UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THREE ESSAYS IN ECONOMICS OF INNOVATION AND INDUSTRIAL ORGANIZATION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

JI GU
Norman, Oklahoma
2018

THREE ESSAYS IN ECONOMICS OF INNOVATION
AND INDUSTRIAL ORGANIZATION

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF ECONOMICS

BY

Dr. Qihong Liu, Chair

Dr. Jaeho Kim

Dr. Myongjin Kim

Dr. Georgia Kosmopoulou

Dr. Jeremy Short

Acknowledgements

I wish to express my gratitude to my advisor, Professor Qihong Liu, for his guidance and unconditional support during all these years of graduate school, and for being a excellent mentor. I also wish to thank Jaeho Kim, Myongjin Kim, Georgia Kosmopoulou, Daniel Nedelescu, Jeremy Short, and seminar participants at University of Oklahoma for helpful comments.

Table of Contents

List of Tables

Abstract

My dissertation studies economics of innovation and industrial organization. In the first chapter, I analyze the effect of geographic proximity on knowledge flows by estimating the impact of travel time on patent citation frequency between US Metropolitan Statistical Areas (MSA) for the period of 1976 to 2006. The variation in travel time comes from airlines' entry and exit. Using a gravity model, I show that on average a 1% reduction in travel time leads to a 0.17% increase in patent citations. The intuition is that faster travel facilitates face-to-face meetings, leading to more learning opportunities. My results are robust when I take into account various types of endogenous shocks.

The second chapter introduces two modifications to standard models of media markets with advertisers on one side and readers/viewers on the other. In the first modification, advertisers make strategic choices on the quality of their ads which affect the utility of readers/viewers joining the same platform. I show that this feature of strategic agents leads to qualitatively different econometric specifications for the estimation of group externality parameters. Relative to benchmark case of passive agents, prices on both sides are lower under strategic agents, benefiting the agents at the cost of platforms. In the other modification, I introduce independent retailers between platforms and readers/viewers. Our results suggest that this modification has no impact on estimating the group externality parameters. However, equilibrium price on either side depends on group externality parameters at both sides. This is in sharp contrast to standard results where prices on one side depend only on group externality parameter of the other

side. In the special case where each platform is split into two independent divisions, I find that equilibrium price is the same across all four divisions, and this common price depends on the product of the group externality parameters of the two sides.

The third studies the impacts of innovations from upstream and downstream industries. Using input-output account and patent citation data, I identify vertical relations between industries in both intermediate good markets and the patent system. My results show that in intermediate good markets, a 1% increase in research and development (R&D) expenditures in upstream industries will decrease a downstream firm's R&D expenditure by 0.729%, while a 1% increase in R&D of the downstream will reduce a upstream firm's market value by 0.907%. Meanwhile, in the patent system, the upstream to downstream R&D and market value elasticities are estimated to be 0.983 and 1.053.

Chapter 1

Learning by Meeting: Assessing the Impact of Travel Time on Patent Citations

1.1 Introduction

The study of industrial agglomeration dates back to Marshall (1890), who explains several reasons why firms are geographically concentrated, one of which is that firms located in industrial clusters can easily learn from the firms nearby, indicating an positive correlation between geographic proximity and knowledge flows. Jaffe et al. (1993) are the first to use patent citations as a proxy for knowledge spillovers and show empirical evidence supporting Marshall's claim. Using a matching strategy, they find pairs of patents with a citation link are more likely to be applied in the same country, state or metropolitan area.

Although numerous subsequent studies using the same methodology obtain similar results (Almeida and Kogut, 1997, 1999; Hicks et al., 2001; Sonn and Storper, 2008), we still have little knowledge about the causal relationship between geographic proximity and knowledge flows. One question people might ask is: do knowledge flows happen more frequently when firms are physically close to one another to begin with? The question is difficult to answer because some unobserved factors might cause two firms to co-locate and cite each other. For example, Thompson and Fox-Kean (2005) find that localization of patent citations somewhat disappears after they further require that patents in the control group and the treatment group are from the same technology subclass.

In this article I study how geographic proximity affects knowledge flows from a different perspective. I examine the impact of travel time on patent citation frequency between two US Metropolitan Statistical Areas (MSAs). In particular, I construct a unique panel dataset that links inter-MSA travel time to their patent citations. There are several benefits from this strategy. First, since travel time varies over time (while physical distance does not), it allows me to estimate a fixed-effect model to account for any time-invariant unobservables at MSA-pair level. Second, the variation of travel time between MSAs is from airlines' route choices, which can be considered as external forces that bring a few firms closer to one another, which mitigates concerns associated with firms' endogenous co-localization decisions. Moreover, it allows me to look into some special cases in which travel time reductions are due to some events that are arguably exogenous to patent citations, such as the opening of a new airline hub. Lastly, Giroud (2013) argues that travel time is a better proxy for geographic proximity than physical distance is, because it is possible that two cities are physically close, but traveling between them involves a long road trip because there is no direct flight.

Another novelty about this paper is that I adopt the gravity model from the international trade literature. The motivation is that trade is similar to knowledge flows in a sense that they are both examples of economic interactions that are negatively correlated with distance. Nonetheless, this article deviates from a traditional gravity model in that it utilizes the time-varying nature of travel time to conduct a panel analysis.

My results show that travel time has a negative impact on inter-MSA patent citation frequency. On average a 1% reduction in travel time increases citation frequency between two MSAs by 0.17%. The intuition is that faster travel makes business interactions more convenient and therefore facilitates face-to-face meetings, leading to more opportunities of knowledge spillovers. Arvey (2009) argues that face-to-face meetings allow people to capture verbal and non-verbal behavioral styles typically not observed on telecommunications equipment, which sometimes can result in breakthrough thinking.

Next, I explore if there exists a non-log-linear relationship between travel time and patent citations, because in reality travelers are indifferent toward travel time reductions that only save them a little time. For this issue, I consider reductions in travel time as treatments, and divide the treatments into several groups, each corresponding to a certain amount of travel time reduction. I find the impacts are significant only when travel time reductions are greater than one hour.

It is possible, however, that sometimes airlines' entry and exit are endogenous to patent citations. For instance, suppose an airline started to operate a new direct flight route between MSA i and j in year t, which reduced the travel time, and we observed an increase in citation frequency between the two MSAs afterwards. I would like to attribute the increase in patent citations to the reduction in travel time, but there is an alternative story: there might be a time-varying unobserved shock, say α_{ijt}, that affects both airlines' route choices and patent citations. For robustness checks, I first assume the unobserved shock, α_{ijt}, can be decomposed into α_{it} and α_{jt}. This happens when there are two separate shocks to MSA i and j. I find my results don't change much when I control for those types of shocks in my specification.

What if α_{ijt} cannot be decomposed? α_{ijt} cannot be decomposed if the shock is due to some interactions between two MSAs. An example could be the following: a large firm in MSA i recently acquired another large firm in MSA j, which resulted in an increase in patent citations between the two MSAs. Meanwhile, the firms' lobbyists successfully convinced an airline to start operating direct flights between the two MSAs. For this type of shock, I examine the cases in which reductions in travel time are due to the opening of a new hub. Arguably it is unlikely that a non-decomposable endogenous shock to the patent citations between some MSA pairs such as a merger and acquisition is strong enough to trigger a hub opening. I find the impacts of travel time reductions due to hub openings are all positive and statistically significant in my results.

Lastly, I check if there exists a reverse causality problem. It is likely that airlines are more willing to operate a new flight route after they observe a lot of patent citations between two MSAs. I argue that if this is the case, then we should see a correlation between current patent citations and future travel time reductions, because there is on average a two-year gap between the time a patent is applied (the time it makes citations) and the time the patent is granted (the time its information becomes publicly accessible and airline carriers observe the citations). I perform the robustness check by including future travel time reductions as explanatory variables in my specification. I find the coefficients on future travel time reductions are all small and insignificant.

Next, I show that the impacts of travel time are heterogeneous across technology fields, with the impacts being the strongest on citations within Chemical patents, and insignificant on citations within Drugs & Medical patents. This can be due to the fact that knowledge in Drugs & Medical patents (such as compound formulas) is more self-evident than other technology fields. Besides that, I don't find evidence that the impacts are heterogeneous over different time periods.

I don't directly test Marshall's theory, which hints at within-region knowledge spillovers. That being said, my study still has important policy implications on air transportation infrastructure, innovation, and regional economic development. It shows that there are empirically undocumented positive externalities of transportation infrastructure on knowledge spillovers across regions.

The rest of the chapter is organized as follows. Section 1.2 reviews the related literature. Section 1.3 discusses the empirical methodology, followed by Section 1.4, in which I describe the data. Results are presented in Section 1.5. Section 1.6 performs multiple robustness checks and Section 1.7 discusses several extensions. Section 1.8 concludes.

1.2 Related Literature

Due to the non-rival property of knowledge, a few firms conducting R&D activities will benefit other firms unintentionally, which results in positive externalities. Since the seminal work of Nelson (1959) and Arrow (1962), the extent to which private investments in research and development create knowledge spillovers has generated an extensive body of literature.

What receives much less attention is the factors determining knowledge spillovers. What makes them happen? A challenge encountered by researchers is that knowledge flows are invisible. Krugman (1991) claims that it would be really hard to measure knowledge flows as they "leave no paper trail".

Jaffe, Trajtenberg and Henderson (1993, hereafter JTH) are the first to solve the problem by using patent citations as a proxy for knowledge flows. For a patent to be granted in the US, inventors are required by law to report any known prior art by citing the relevant patents. If Patent A cites Patent B, it means the inventor of patent A probably received some knowledge from patent B. In their path-breaking study on the geographic localization of knowledge flows, JTH construct three sets of patents: 1) a set of originating patents; 2) a set of patents citing the originating patents; 3) and a set of control patents that are similar to the set of citing patents in that they are from the same technology class and are applied in the same year, except that they don't cite the originating patents. They compare the probability that a citing patent matches the originating patent geographically (they come from the region), with the probability that a non-citing control patent matches the originating patent geographically. JTH find that patent pairs with a citation link are more likely to come from the same country, state or metropolitan area.

