	3.937
	(4.527)
EmployeesQ62	-4.09e-05***
	(1.41e-05)
HHI	-21.44
	(39.05)
MeasurePerformance	-3.421
	(2.277)
LicensingRevenue	1.162
	(1.175)
CrossLicensing	0.924
	(0.934)
PreventLitigation	1.309
	(1.054)
PreventCopying	0.282
	(1.907)
PreventOtherPatents	3.731***
	(1.229)
EnhanceReputation	3.210**
	(1.457)
InventionOverlap	-1.808***
	(0.644)
GraduateDegree	0.634
	(6.056)
Constant	5.329
	(12.26)
Observations	7,556
R-squared	0.102

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Reference List:

Abrams, D., Akcigit, U., & Popadak, J. (2013). Patent Value and Citations: Creative Destruction or Strategic Disruption? NBER Working Paper.

Abrams, D. S., & Sampat, B. N. What's the Value of Patent Citations? Evidence from Pharmaceuticals.

Alcacer, J., & Gittelman, M. (2004). How do I know what you know? Patent examiners and the generation of patent citations.

Alcacer, J., Gittelman, M., & Sampat, B. (2009). Applicant and examiner citations in US patents: An overview and analysis. Research Policy, 38(2), 415-427.

Belenzon, S. (2012). Cumulative innovation and market value: evidence from patent citations. The Economic Journal, 122(559), 265-285.

Chen, Y. S., & Chang, K. C. (2010). Exploring the nonlinear effects of patent citations, patent share and relative patent position on market value in the US pharmaceutical industry. Technology Analysis & Strategic Management, 22(2), 153-169.

Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2000). Protecting their intellectual assets: Appropriability conditions and why US manufacturing firms patent (or not) (No. w7552). National Bureau of Economic Research.

Fischer, T., & Leidinger, J. (2014). Testing patent value indicators on directly observed patent value—An empirical analysis of Ocean Tomo patents auctions. Research Policy, 43(3), 519-529.

Hall, B., Jaffe, A., & Trajtenberg, M. (2005). Market Value and Patent Citations. RAND Journal of Economics,36(1), 16-38.

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

Hegde, D., & Sampat, B. (2009). Examiner citations, applicant citations, and the private value of patents. Economics Letters, 105(3), 287-289.

Keil, Jan, The Trouble with Approximating Industry Concentration from Compustat (May 19, 2017). Journal of Corporate Finance, Vol. 45, No. August, 2017

Lampe, R. (2012). Strategic citation. Review of Economics and Statistics, 94(1), 320-333.

Leslie, C. R. (2006). The Anticompetitive Effects of Unenforced Invalid Patents. Minn. L. Rev., 91, 101

Merges, R. P., & Nelson, R. R. (1990). On the complex economics of patent scope. *Columbia Law Review*, *90*(4), 839-916.

Nikulainen, T., Hermans, R., & Kulvik, M. (2008). Patent citations indicating present value of the biotechnology business. International Journal of Innovation and Technology Management, 5(03), 279-301.

Roach, M., & Cohen, W. M. (2013). Lens or prism? Patent citations as a measure of knowledge flows from public research. *Management Science*, *59*(2), 504-525.

Sampat, B. N., & Ziedonis, A. (2002). Cite seeing: Patent citations and the economic value of patents. Unpublished manuscript, from the author.

Sampat, B. N., & Ziedonis, A. A. (2005). Patent Citations and the Economic Value of Patents. Handbook of Quantitative Science and Technology Research,277-298.

Sampat, 2011, "USPTO Patent and Citation Data, hdl: 1902.1/16412, Harvard Dataverse, V4, UNF:5:ERqP27enbwBRimghqDD4gQ==

Trajtenberg, M. (1990). A Penny for Your Quotes: Patent Citations and the Value of Innovations. The RAND Journal of Economics,21(1), 172.

Appendix I:

Variable	**Definition**
ForwardCite	The number of times a patent is cited by other patents.
Industry	The four-digit SIC code for the industry to which the patenting firm belongs.
Year	The year a patent was granted.
USCompete	How many firms are able to introduce competing innovations in time to effectively diminish your firm's profits from your innovations?
HHI	An annual measure of concentration for the industry in which the patenting firm belongs. The year is the year the patent was granted.
Employees	For your focus industry, please estimate how many employees your firm has working in that industry in the U.S.
BackwardCite	How many patents a patent in the data cites as prior art.
PatentEffective	During the last three years, for what percent of your product innovations was [patent protection] effective in protecting your firm's competitive advantage from those innovations?
MeasurePerformance	For your R&D unit's most recent application for a product or process patent, which of the following reasons motivated your decision to apply for a patent—to measure the performance of your researchers.
LicensingRevenue	For your R&D unit's most recent application for a product or process patent, which of the following reasons motivated your decision to apply for a patent—to obtain revenue through licensing the invention.
CrossLicensing	For your R&D unit's most recent application for a product or process patent, which of the following reasons motivated your decision to apply for a patent—to improve your position in negotiations with other firms, for example, in cross-licensing agreements.
PreventCopying	For your R&D unit's most recent application for a product or process patent, which of the following reasons motivated your decision to apply for a patent—to prevent other firms from copying your invention.
PreventLitigation	For your R&D unit's most recent application for a product or process patent, which of the following reasons motivated your decision to apply for a patent—to prevent patent infringement suits against your firm.
PreventOtherPatents	For your R&D unit's most recent application for a product or process patent, which of the following reasons motivated your decision to apply for a patent—to prevent other firms from patenting a related invention.
EnhanceReputation	For your R&D unit's most recent application for a product or process patent, which of the following reasons motivated your decision to apply for a patent—to enhance the reputation of the firm or its R&D employees.
InventionOverlap	What percent of the projects started by your R&D unit in the last three years have the same technical goals as an R&D project conducted by at least one of your competitors?
GraduateDegree	Of the [number of professional and technical R&D employees], how many are Ph.D. or M.D. scientists?"