JTH's novel empirical strategy is widely applied in the subsequent literature. Almeida and Kogut (1997) find that patent citations of small firms are more geographically localized than large firms in the semiconductor industry. Almeida and Kogut (1999) show evidence that geographic localization of patent citations varies across regions and ideas are transmitted through labor markets by the inter-firm mobility of engineers. Hicks et al. (2001) conclude that patents assigned to US companies are more likely to cite research papers produced by the nearby public universities. Sonn and Storper (2008) show that localization of knowledge spillovers has been increasing over time.

Nevertheless, to my best knowledge almost all of the studies in the related literature are cross-sectional, making causal inference difficult. Any unobserved common shock to both firms' locality and their citations can compromise the validity of empirical results. Due to the constraint of patent data, the control variables used in most studies are primary technology class and application year, but many other factors might cause firms to cluster and cite one another (for example, two firms may be in the same business alliance).

Maurseth and Verspagen (2002) regress citation frequency among European regions on their physical distance. They conclude that physical distance has a negative impact on inter-regional citation frequency. Their analysis has similar endogeneity issues: countries close to one another may have similar unobserved characteristics (such as similar cultures and social values), which is why we observe more citations.

Agrawal et al. (2016) find the the stock of highways in a state has positive impact on its innovation activities and knowledge flows. While their study sheds light on knowledge spillovers *within* a region, this paper focuses on *across* regions.

This paper contributes to the literature in the following aspects. First, unlike physical distance, travel time varies over time, which allows me to estimate a MSA-pair fixed-effect model to take into account any time-invariant MSA-pair heterogeneity. Second, the source of variation in travel time is from airlines' route choices, which can be considered as external forces that bring a few MSAs closer to one another. It mitigates concerns associated with firms' endogenous co-localization. Moreover, it allows me to look into some special cases that travel time reductions are due to some events exogenous to patent citations, such as the opening of a new airline hub.

1.3 Identification and Methodology

In international trade, a traditional gravity model can be written as

$$X_{ij} = \alpha_0 Y_i^{\alpha_1} Y_j^{\alpha_2} D_{ij}^{\alpha_3} \epsilon_{ij}, \tag{1.1}$$

where X_{ij} is the trade volume between country i and j, Y_i and Y_j are the economy size of country i and j, D_{ij} is the physical distance between i and j, and ϵ_{ij} is the error term. There is a tradition in the profession that Equation 1.1 is estimated in log-form:

$$\ln X_{ij} = \ln \alpha_0 + \alpha_1 \ln Y_i + \alpha_2 \ln Y_j + \alpha_3 \ln D_{ij} + \ln \epsilon_{ij}, \tag{1.2}$$

Silva and Tenreyro (2006) point out that estimating Equation 1.2 using OLS has two problems. First, in many datasets, ϵ_{ij} is heteroskedastic. Since $E(\ln \epsilon_{ij})$ depends on the variance of ϵ_{ij}, $E(\ln \epsilon_{ij})$ is correlated with other regressors in the presence of heteroskedasticity, which violates the classical assumptions of OLS. Second, in many observations, X_{ij} is zero, which creates a problem when taking logarithm. Silva and Tenreyro (2006) show that it is appropriate in this case to estimate Equation 1.2 with a Poisson regression.

I follow their strategy and estimate a fixed-effect Poisson model. Another reason for doing this is that Poisson regression is a natural candidate for estimating count data (patent citation frequency). Here I assume the arrival rate of citation frequency from MSA i to MSA j in year t, λ_{ijt}, is a deterministic function,

$$ln(\lambda_{ijt}) = \alpha_{ij} + \alpha_t + \beta \cdot ln(travel_{ijt}) + \boldsymbol{\theta}' \boldsymbol{X_{ijt}}, \tag{1.3}$$

where α_{ij} and α_t are MSA-pair fixed effects and year fixed effects respectively, $travel_{ijt}$ is the travel time between MSA i and MSA j in year t,$\boldsymbol{X_{ijt}}$ is a vector of control variables, and β is the parameter of interest. The randomness of this specification comes from the Poisson distribution: the probability of observing c_{ijt}, the citation frequency from from MSA i to MSA j in year t, is

$$pr(c_{ijt}) = \frac{e^{-\lambda_{ijt}} \lambda_{ijt}^{c_{ijt}}}{c_{ijt}!} \tag{1.4}$$

However, using traditional maximum likelihood methods by maximizing the log-likelihood function $\sum_i \sum_j \sum_t ln(pr(c_{ijt}))$ will generate inconsistent estimates of β due to the incidental parameters problem. Instead, I use a conditional fixed-effect maximum likelihood estimator by Hausman et al. (1984, here after HHG), in which MSA-pair dummies are dropped by conditioning c_{ijt} on the sum of citation frequencies over time ($\sum_t c_{ijt}$).

Wooldridge (1999) proves that HHG's conditional fixed-effect Poisson estimator is consistent whether the data is over-dispersed or under-dispersed, as long as Equation 1.3 holds. He also proposes an estimator of $var(\hat{\beta})$ that is robust to heteroskedasticity and serial correlation of the dependent variable, which is used in this paper. Standard errors are clustered within MSA pairs.

1.4 Data

1.4.1 Travel Time

Ideally, to calculate the travel time between two MSAs, we want to survey some travelers and ask how much time they spent on the trips. In reality, such data are missing for most MSA pairs. I follow the same steps as Giroud (2013) to calculate travel time. I first assume travelers either drive or fly, and they always pick the routes and means of transportation that minimize travel time.

Specifically, to measure the travel time between MSA i and MSA j in year t, I first calculate the driving time between MSA i and MSA j using HERE Maps. In Appendix A I provide details on how this driving time is calculated. Driving time is used as travel time if it is faster than flying between MSAs.

Next, I calculate the air travel time, which has three components: (1) the driving time between MSA i and its corresponding airport; (2) the driving time between MSA j and its corresponding airport; (3) the time it takes to fly between the two airports.

(1) and (2) are calculated using HERE Maps. (3) is equal to the sum of wait time at the origin airport, layover time (only for non-direct flights), and flight time. Wait time and layover time are both assumed to be 1 hour. Flight time data is from the T-100 Domestic Segment Database for period of 1990-2006 and ER-586 for the period of 1976-1989, which include all domestic flights.[1] The flight time for each route is equal to "ramp-to-ramp time", which is defined as "the time computed from the moment an aircraft first moves under its own power for purposes of flight, until it comes to rest at the next point of landing", according to the Bureau of Transportation Statistics.

When an MSA is served by multiple airports, I search for the fastest flight route and compute the air travel time. The travel time between MSA i and MSA j equals either the air travel time or the driving time, whichever is faster.

[1] ER-586 data is available on the National Archives and Records Administration website.

1.4.2 Patent Data

According to the US Patent and Trademark Office (USPTO), a patent has to be non-obvious, novel and useful. Once a patent is granted, a public document is released with information about the patent. One section of the document is "reference cited", which lists all the citations the patent makes.

There have been disputes over the validity of using patent citations as an indicator of knowledge flows. In fact, patent citations can be added by the patent examiner, if she considers it necessary, or even influenced by the law firms hired to file the patent (Wagner et al., 2014), but overall patent citations are "noisy" but meaningful indicators of knowledge flows, according to the surveys of inventors conducted by Jaffe et al. (2000). If patent A cites patent B, it is likely that patent A's inventor received some sort of knowledge from patent B.

The patent data is from the NBER Patent Data Project, which is an updated version of a older dataset constructed by Hall et al. (2001). It includes all the patents granted by the USPTO from 1976 to 2006. For each patent, I observe the patent identifier number, application date, grant date, citations and technology class. Hall et al. (2001) categorize all patents into six technology classes: Chemical, Computers & Communication, Drugs & Medical, Electrical & Electronics, Mechanical, and Others.[2] I complement the NBER's patent data with another dataset constructed by Li et al. (2014), which also records patent inventors' addresses of residence(city, state, ZIP

[2]The category Others includes several sub-categories, including Agriculture, Husbandry, Food, Amusement Devices, Apparel & Textile, Earth Working & Wells, Furniture, House Fixtures, Heating, Pipes & Joints, Receptacles, and Miscellanenous-Others. Note most of these sub-categories are technologically unrelated.

code and coordinate).[3] Finally, I keep all the utility patents applied by inventors that reside in the contiguous 48 US states and District of Columbia.

Citation frequency between MSAs serves as a proxy for knowledge flows. I use the Metropolitan Statistical Area defined by the Office of Management and Budget in 2009 as my basic geographic unit. There are a total of 363 MSAs in the lower 48 states and District of Columbia. I assign each patent in my sample to an MSA according to its inventor's address. Finally, I count patent citation frequency at the MSA-pair-year level to construct the variable c_{ijt}, the number of citations made from MSA i to MSA j in year t. For example, if patent A applied in Oklahoma City in 1995 cites patent B applied in Boston, then I will add one to the citation frequency from Oklahoma City to Boston in 1995. I use the application year rather than the grant year of the citing patent, because the former better approximates the time when knowledge flows take place. In practice, the construction of c_{ijt} is further complicated by the fact that sometimes a patent has multiple inventors residing in different MSAs. For this issue, I follow the same rule as Sonn and Storper (2008): suppose patent A cites patent B, and patent A has M inventors and patent B has N inventors, then each inventor pair carries a fraction of $1/(M \cdot N)$ citations.

For example, suppose Patent A has three inventors, one living Dallas and two living in Memphis, and patent B has two inventors, one living in Seattle and the other living in Portland. The citation from Dallas to Seattle equals $1/3 * 1/2 = 1/6$ and the citation from Dallas to Portland equals $1/3 * 1/2 = 1/6$. The citation from Memphis to Seattle equals

[3] The matching of the two datasets performs well. Less than 0.01% observations in the NBER patent data were not able to find a match in the dataset by Li et al. (2014)

to $2/3 * 1/2 = 1/3$, and the citation from Memphis to Portland equals $2/3 * 1/2 = 1/3$. In total, they sum up to 1.[4]

I include log of patents applied in the citing MSAs (app_{it}) and log of patent stock in the cited MSAs (stk_{jt}) as two of my control variables. The reason is that MSAs where a lot of patents are applied in a year are more likely to cite other MSAs, and MSAs with a large amount of patent stock are more likely to be cited. These two variables control for the overall innovation activities of the citing and cited MSAs.

For app_{it}, I count the number of patents applied in the citing MSA i in year t. For stk_{jt}, I first count the number of patent granted in cited MSA j by year t and then use a perpetual inventory method to calculate patent stock with a depreciation rate of 15%. I use the grant year to calculate stk_{jt}, because only after a patent is granted, its information become publicly accessible and the patent starts to receive citations (Mehta et al., 2010). The same as c_{ijt}, when a patent has N inventors residing in different MSAs, each inventor carries a share of $1/N$. Finally, I add one to both variables before taking the log to deal with the zeroes.

Another important control variable is the technological similarity of the cited and citing MSAs. MSAs working on R&D in related fields are more likely to cite one another. In the spirit of Jaffe (1988), I construct two vectors, $\boldsymbol{p_{it}}$ and $\boldsymbol{q_{jt}}$. $\boldsymbol{p_{it}} = (p_{1it}, p_{2it}, p_{3it}, p_{4it}, p_{5it})'$, where p_{cit} is the share of *patents applied* in category c ($c = 1, 2, 3, 4, 5$) in the citing MSA i in year t, and $\boldsymbol{q_{jt}} = (q_{1it}, q_{2it}, q_{3it}, q_{4it}, q_{5it})'$, where q_{cit} is the share of *patent*

[4]This would mean that some observations of c_{ijt} are not integers. This is not a problem, as Wooldridge (1999) shows that the conditional fixed-effect Poisson estimators are consistent as long as the dependent variable is non-negative.

stock in category c in cited MSA j. $\boldsymbol{p_{it}}$ captures the technology structure of the *ongoing* innovation activities of MSA i in year t, while $\boldsymbol{q_{it}}$ captures the technology structure of the *historical* innovation activities in MSA j by year t. [5] Finally, the technological similarity between MSA i and MSA j in year t is measured by the technological distance, d_{ijt}, the euclidean distance between the two coordinates.

$$d_{ijt} = \sqrt{\Sigma_{c=1}^{5}(p_{cit} - q_{cit})^2} \tag{1.5}$$

Higher d_{ijt} means lower technological similarity, with $d_{ijt} = 0$ indicating the best technological match. Unlike Jaffe (1988), this measure of technology similarity captures the dynamic feature of patent citations. It means if MSA i's ongoing innovation activities are similar to the historical innovation activities in MSA j, then patents in MSA i will cite patents in MSA j more often.

1.4.3 Other control variables

I also control for the overall economic activities of MSAs. The variables I use are per capita personal income and population, both of which are from the US Bureau of Economic Analysis Regional Economic Accounts. Per capita income is deflated in real term using 2009 US dollar.

[5] I don't include the sixth category "Others" in constructing the vectors because MSAs with the same share of Others patents can still have very different technological structure, since the category Others is composed of various technologically unrelated sub-categories.

1.4.4 Summary statistics

Figure 1 depicts the citation (c_{ijt}) trend over years. We see citations steadily increase over years and peak in year 2002 at around 800,000. There is a drastic drop after 2002, which can be explained with the following example: my sample includes patents granted until 2006. For a patent granted in 2006 to make any citations in 2006, it has to be applied in 2006 as well (because I use the application year of the citing patent as the citation year), which is rare because there is on average a two-year gap between the time a patent is applied and the time it is granted.

Table 1.1 presents the distribution of citation frequency over technology fields. Most of the citations are within-category (both the citing and cited patents are from the same category). Citations within Computers & Communication ranks the first, accounting for 22.58% of the citations in my sample, followed by Electrical & Electronics, and Drugs & Medical. The most frequent across-category citations are made from Computers & Communication patents to Electrical & Electronics patents.

Table 1.2 ranks citation frequency by MSA-pair. I find that the top 10 MSA-pairs are all large and populous MSAs, including some famous national R&D centers, such as Silicon Valley and Boston's Route 128 region. Table 1.3 provides summary statistics. There are 363 MSAs and 31 years in my sample, resulting in a total of $363*362*31 = 4,073,586$ MSA-pair-year observations. Inter-MSA citation frequency is over-dispersed with a mean of 1.548 and a standard deviation of 28.814. The average travel time between MSAs is about 386 minutes. The average number of patents applied is 110 and the average patent stock is 566. MSAs in the sample have an average

Table 1.1: Citations across technology fields

Cited category	**Citing category**	**Frequency**	**Percentage**
Comp. & Comm.	Comp.s & Comm.	1,793,259	22.58%
Electrical	Electrical	1,078,642	13.58%
Drugs & Medical	Drugs & Medical	975,358	12.28%
Chemical	Chemical	869,150	10.94%
Others	Others	673,606	8.48%
Mechanical	Mechanical	551,708	6.95%
Electrical	Comp. & Comm.	198,897	2.50%
Comp. & Comm.	Electrical	148,067	1.86%
Chemical	Drugs & Medical	133,700	1.68%
Chemical	Others	123,972	1.56%
The rest		1,395,843	17.58%
Total		7,942,202	100%

This table shows the distribution of citations across technology fields. Patent categories are defined by (Hall et al., 2001). The category "Others" includes Agriculture, Husbandry, Food, Amusement Devices, Apparel & Textile, Earth Working & Wells, Furniture, House Fixtures, Heating, Pipes & Joints, Receptacles, and Miscellaneous-Others.

population of 581,726 and an average per capita income of $27,085.

1.5 Results

1.5.1 Impacts of travel time on citations

Main results are presented in Table 1.4. The dependent variable is inter-MSA citation frequency, and the independent variable of interest is log of travel time. All the regressions in this table include log of patents applied in the citing MSA, log of patent stock in the cited MSA, and technological distance as control variables. The coefficients on the variables in log-form can be interpreted as elasticities. In column (1), no fixed effects are used. It is estimated that a 1% reduction in travel time is associated with an 0.35% increase in citation frequency. The elasticities of patents applied and patent stock are all greater than one and statistically significant. The coefficient on technological distance is negative and significant, which is within my expectation as higher technological distance corresponds to lower technological similarity.

In column (2), I additionally control for MSA-pair fixed effects and year fixed effects. The results change a lot compared to column (1), indicating potential endogeneity issues in the previous specification. The magnitude of the elasticity of travel time decreased drastically to 0.194, but it is still highly statistically significant at 1% level. Moreover, the elasticity of patents applied remains similar to column (1), but the coefficient on patent stock decreased significantly (from 1.055 to 0.625). The coefficient on technological distance changed a lot as well (from -1.622 to -0.938).

Table 1.2: MSA-pairs with the most citation frequency

Cited MSA	**Citing MSA**	**Frequency**
San Jose-Sunnyvale-Santa Clara, CA	San Francisco-Oakland-Fremont, CA	73,643
San Francisco-Oakland-Fremont, CA	San Jose-Sunnyvale-Santa Clara, CA	62,907
New York-Northern New Jersey-Long Island	San Jose-Sunnyvale-Santa Clara, CA	47,851
Boston-Cambridge-Quincy, MA-NH	San Jose-Sunnyvale-Santa Clara, CA	38,275
New York-Northern New Jersey-Long Island	Boston-Cambridge-Quincy, MA-NH	32,055
New York-Northern New Jersey-Long Island	San Francisco-Oakland-Fremont, CA	31,623
San Jose-Sunnyvale-Santa Clara, CA	Boston-Cambridge-Quincy, MA-NH	27,751
New York-Northern New Jersey-Long Island	Chicago-Naperville-Joliet, IL-IN-WI	27,088
Los Angeles-Long Beach-Santa Ana, CA	San Jose-Sunnyvale-Santa Clara, CA	25,027
New York-Northern New Jersey-Long Island	Los Angeles-Long Beach-Santa Ana, CA	24,001

This table lists MSA-pairs with the most patent citations. MSAs are defined the by the Office of Management and Budget in 2009.

Table 1.3: Summary statistics

Variable	Mean	Std. Dev.	Min	Max
Citation frequency	1.548	28.814	0	10,865
Travel time	385.769	119.6565	18	2,542
Patents applied	110.334	368.364	0	7,225
Patent stock	566	1,856	0	33,789
Technology distance	0.427	0.204	0	1.414
Population	581,726	1,390,271	8172	19,253,644
Per capita income	27,085	6,922	10,659	87,477

The construction of all the variables in this table is discussed in Section 4. Travel time (one-way) is in minutes. Per capita income is deflated using 2009 dollar.

A key issue in column (2) is that I don't take into account the overall economic activities of the MSAs. In column (3), I include population and per capita income of both citing and cited MSAs in my regression. I found the coefficients on travel time, patents applied, patent stock and technological distance are similar to those of column (2). Looking at the coefficients on MSA population and income, the elasticities of cited MSA population and income are all significant and positive. This is within my expectation, as larger or wealthier MSAs generally have more high-quality patents and more business travelers. The elasticities of citing MSA population and income are all insignificant. This seems counterintuitive because we expect larger MSAs to cite more for the same reason as above. A possible explanation is that larger firms, typically located in populous MSAs, tend to be less research-active than startups (Almeida and Kogut, 1997; Breitzman, 2013).