Appendix II:

Specification (2) uses an interaction-term approach to test my hypothesis. It is similar to Specification (1), except that it converts competitors into a continuous variable using the midpoint of each dummy variable's range, and then includes interaction terms between competitors and citations and HHI and citations. If my hypothesis is correct, B_5 should be positive and B_7 negative, indicating that forward citations better capture renewal rates when competition increases. In a third specification, I divide forward citations by the lab's number of competitors to generate a scaled citations variable. This means dropping the competitors-citations interaction term, because by definition that would equal forward citations.

$$(2)\ Logit(Renewal)_{ijt} = \beta_1 ForwardCite_i + \beta_2 Industry_{jt} + \beta_3 Year_i + \beta_4 USCompete_j + \beta_5 USCompete_j * ForwardCite_i + \beta_6 HHI_{jt} + \beta_7 HHI_{jt} * Cite_i + \beta_8 Employees_j + \beta_9 BackwardCite_i + \beta_{10} PatentEffective_j + \beta_{11} MeasurePerformance_j + \beta_{12} LicensingRevenue_j + \beta_{13} PreventLitigation_j + \beta_{14} CrossLicensng_j + \beta_{15} PreventCopying_j + \beta_{16} PreventOtherPatents_j + \beta_{17} EnhanceReputation_j + \beta_{18} InventionOverlap_j + \beta_{19} GraduateDegree_j + u_{ijt}$$

The results are shown in Table 5 on the next page. Models 1 and 3 use unscaled citations, and Models 2 and 4 use scaled citations. The relevant interaction terms are not statistically significant in any of the four models, which is consistent with the non-results described in the main body of my paper.

Table 5: Testing Hypothesis with Interaction Terms

VARIABLES	(1) Model 1 (Renewed8)	(2) Model 2 (Renewed8)	(3) Model 3 (Renewed12)	(4) Model 4 (Renewed12)
ForwardCitations	0.0252***		0.0364***	
	(0.00852)		(0.00738)	
BackwardCitations	0.00180	0.00358	0.00557**	0.00740***
	(0.00261)	(0.00279)	(0.00256)	(0.00276)
PatentEffectiveness	0.124**	0.126**	0.169***	0.162***
	(0.0509)	(0.0520)	(0.0465)	(0.0465)
EmployeesQ62	1.52e-06	1.42e-06	1.22e-06	9.56e-07
	(1.05e-06)	(1.00e-06)	(8.91e-07)	(8.67e-07)
USCompetitorsQ34A	-0.0605**	-0.00423	-0.0566***	0.000514
	(0.0237)	(0.0204)	(0.0214)	(0.0187)
USCompetitorsQ34ACitations	0.000882		0.000856	
	(0.00112)		(0.00102)	
HHI	8.494	8.565	7.179	6.610
	(7.052)	(6.962)	(5.449)	(5.421)
HHICitations	-0.0414		-0.150	
	(0.0897)		(0.0889)	
MeasurePerformance	-0.0379	0.124	1.169***	1.136***
	(0.300)	(0.321)	(0.298)	(0.320)
LicensingRevenue	-0.00307	-0.123	0.0837	0.00477
	(0.131)	(0.143)	(0.122)	(0.131)
CrossLicensing	-0.324***	-0.398***	-0.315***	-0.401***
	(0.125)	(0.127)	(0.114)	(0.119)
PreventLitigation	0.280	0.167	0.0642	-0.0128
	(0.183)	(0.183)	(0.166)	(0.166)
PreventCopying	0.220	0.464	0.0654	0.355
	(0.336)	(0.325)	(0.314)	(0.316)
PreventOtherPatents	0.317**	0.369**	0.591***	0.676***
	(0.151)	(0.153)	(0.144)	(0.145)
EnhanceReputation	-0.337***	-0.284**	-0.623***	-0.579***
	(0.123)	(0.123)	(0.116)	(0.116)
InventionOverlap	0.197***	0.163**	0.279***	0.252***
	(0.0696)	(0.0705)	(0.0628)	(0.0635)
GraduateDegree	1.293**	1.288**	1.649***	1.585***
	(0.524)	(0.539)	(0.476)	(0.484)
ScaledForwardCitations		0.0438***		0.0472***
		(0.0148)		(0.0130)
HHIScaledCitations		-0.110		-0.116
		(0.199)		(0.167)
Constant	-0.803	-1.010	-1.532	-1.827
	(1.344)	(1.346)	(1.398)	(1.413)
Industry Dummies	Y	Y	Y	Y
Year Granted Dummies	Y	Y	Y	Y
Pseudo R^2	.1259	.1209	.1608	.1530
Observations	7,490	7,075	7,522	7,107

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1