Table 1.4: Impacts of travel time on citations

Variable	(1)	(2)	(3)	(4)
Travel time	-0.353***	-0.194***	-0.172***	-0.166***
	(0.019)	(0.043)	(0.044)	(0.041)
Patents applied	1.041***	1.018***	1.015***	0.759***
	(0.003)	(0.011)	(0.012)	(0.015)
Patent stock	1.055***	0.625***	0.561***	0.528***
	(0.003)	(0.016)	(0.020)	(0.020)
Tech distance	-1.622***	-0.938***	-0.926***	-1.024***
	(0.030)	(0.041)	(0.041)	(0.039)
Cited MSA income			0.336***	0.290***
			(0.080)	(0.078)
Cited MSA pop.			0.437***	0.309***
			(0.071)	(0.076)
Citing MSA income			-0.023	0.047
			(0.071)	(0.067)
Citing MSA pop.			-0.024	-0.456***
			(0.068)	(0.071)
$\log \omega_{it}$				0.335***
				(0.010)
$\log \omega_{jt}$				0.080***
				(0.013)
fixed effects	No	Yes	Yes	Yes
Log likelihood	-3152478.9	-1616280.8	-1615163.9	-1600267.5
N	4,073,586	4,073,586	4,073,586	4,073,586

Standard errors in parentheses.* ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). The dependent variable is citation frequency between MSAs.

1.5.2 Non-log-linear relationship between travel time and patent citations

The gravity model assumes a log-linear relationship between travel time and patent citations. What if the impacts are not log-linear? In reality travelers are generally indifferent toward travel time reductions that only save them a couple minutes. To explore this possibility, I write an alternative specification,

$$ln(\lambda_{ijt}) = \alpha_{ij} + \alpha_t + \boldsymbol{\beta}' \boldsymbol{trt_{ijt}} + \boldsymbol{\theta}' \boldsymbol{X_{ijt}}, \tag{1.6}$$

where I replace travel time with $\boldsymbol{trt_{ijt}}$, a vectors of treatments, with each element equal to one if (one-way) travel time between MSA i and j is reduced by a certain amount by year t.[6] Specifically, I categorize travel time reductions into five groups in Table 1.5: 0-15 minutes, 15-30 minutes, 30-60 minutes, 60-120 minutes, and more than 120 minutes. Results are presented in Table 1.5. I find the impact of 0-15 minutes' reduction in travel time is small and insignificant. The impacts become marginally insignificant when reductions in travel time gradually increase from 15 minutes to 60 minutes, and finally become significant when reductions are greater than one hour. It is estimated that a 60-120 minutes' reduction in travel time will increase citations by 9.2%, while a more-than-two-hour reduction leads to a 28.4% increase in patent citations.

[6] Appendix B provides details on the construction of $\boldsymbol{trt_{ijt}}$.

Table 1.5: Impacts of reduction in travel time on citations

Variable	(1)
Travel time reductions	
0-15 minutes	-0.002
	(0.016)
15-30 minutes	0.028
	(0.018)
30-60 minutes	0.031
	(0.020)
60-120 minutes	0.092***
	(0.022)
more than 120 minutes	0.284***
	(0.073)
Patents applied (log)	1.015***
	(0.012)
Patent stock (log)	0.560***
	(0.020)
Technological distance	-0.925***
	(0.041)
MSA-pair fixed effects and year fixed effects	Yes
Log likelihood	-1615278.7
Number of observations	4,073,586

Standard errors in parentheses.* ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). The depedent variable is citation frequency between MSAs. Coefficients on MSA population and per capita income are available upon request.

1.6 Robustness checks

The source of variation in travel time in my data comes from changes in airlines' route choices. There is a potential endogeneity problem. For example, suppose an airline started to operate a new direct flight route between MSA i and j in year t, which reduced the travel time, and we observe an increase in citation frequency between MSA i and j afterwards. This could be because there is an unobserved endogenous shock, say α_{ijt}, that affects both airlines' route choices and patent citations. In this section, I perform several robustness checks to address these concerns.

1.6.1 Decomposable endogeneous shocks

One possibility is that α_{ijt} can presumably be decomposed into α_{it} and α_{jt}. This happens when there are two separate shocks to MSA i and j. The most straightforward solution is to include α_{it} and α_{jt} as dummies in my specification. Unfortunately, that would mean including additional 363 MSAs $\times$ 31 years $\times$ 2 = 22506 dummies in my regression. Even the best computing power available today can't handle this. To bypass this problem, I introduce two variables, ω_{it} and ω_{jt}, where $\omega_{it} = 1/362 \sum_{s \neq j}^{363} c_{ist}$, the average citations that MSA i (the citing MSA) makes to other MSAs (excluding MSA j) in year t, and $\omega_{jt} = 1/362 \sum_{s \neq i}^{363} c_{sjt}$, the average citations that MSA j (the cited MSA) receives from other MSAs (excluding MSA i) in year t. The idea is that if there is a positive shock to α_{it}, then it should not only increase c_{ijt}, the citations MSA i makes to MSA j, but also increases $c_{ist}(s \neq j)$, the citations MSA i makes to other MSAs, since $\alpha_{ist} = \alpha_{it} + \alpha_{st}$ by construction. Therefore, a shock to α_{it} can be taken into account by controlling for the average citations MSA i makes to other MSAs, and similarly, a shock to α_{jt} can be traced by controlling for the average citations MSA j receives from other MSAs.

In column (4) of Table 1.4, I present results with ω_{it} and ω_{jt} included as controls. The coefficient on travel time is similar compared to column (3). Besides, the coefficients on ω_{it} and ω_{jt} are positive and statistically significant.

1.6.2 Non-decomposable endogeneous shocks

What if α_{ijt} is non-decomposable? This happens when the endogenous shock is the result of some interactions between the two MSAs. For example, a large firm in MSA i recently acquired another large firm in MSA j, which resulted in an increase in patent citations between the two MSAs. Meanwhile, the firms' lobbyists successfully convinced an airline to start operating direct flights between the two MSAs. For this type of shock, I examine the cases in which reductions in travel time are due to the opening of a new hub. Giroud (2013) is the first to use hub openings as an exogenous source of variation in travel time. Similar to his argument, it is less likely that a non-decomposable endogenous shock, such as a merger and acquisition, is strong enough to trigger a hub opening. Here I consider the following specification,

$$ln(\lambda_{ijt}) = \alpha_{ij} + \alpha_t + \beta \cdot hubtrt_{ijt} + \boldsymbol{\theta}^{'} \boldsymbol{X_{ijt}} \tag{1.7}$$

$hubtrt_{ijt}$ equals 1 if travel time is reduced by more than one hour due to the opening of a new hub, and 0 otherwise. If the results I get in the previous sections are due to some non-decomposable endogenous shock α_{ijt}, then my estimate of β should be insignificant, since $hubtrt_{ijt}$ is uncorrelated with a non-decomposable α_{ijt}. In Table 1.6, I represent the results. The control variables are identical to those of column (4) of Table 1.4, in which both ω_{it} and ω_{jt} are included. The coefficient of $hubtrt_{ijt}$ is positive and significant at 1% level.

Table 1.6: Reduction in travel time due to hub openings

Variable	(1)
$hubtrt_{ijt}$	0.221**
	(0.062)
Patents applied (log)	0.759***
	(0.015)
Patent stock (log)	0.528***
	(0.020)
Technological distance	-1.024***
	(0.039)
Average citations made ($\log \omega_{it}$)	0.336***
	(0.013)
Average citations received ($\log \omega_{jt}$)	0.080***
	(0.010)
MSA population and income	Yes
MSA-pair and year fixed effects	Yes
Log likelihood	-1600407.6
Number of observations	4,073,586

Standard errors in parentheses.* ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). Coefficients on MSA population and income are available upon request.

1.6.3 Simultaneity

The above robustness checks don't rule out the possibility of reverse causality. Suppose the citations from MSA i to MSA j increase in year t. Seeing there is a demand for knowledge flows, airlines decide to operate a direct flight between the two MSAs. I argue that if this is the case, there should be a correlation between current patent citations and future travel time reductions. The reason is the following. By construction, my dependent variable c_{ijt}, citations from MSA i to MSA j in year t are the citations made by patents *applied* in year t. A patent's information is publicly accessible only after it is granted and the grant process typically takes two years, which means if there is a positive shock to c_{ijt}, airlines should normally observe it two years later on average.

In Table 1.7, I test such claim by including both past and future travel time reductions in my specification. The dummies equal one if travel time was reduced by more than one hour in year $t \pm s$ and the dependent variable is c_{ijt}, citations in year t. In column (1), I first add past travel time reductions. I find that reductions in travel time in year t increase citations in the same year by 5.8%. The impacts become stronger a year later and peak at 11.1% two years later. After two years, the impacts gradually diminish. In column (2), I additionally include future time reductions as explanatory variables, travel time reductions in year $t + 1$ and $t + 2$. All coefficients on future time reductions are insignificant while the coefficients on past reductions remain similar to column (1).

Table 1.7: Impacts of lags and leads of travel time on citations

Variable	**(1)**	**(2)**
Travel time reduction $(t+2)$		-0.011
		(0.019)
Travel time reduction $(t+1)$		0.022
		(0.026)
Travel time reduction (t)	0.058**	0.061**
	(0.023)	(0.027)
Travel time reduction $(t-1)$	0.084***	0.086***
	(0.028)	(0.032)
Travel time reduction $(t-2)$	0.111***	0.114***
	(0.031)	(0.035)
Travel time reduction $(t-3)$	0.099***	0.101***
	(0.030)	(0.032)
Travel time reduction $(t-4)$	0.083***	0.085***
	(0.027)	(0.030)
Travel time reduction (+4 yrs ago)	0.064***	0.066***
	(0.024)	(0.027)
Log likelihood	-1615119.4	-1615113.4
Number of observations	3,679,368	3,416,556

Standard errors in parentheses.* ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). All regressions control for patents applied in the citing MSA, patent stock in the cited MSA, population and income of both citing and cited MSAs, MSA-pair fixed effects, and year fixed effects

1.7 Extensions

1.7.1 Heterogeneous impacts of travel time across technology fields

Are the impacts different across technology fields? In column (1) to (6) of Table 1.8, I present results when the dependent variables are citations within certain technology fields, that is, both the citing and cited patent come from the same patent category. Looking at column (1) to (5), the impact is the strongest within Chemical category at -0.45, followed by Mechanical, Electrical & Electronics, and Computers & Communication. Surprisingly, the impact within Drugs and Medical category is insignificant. This can be due to the fact that the knowledge in drugs and medical patents (such as compound formulas) is more self-evident than other industries. In column (6), the dependent variable is the citations whose citing and cited patents are from the same category, the sum of the dependent variables from column (1) to (5). I find that on average a 1% reduction in travel time will increase within-category citations by 0.23%. In column (7), the dependent variable is citations across categories. The elasticity of travel time is smaller compared to column (6). It is estimated that a 1% reduction in travel time leads to a 0.12% increase in citation frequency across patent categories.

1.7.2 Early years vs. later years

It is natural to expect the impacts of travel time reductions to be stronger in the early years, since the recent development of communication technology

Table 1.8: Impacts of travel time on citations: different categories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Categories	Chemical	Comp.&Comm.	Drugs&Med.	Elec.	Mechanical	Within	Across
Travel time (log)	-0.450***	-0.146**	0.176	-0.150**	-0.168**	-0.227***	-0.123***
	(0.109)	(0.069)	(0.113)	(0.074)	(0.067)	(0.046)	(0.068)
Patents applied (log)	1.142***	1.092***	1.081***	0.978***	1.190***	1.050***	0.988***
	(0.005)	(0.015)	(0.022)	(0.015)	(0.015)	(0.014)	(0.015)
Patent stock (log)	0.643***	0.630***	0.697***	0.626***	0.736***	0.611***	0.559***
	(0.028)	(0.022)	(0.030)	(0.026)	(0.031)	(0.022)	(0.022)
Technological distance	-0.581***	-0.591***	-0.387***	-0.237***	-0.367***	-0.858***	-0.971***
	(0.069)	(0.068)	(0.083)	(0.070)	(0.069)	(0.013)	(0.053)
MSA-pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log likelihood	-430330.55	-438977.47	-323018.69	-459107.53	-391332.04	-1189471.7	-761101.95
Number of observations	4,073,586	4,073,586	4,073,586	4,073,586	4,073,586	4,073,586	4,073,586

Standard errors in parentheses.* ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). Coefficients on MSA population and per capita income are available upon request.

makes remote business cooperation much easier. To test such claim, based on Equation 1.3, I interact $ln(travel_{ijt})$ with a set of four dummies, each corresponding to a different time period (1970s, 1980s, 1990s and 2000s). The impacts of travel time reductions seem to be decreasing starting from 1970s to 1990s and bounce back a little in the 2000s. However, if I pick any two of the four coefficients and perform a t-test, none of them are statistically different at 10% level. Therefore, I don't have evidence that the impacts of travel time are heterogeneous over time.

1.8 Conclusion

The claim that geographic proximity facilitates knowledge flows has been found in many contexts. However, most of the evidence in the literature comes from cross-sectional studies, in which endogeneity problems are inevitable. Any unobserved common shock to firms' locality and knowledge flows makes empirical results hard to trust. In this paper, I attempt to address this issue by studying the effects of changes in travel time on inter-MSA patent citations. I find that reductions in travel time have positive impacts on inter-regional patent citation frequency. Besides that, I also find travelers are indifferent toward travel time reductions that save them less than one hour, and the impacts of travel time reductions are the strongest two years later. Most importantly, my results are robust when I control for various types of endogenous shocks, when I examine travel time reductions due to hub openings, and when I examine the dynamic impacts of travel time reductions. Finally, I find the impacts are heterogeneous across technology fields, with the impacts being the strongest on citations within

Table 1.9: Earlier years vs. later years

Variable	(1)
Travel time (log) * 1970s dummy	-0.200***
	(0.046)
Travel time (log) * 1980s dummy	-0.174***
	(0.043)
Travel time (log) * 1990s dummy	-0.155***
	(0.041)
Travel time (log) * 2000s dummy	-0.179***
	(0.042)
Patents applied (log)	0.758***
	(0.015)
Patent stock (log)	0.526***
	(0.020)
Technological distance	-1.021***
	(0.039)
Log likelihood	-1600133.7
Number of observations	4,073,586

Standard errors in parentheses.* ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). Coefficients on MSA population and per capita income are available upon request.

Chemical patents and the weakest within Drug & Medical patents.

Chapter 2

Strategic Agents and Vertical Relationships in Media Markets

2.1 Introduction

Two-sided markets have attracted increasing attention among economists. In a two-sided market, platforms serve two (or multiple) sides, and the utility of an agent joining a platform depends on how many agents on the other side join the same platform. The most studied two-sided markets are probably media markets where the group externality is negative on one side but positive on the other side. For example, in the case of TV (magazine) markets, the two sides of agents are viewers (readers) and advertisers. For advertisers, the value of placing an ad increases when the TV program or magazine reaches more readers (viewers). In contrast, it is often assumed that ads lead to nuisance cost for viewers (readers), so having more advertisers on the same platform lowers their utility.

Earlier studies on two-sided markets have analyzed various pricing issues (fixed fee vs per-unit price, free vs. paid etc.) and market features (single-homing vs. multi-homing).[1] There are two common assumptions in these studies. First, all agents are *passive* in the sense that their only decision is which platform(s) to join (participation decision). Second, platforms serve the agents directly (no middleman). Both assumptions may be violated in practice. In media markets, advertisers typically are not passive, and need to make strategic decisions such as choosing ad quality. This is because, advertisers are interested in not only how many readers will see their ads but also how effective their ads will be in influencing sales. The success of ads depend heavily on "the actual content of your commercial, the production quality".[2] And production quality comes at a cost. According to Chron, production costs of a 30-second commerical in 2008 range in price from free, to \$200-\$1,500 produced by local television statsions, to \$342,000 on average produced by an advertising agency for national commercial.[3] Second, platforms may not directly sell to the agents. For example, magazines may be sold to retailers first, who then resell the magazines to final readers. Similarly, ESPN sells its content to cable providers who as middlemen then sell to viewers.

In this paper, we first consider the case of strategic agents (e.g., advertisers who make strategic decisions such as ad quality). We identify two complications if the strategic feature of advertisers is ignored. First, wrong estimates of group externality parameters will be obtained. Advertisers'

[1] See, for example, Caillaud and Jullien (2003); Armstrong (2006) and Choi (2006).

[2] "How Much Does TV Commercial Production Cost?", Kelly McCaughey, August 8, 2016, http://www.greyskyfilms.com/tv-commercial-production-cost/. Accessed on January 25, 2017.

[3] "How Much Does Television Advertising Really Cost?," by Nancy Wagner, http://smallbusiness.chron.com/much-television-advertising-really-cost-58718.html. Accessed on January 25, 2017.

optimal ad quality choice depends on the anticipated market share of the platform on the reader side. Once this is taken into account, the econometric model will differ qualitatively from the one under passive agents. Therefore, the passive agents model will lead to model specification errors, and in turn wrong estimates.

Second, the equilibrium prices will differ. In particular, prices on both sides of the market will be upward biased so taking into account the strategic feature of advertisers lowers equilibrium prices on both sides. On the reader side, competition is now more intense – more readers affect advertiser utility not only directly, but also indirectly through its impact on ad quality. Price must be lower with more intense competition. On the advertiser side, endogenous ad quality essentially reduces the magnitude of group externality parameter on the advertiser side. Since this group externality parameter is negative, it raises equilibrium prices. Correspondingly, a reduction in the magnitude of group externality parameter leads to a smaller increase of price, relative to the case of passive agents.

We also analyze the case of vertical relationship where there are middlemen between platforms and agents on the reader side.[4] That is, platforms sell to (dedicated) retailers at wholesale prices, and retailers then sell to readers at retail prices. Our results show that ignoring this vertical structure has no impact on estimating the group externality parameters. In particular, even if one uses wholesale prices rather than retail prices on the

[4]Many two-sided markets have middlemen. In our setting, readers may buy magazines from a retailer rather than the publisher (manufacturer) directly. In terms of credit cards, a consumer typically receives credit card from an issuer rather than a credit card agency. For example, consumers may receive Visa cards not from Visa, but from airlines, banks or hotels etc.

reader side, one can still obtain the correct group externality parameters. We also solve for equilibrium prices and find that in general equilibrium price on either side depends on both sides' group externality parameters. This is in sharp contrast to findings in standard two-sided market models where equilibrium price on either side depends on the group externality on the other side only. Interestingly, we consider the case where each platform is split into two independent divisions who maximizes its own profit. We find that the two sides would command the same equilibrium price, which involves the product of the group externality parameters of both sides.

To our knowledge this is the first paper concerned with these two modifications of strategic agents and vertical relationship. For example, in two of the most influential papers on two-sided markets, Caillaud and Jullien (2003) and Armstrong (2006), platforms are allowed to choose different prices across groups but not within a group, similar to our paper. They do not consider strategic agents or vertical relationship, but analyze issues such as single-homing vs. multi-homing.

Our paper is most closely related to the literature estimating two-sided market models. Most look at media markets where one of the two sides is advertisers,. For example, see Kaiser and Wright (2006) and Kaiser and Song (2009) for magazine market, Argentesi and Filistrucchi (2007) for newspaper market and Wilbur (2008) for TV advertising.[5] Kaiser and Wright use a model similar to Armstrong (2006) with two platforms and

[5]There are also earlier studies estimating networks effects, for example, Rysman (2004). There, advertisers choose the size of their ad. But this is more like the platform chooses multiple qualities (versions) or their product and let the advertisers self-select. This is qualitatively different from our strategic agent case where advertisers determine how much to invest in making their ad, choosing ad quality.

specific consumer distribution (Hotelling model). They adapt the model and derive the demand on both sides of market. They also construct various instrument variables to deal with endogeneity issues. Our benchmark model largely follows that in Kaiser and Wright with some simplification (see Section 2.2 for more details). Our focus is on the two modifications of strategic agents and vertical relationship, and how they affect estimation and comparative statics. Neither of the aforementioned studies considers any of these two modifications.

There are also various theory studies analyzing media markets. See, for example, Anderson and Coate (2005), Kind et al. (2007) and Reisinger (2012). Similar to the empirical literature mentioned above, in these studies, agents only make platform participation decisions (passive agents) and platforms directly sell to agents (no middleman). Advertisers in our model viewing the competing platforms as differentiated, but they all value consumers the same. In contrast, in Athey et al. (2013), advertisers have heterogeneous valuations for reaching consumers.[6]

The rest of the paper is organized as follows. We present and analyze the benchmark model in Section 2.2. Section 2.3 considers the case of strategic agents where agents on one side (advertisers) make investment decisions which directly affect the utility level of readers joining the same platform. In Section 2.4 we analyze a vertical structure where platforms sell to middlemen (e.g. retailers) instead of selling to agents directly. We conclude in Section 2.5. Proofs of lemmas and propositions can be found

[6]This is also assumed in various other studies. See Anderson and Jullien (2015) for a survey of this literature. Note that while advertisers may have different valuations for reaching viewers, each viewer values all advertisers the same. In contrasts, in our model, a viewer values different ads differently, depending on the quality of these ads.

in the appendix.

2.2 The benchmark model

Our benchmark model follows Kaiser and Wright (2006) closely. Two platforms (e.g., magazines) $i = 1, 2$, are located at the two end points of a Hotelling line with platform 1 located at 0. Platforms serve two groups of agents: readers (denoted by superscript r) and advertisers (denoted by superscript a). There is a continuum of each group of agents (with mass 1), uniformly distributed on the Hotelling line. Transport cost is linear in distance traveled and unit transport cost t is the same for the two groups of agents. We further assume $t = \frac{1}{2}$ for simplicity. Consider a reader located at x. If she joins platform 1, she enjoys a utility of

$$u_1^r = \theta^r + \gamma N_1^a - p_1 + \varepsilon_1^r - \frac{1}{2}x.$$

In the above expression, θ^r is the reservation value from consuming the content provided by the platform. p_1 is the price a reader has to pay to joining platform 1 (e.g., buying a copy of its magazine). N_1^a is the number of advertisers who join platform 1. γ measures the externality of ads on readers. We follow the common assumption $\gamma \leq 0$ in the literature studying media markets. That is, going through ads is a nuisance cost which viewers/readers need to bear to access the content provided by the platforms. ε_1^r captures a platform specific shock to all readers' utility which is unobserved with mean zero.

If this reader joins platform 2 instead, her utility will be

$$u_2^r = \theta^r + \gamma N_2^a - p_2 + \varepsilon_2^r - \frac{1}{2}(1-x).$$

Similarly, an advertiser located at x will have the following utilities

$$u_1^a = \theta^a + \rho N_1^r - a_1 + \varepsilon_1^a - \frac{1}{2}x,$$

$$u_2^a = \theta^a + \rho N_2^r - a_2 + \varepsilon_2^a - \frac{1}{2}(1-x),$$

where θ^a is the reservation utility and the group externality parameter is $\rho > 0$.

By looking at the u_i^j expressions, $i = 1, 2$, $j = a, r$, we can see that for agents on either side, only group externality parameter on the same side enters into the utility function, and enters in a linear fashion. For example, only ρ enters into the u_i^a expressions and does so linearly.

The stage game we analyze is the following. In stage 1, platforms choose prices p_i and a_i simultaneously. Observing these prices, readers and advertisers choose which platform to join simultaneously in stage 2.

This is a simplified version of the model in Kaiser and Wright (2006), with $\theta_i^j = \theta^j$, $i = 1, 2$, $j \in \{r, a\}$. We also assume that the coefficients in front of a_i and p_i are 1 (marginal utility of income is normalized to 1). The marginal advertiser x^a, who is indifferent between joining either platform, can be derived as

$$u_1^a = u_2^a \Rightarrow x^a = \frac{1}{2} + \rho(N_1^r - N_2^r) - (a_1 - a_2) + (\varepsilon_1^a - \varepsilon_2^a). \tag{2.1}$$

The number of advertisers joining either platforms is given by

$$N_1^a = x^a, \quad N_2^a = 1 - x^a.$$

Similarly we can obtain the marginal reader

$$x^r = \frac{1}{2} + \gamma(N_1^a - N_2^a) - (p_1 - p_2) + (\varepsilon_1^r - \varepsilon_2^r), \tag{2.2}$$

and

$$N_1^r = x^r, \quad N_2^r = 1 - x^r.$$

Estimating group externality parameters:

Let $n_1^a = \frac{N_1^a}{N_1^a + N_2^a}$ and $n_1^r = \frac{N_1^r}{N_1^r + N_2^r}$. Since $N_i^j + N_{-i}^j = 1$ (unit mass of agents on either side), we have

$$n_1^a = \frac{1}{2} + \rho(N_1^r - N_2^r) - (a_1 - a_2) + \varepsilon_1^a - \varepsilon_2^a, \tag{2.3}$$

$$n_1^r = \frac{1}{2} + \gamma(N_1^a - N_2^a) - (p_1 - p_2) + \varepsilon_1^r - \varepsilon_2^r. \tag{2.4}$$

With data on sales N_i^j, market share n_i^j and prices (p_i and a_i), these two equations can be used to estimate group externality parameters ρ and γ respectively.[7]

Equilibrium prices and comparative statics:

Setting $\varepsilon_i^j = 0$ (their mean value), platforms' profit maximization prob-

[7]Prices and sales on the right hand side are likely to be endogenous. Kaiser and Wright propose a series of instruments to solve this problem.

lems are,

$$\max_{p_i,a_i} \quad \pi_i = a_i N_i^a + p_i N_i^r, \quad i = 1, 2.$$

Solving the FOCs, we can obtain equilibrium prices and profit as the following:

$$p_i = -\rho + \frac{1}{2}, \quad a_i = -\gamma + \frac{1}{2}, \quad \pi_i = -\frac{1}{2}\gamma + \frac{1}{2} - \frac{1}{2}\rho. \tag{2.5}$$

Taking derivatives, we have

$$\frac{\partial p_i}{\partial \gamma} = 0, \quad \frac{\partial p_i}{\partial \rho} = -1; \quad \frac{\partial a_i}{\partial \gamma} = -1, \quad \frac{\partial a_i}{\partial \rho} = 0. \tag{2.6}$$

That is, own derivatives ($\frac{\partial p_i}{\partial \gamma}$ and $\frac{\partial a_i}{\partial \rho}$) are always zero and cross derivatives ($\frac{\partial p_i}{\partial \rho}$ and $\frac{\partial a_i}{\partial \gamma}$) are always -1 (similar to the results in Armstrong (2006)).

All these results are straight from Kaiser and Wright (2006). Next, we introduce two new modifications to the model, one at a time. First, we allow agents on one side to make a strategic choice, which affects their own utility as well as the utility of agents on the other side. Second, we allow a vertical structure on one side with an intermediary between the platform and the agents. For example the magazines may be sold to retailers first which then sell the magazines to readers.

2.3 Strategic agents

The benchmark model assumes that all agents are passive, i.e., they only determine which platform to join. What if some agents make strategic decision (beyond the simple participation decisions), which in turn affect the well being of agents on the other side and the platforms?

Consider advertisers for example. Suppose that in addition to platform participation decisions, they also decide what types of ads to show to readers on a platform. For simplicity, we characterize the type of ad with single parameter $\kappa \geq 0$ representing of the quality of ad, and advertisers' strategic decision becomes ad quality decision. Better quality ad benefits both readers (e.g., more fun, less nuisance) and advertisers (e.g., attract more attention from readers). But investment is also costly, entailing a fixed investment cost of $c(\kappa)$, where $c'(\kappa) > 0$ and $c''(\kappa) > 0$. For tractability, we assume that ad cost is a fixed cost in the common quadratic form $c(\kappa) = \frac{1}{2}\kappa^2$.

With investment, a representative advertiser's utility of joining platform 1 becomes

$$u_1^a = \theta^a + f(\rho, \kappa, N_1^r) - a_1 + \varepsilon_1^a - \frac{1}{2}x - \frac{\kappa_1^2}{2},$$

where $f(\rho, \kappa, N_1^r)$ increases with ρ, κ and N_1^r. For simplicity, we assume that

$$f(\rho, \kappa, N_1^r) = \rho(1+\kappa)N_1^r.$$

Note that if the advertiser does not invest in ad quality ($\kappa = 0$), then we recover the utility in the benchmark model. The advertiser's utility of joining either platform is,

$$\begin{aligned} u_1^a &= \theta^a + \rho(1+\kappa)N_1^r - a_1 + \varepsilon_1^a - \frac{1}{2}x - \frac{\kappa_1^2}{2}, \\ u_2^a &= \theta^a + \rho(1+\kappa)N_2^r - a_2 + \varepsilon_2^a - \frac{1}{2}(1-x) - \frac{\kappa_2^2}{2}. \end{aligned} \tag{2.7}$$

For either platform the advertiser joins, it chooses κ optimally to maximize u_i^a. Solving $\frac{\partial u_i^a}{\partial \kappa} = 0$, we obtain the results in the following Lemma.

Lemma 1. *Optimal ad quality is given by $\kappa_i^* = \rho N_i^r$ where N_i^r is platform $i = 1, 2$'s expected market share on the reader side.*

Note that all advertisers joining the same platform will choose the same investment level. Correspondingly, the average ad quality is also κ_i for platform i.

2.3.1 Estimating group externality parameters

Having derived advertiser's optimal investment decisions, next we investigate how such investment decisions affect the estimation of group externality parameters.

Substituting $\kappa_i^* = \rho N_i^r$ into advertiser's utility functions, we can obtain

$$\begin{aligned} u_1^a &= \theta^a + \rho N_1^r + (\rho N_1^r)^2 - a_1 + \varepsilon_1^a - \frac{1}{2}x - \frac{1}{2}(\rho N_1^r)^2 \\ &= \theta^a + \rho N_1^r + \frac{\rho^2}{2}(N_1^r)^2 - a_1 + \varepsilon_1^a - \frac{1}{2}x; \\ u_2^a &= \theta^a + \rho N_2^r + \frac{\rho^2}{2}(N_2^r)^2 - a_2 + \varepsilon_2^a - \frac{1}{2}(1-x). \end{aligned} \tag{2.8}$$

On the advertiser side, it remains that only the own group externality parameter ρ enters into the utility function, except that it now enters in a nonlinear fashion.

Similar to the benchmark model, we derive the marginal advertiser x^a and the number of advertisers joining either platform N_i^a. We can then obtain

$$n_1^a = \frac{1}{2} + \rho(N_1^r - N_2^r) + \underbrace{\frac{\rho^2}{2}\left[(N_1^r)^2 - (N_2^r)^2\right]}_{\text{additional term}} - (a_1 - a_2) + \varepsilon_1^a - \varepsilon_2^a, \quad (2.9)$$

where $n_1^a = \frac{N_1^a}{N_1^a + N_2^a}$.

This seems different from equation (2.4) in the benchmark model. In particular, advertiser demand now is linear-quadratic in ρN_i^r, rather than just linear.[8]

Ad quality affects reader utility as well. This may be because readers enjoy more an ad with higher quality, or a higher quality ad may be more informative which allows readers to make better purchasing decisions. Taking this into account, a representative reader's utility becomes,

$$u_i^r = \theta^r + g(\gamma, \bar{\kappa}_i, N_i^a) - p_i + \varepsilon_i^a - \frac{1}{2}x,$$

where $\bar{\kappa}_i$ is the average ad quality on platform i, and $g(\gamma, \bar{\kappa}_i, N_i^a)$ increases with γ and $\bar{\kappa}_i$ but decreases with N_i^a.

For simplicity, assume that

$$g(\gamma, \bar{\kappa}_i, N_i^a) = \gamma(1 - \bar{\kappa}_i)N_i^a, \quad i = 1, 2.$$

Because all advertisers on the same platform make the same investment, we have $\bar{\kappa}_i = \kappa_i^*$. A representative reader's utility of joining either platform

[8]This can also be seen with general $f(\rho, \kappa_i, N_i^r)$ and $c(\kappa)$, since κ^* in general will be a function of N_i^r. Once κ_i^* is substituted, then n_i^a will not be linear in N_i^r anymore.

becomes,

$$
\begin{aligned}
u_1^r &= \theta^r + \gamma(1-\kappa_1^*)N_1^a - p_1 + \varepsilon_1^r - \frac{1}{2}x, \\
&= \theta^r + \gamma(1-\rho N_1^r)N_1^a - p_1 + \varepsilon_1^r - \frac{1}{2}x; \\
u_2^r &= \theta^r + \gamma(1-\rho N_2^r)N_2^a - p_2 + \varepsilon_2^r - \frac{1}{2}(1-x). \qquad (2.10)
\end{aligned}
$$

From equation (2.10), we can see that on the reader side, now group externality parameters of both sides (ρ and γ) enter into their utility functions. In particular, it contains the same linear term γN_i^a, but also an interaction term $\rho\gamma N_i^r N_i^a$.

Using the new u_i^r expressions, we can derive the marginal readers x^r and N_i^r, which then lead to

$$
\begin{aligned}
n_1^r &= \frac{1}{2} + \gamma[(1-\kappa_1^*)N_1^a - (1-\kappa_2^*)N_2^a)] - (p_1 - p_2) + \varepsilon_1^r - \varepsilon_2^r \\
&= \frac{1}{2} + \gamma(N_1^a - N_2^a) - \underbrace{\gamma \cdot \rho \cdot (N_1^r \cdot N_1^a - N_2^r \cdot N_2^a)}_{\text{additional term}} \\
&\quad -(p_1 - p_2) + \varepsilon_1^r - \varepsilon_2^r. \qquad (2.11)
\end{aligned}
$$

It is also different from (2.3) of the benchmark model. Combined with the results for estimating ρ using advertisers' side, we have the following results.

Proposition 1. *(Estimating group externality parameters) Ignoring advertisers' strategic investment decisions will lead to wrong econometric models for estimating the group externality parameters on both the advertiser and reader sides.*

If the underlying model features strategic agents yet it is not modeled, then the consequent econometric models are likely to be wrong as we have shown, leading to inaccurate estimates.

2.3.2 Equilibrium prices and comparative statics

Setting $\varepsilon_i^j = 0$, $i = 1, 2$, $j = a, r$, we can solve platforms' FOCs and obtain equilibrium prices and profits, as given in the next Proposition.

Proposition 2. *(Strategic agents) When advertisers make strategic investment in ad qualities, the unique SPNE is characterized by*

$$p_i = \frac{1}{2} - \rho + \frac{1}{2}\rho(\gamma - \rho), \quad a_i = \frac{1}{2} - \gamma + \frac{1}{2}\rho\gamma,$$

$$\pi_i = \frac{1}{2}\gamma\rho - \frac{1}{2}\gamma + \frac{1}{2} - \frac{1}{2}\rho - \frac{1}{4}\rho^2, \quad i = 1, 2.$$

Proof. See the Appendix. □

Using the equilibrium price expressions, one can easily verify that

$$\begin{aligned} \text{Own derivatives} \quad &: \quad \frac{\partial p_i}{\partial \gamma} = \frac{\rho}{2} > 0, \quad \frac{\partial a_i}{\partial \rho} = \frac{\gamma}{2} < 0; \\ \text{Cross derivatives} \quad &: \quad \frac{\partial p_i}{\partial \rho} = -\rho + \frac{\gamma}{2} - 1 < -1, \quad \frac{\partial a_i}{\partial \gamma} = \frac{\rho}{2} - 1 \in (-1, 0), \end{aligned}$$

which lead to the following Corollary.

Corollary 1. *(Strategic agents) Different from the benchmark case, with strategic agents, own derivatives ($\frac{\partial p_i}{\partial \gamma}$ and $\frac{\partial a_i}{\partial \rho}$) are not zero and the cross derivatives ($\frac{\partial p_i}{\partial \rho}$ and $\frac{\partial a_i}{\partial \gamma}$) are not -1.*

In standard two-sided markets, the own partial derivatives are always zero and cross partial derivatives are always -1. This is not true anymore under strategic agents. Why? The answer lies in the different forms of utility functions. First, note that ρ enters into u_i^a expressions nonlinearly. Second, both γ and ρ enters into the u_i^r expressions.[9]

We can also explore welfare impacts of strategic agents. The results are summarized in the next Proposition.

Proposition 3. *Relative to the benchmark case of passive agents, in the strategic agents case:*

(i) Equilibrium prices on both sides are lower;

(ii) Platforms are worse off while advertisers and readers are better off.

Proof. See the Appendix. □

(i) can be easily verified and is rather counterintuitive. After all, advertisers' investment in ad quality helps themselves and readers. Let us see why. When platforms choose prices on the reader side, there is more to lose when advertisers are strategic – ρ affects u_i^a not only directly, but also indirectly by affecting advertiser investment. This gives each platform more incentive to attract readers which intensifies competition at the reader side and leads to lower prices. Prices on the advertisers side are determined by group externality at the reader side. We can see it from reader's utility

[9]To see this, we consider a hypothetical situation where advertiser's ad quality affects their own utilities but not readers' utilities. We find the equilibrium prices to be $p_1 = -1/2\rho^2 - \rho + 1/2$ and $a_1 = 1/2 - \gamma$. Note that own derivatives are still zero, since only group externality parameter of the same side enters into an agent's utility function. However, ρ now enters into p_1 nonlinearly, because ρ enters into u_i^a expressions nonlinearly.

function,

$$\begin{aligned} u_1^r &= \theta^r + \gamma(1-\kappa_1)N_1^a - p_1 + \varepsilon_1^r - \frac{1}{2}x \\ &= \theta^r + \tilde{\gamma}N_1^a - p_1 + \varepsilon_1^r - \frac{1}{2}x, \end{aligned}$$

where $\tilde{\gamma} \equiv (1-\kappa_1^*)\gamma$. It look just like the u_1^r except γ is replaced by $\tilde{\gamma}$. Correspondingly, on the advertiser side, platforms will raise prices by $|\tilde{\gamma}|$. Prices on the advertiser side are also lower than those in the benchmark model since

$$|\tilde{\gamma}| = |(1-\kappa_1^*)\gamma| = \left|\left(1-\rho\cdot\frac{1}{2}\right)\gamma\right| < |\gamma|.$$

With lower prices on both sides, platforms must be worse off. However, advertisers and readers are better off because they both enjoy lower prices. Additionally, advertisers can mimic the benchmark case by choosing zero investment. The fact that they choose positive investment must make them strictly better off. For readers, investment by advertisers directly raises their utilities, everything else the same. Combined with lower prices, readers must be better off as well.

To summarize, if the strategic agent aspect is ignored, we find that: (i) Estimates of the group externality parameters will be biased; (ii) Equilibrium prices are upward biased, and their derivatives with respect to group externality parameters differ; (iii) Platform profits are over-reported while advertiser and reader surplus are under reported.

2.4 Vertical relationship

In the previous section, the deviation from standard two-sided market models is strategic agents on one side. In this section, we go back to passive agents who do not make strategic decisions. Instead, we introduce an extra layer on one side of the market. On the reader side, we assume that platforms do not directly sell to readers, but rather through independent retailers. In particular, there are two (dedicated) retailers, with retailer $i = 1, 2$ serving platform $i = 1, 2$ only.[10]

Representative advertiser and reader's utilities from joining either platform are the same as in the benchmark model,

$$u_i^a = \theta^a + \rho N_i^r - a_i + \varepsilon_i^a - t_i(x), \quad u_i^r = \theta^r + \gamma N_i^a - p_i + \varepsilon_i^r - t_i(x), \quad i = 1, 2,$$

where $t_1(x) = \frac{1}{2}x$ and $t_2(x) = \frac{1}{2}(1-x)$. Note that retail prices are now chosen by retailers, not platforms.

Let w_i denote platform i's wholesale price charged to retailer i. Retailer i's profit maximization problem is

$$\max_{p_i} \quad \pi_i^R = (p_i - w_i)N_i^r, \quad i = 1, 2.$$

[10] There are alternative ways to model the retail sector. For example, there may be a single retailer serving both platforms. Or there are may be two undedicated retailers, each serving both platforms. The former case introduce double marginalization, but also the problem of sharing a single downstream distribution. In the latter case, one will need to introduce retailer differentiation on top of platform differentiation. Our specification of two dedicated retailers introduces double marginalization only.

Platform i's profit is

$$\pi_i^M = w_i N_i^r + a_i N_i^a, \quad i = 1, 2.$$

The stage game is as follows. In stage 1, platforms choose wholesale prices (w_i) simultaneously. In stage 2, retail prices on both sides (p_i and a_i) are chosen simultaneously. In stage 3, readers and advertisers decide what platforms to join.

2.4.1 Estimating group externality parameters

First note that the agents' utility functions are the same as in the benchmark model case, we still have

$$n_1^a = \frac{1}{2} + \rho(N_1^r - N_2^r) - (a_1 - a_2) + \varepsilon_1^a - \varepsilon_2^a,$$

$$n_1^r = \frac{1}{2} + \gamma(N_1^a - N_2^a) - (p_1 - p_2) + \varepsilon_1^r - \varepsilon_2^r.$$

If one uses final retail prices (p_i and a_i), having a vertical structure would have no impact on the correct estimation of the group externality parameters γ and ρ.

We solve the game backwards, starting with stage 2 where platforms and retailers maximize their respective profits given w_i:

$$\max_{p_i} \ \pi_i^R, \quad \max_{a_i} \ \pi_i^M.$$

Solving the retailers' FOC, we can obtain $p_i(w_1, w_2)$. They lead to[11]

$$p_1 - p_2 = \frac{(8\gamma^2 + 3)(w_1 - w_2)}{9 - 16\gamma\rho}. \tag{2.12}$$

What if one uses wholesale prices (w_i) instead of the retail prices (p_i) on the reader side to estimate γ? To see this, we substitute the p_i expressions as functions of w_1 and w_2, into the n_1^r expression above. We can obtain

$$\begin{aligned} n_1^r &= \frac{1}{2} + \gamma(N_1^a - N_2^a) - (p_1 - p_2) + \varepsilon_1^r - \varepsilon_2^r. \\ &= \frac{1}{2} + \gamma(N_1^a - N_2^a) - \frac{(8\gamma^2 + 3)(w_1 - w_2)}{9 - 16\gamma\rho} + \varepsilon_1^r - \varepsilon_2^r. \end{aligned}$$

If one ignores the endogeneity of wholesale prices, then group externality parameters can be correctly estimated even with wholesale prices.

2.4.2 Equilibrium prices and comparative statics

In the previous section, we have solved for the optimal p_i from retailers' FOCs. From platforms' FOC, we can obtain $a_i(w_1, w_2)$. Substituting the optimal p_i and a_i expressions into π_i^M, we can solve for the optimal wholesale prices which can be substituted back to obtain the retail prices p_i and a_i.

The results are presented in the next Proposition.

Proposition 4. *(Vertical structure) With retailers on the readers side, equilibrium prices are given by:*

$$w_i = \frac{9 - 4\rho - 6\gamma - 16\rho\gamma}{16\gamma^2 + 6}, \tag{2.13}$$

[11] More details are provided in Proof of Proposition 4.

$$p_i = w_i + \left(\frac{1}{2} - 2\rho\gamma\right), \tag{2.14}$$

$$a_i = -2\gamma w_1 + \left(\frac{1}{2} - 2\rho\gamma\right). \tag{2.15}$$

Proof. See the Appendix. □

In the equilibrium, platforms will choose $w_i > 0$. However, if we substitute $w_i = 0$ into the optimal retail price functions, we get the same retail prices as substituting $w_i = 0$ into equations (2.14)-(2.15). That is, $p_i = a_i = \frac{1}{2} - 2 \cdot \rho \cdot \gamma$. In this case, the platforms make profits from the advertiser side only while independent retailers profit from the reader side only. Platform and retailers maximize their own profits, and do not take into account theirs impacts on the other side. This is the same as the case where a platform splits itself into two independent divisions (presented in the next Corollary), one for each side, and each division maximizes its own profit. Thus we have the following results,

Corollary 2. *(Independent division) If each platform is split into two independent divisions, one for each side, then in the unique equilibrium, prices are given by*

$$p_i = a_i = \frac{1}{2} - 2\rho\gamma.$$

In the standard two-sided markets, price on each side depends on the other side's group externality, not on its own side's group externality. As a result, price in the two sides differ from each other ($p_i \neq a_i$). In contrast, under independent division, prices on the two sides are the same and depend on group externalities of both sides. Let us see why. Let π_r^1 and π_a^1 denote the profit of platform 1's reader and advertiser division respectively,

$$\pi_1^r = p_1 n^r, \quad \pi_1^a = a_1 n^a.$$

Their respectively FOCs are

$$\frac{\partial \pi_1^r}{\partial p_1} = n^r + p_1 \frac{\partial n^r}{\partial p_1} = 0, \quad \frac{\partial \pi_1^a}{\partial a_1} = n^a + a_1 \frac{\partial n^a}{\partial a_1} = 0.$$

It can be easily verified that

$$\frac{\partial n^r}{\partial p_1} = \frac{\partial n^a}{\partial a_1} = -\frac{1}{1 - 4\gamma\rho}.$$

In the equilibrium, $n^r = n^a = \frac{1}{2}$. It must be that

$$p_1 = a_1 = -\frac{1}{2 \cdot \frac{\partial n^r}{\partial p_1}} = \frac{1}{2} - 2\rho\gamma.$$

In standard two-sided markets, platform 1 maximizes its joint profit from the two sides,

$$\pi_1 = p_1 n^r + a_1 n^a.$$

Profit maximization requires

$$\frac{\partial \pi_1}{\partial p_1} = n^r + p_1 \frac{\partial n^r}{\partial p_1} + \underbrace{a_1 \frac{\partial n^a}{\partial p_1}}_{\text{cross impact}} = 0,$$

$$\frac{\partial \pi_1}{\partial a_1} = n^a + a_1 \frac{\partial n^a}{\partial a_1} + \underbrace{p_1 \frac{\partial n^r}{\partial a_1}}_{\text{cross impact}} = 0.$$

The platform internalizes the impact of price change in one side on the profit from the other side (cross impacts), so the cross derivatives enter into FOCs. It can be easily verified that the two cross derivatives are unequal,

$$\frac{\partial n^r}{\partial a_1} = -\frac{2\gamma}{1-4\gamma\rho} \neq \frac{\partial n^a}{\partial p_1} = -\frac{2\rho}{1-4\gamma\rho}.$$

This leads to different prices at the two sides ($p_i \neq a_i$).[12]

Using the equilibrium prices in equations (2.14) and (2.15), one can easily verify the results in the next Corollary.

Corollary 3. *(Vertical relationship) Different from the benchmark case, own derivatives ($\frac{\partial p_i}{\partial \gamma}$ and $\frac{\partial a_i}{\partial \rho}$) are not zero, and the cross derivative $\frac{\partial p_i}{\partial \rho}$ in general are not -1.*

Even with $w = 0$, final price on either side depends on the product of the two group externality parameters. On top of that, optimal w_i is a function both group externality parameters and affects the prices on the reader side further. Together they led to the results in Corollary 3.

2.5 Conclusion

This paper considers, one at a time, two modifications to the standard two-sided market models. In the first modification, agents one side of the market (advertisers) make strategic choices on the ad quality, which affects their own utility as well as utilities of readers (i.e., agents on the other side) joining the same platform. We find that having strategic agents leads to

[12]The cross partial derivatives must offset the own derivatives in a way so that the eventual equilibrium price on either side does not depend on the group externality of that side.

qualitative different demand systems for the estimation of group externality parameters. As a result, if the strategic agent feature is not properly accounted for, one would obtain wrong estimates of the group externality parameters. We also solve for equilibrium prices and find that under strategic agents, equilibrium prices on both sides are lower than when agents are passive, and they depend on group externality parameters of both sides in general.

In the second modification, we introduce independent retailers between platforms and readers. We find that this modification has no impact on estimating group externality parameters, even if one is to use the wholesale prices instead of final retail prices. However, the equilibrium prices in general depend on group externality parameters at both sides of the market. One particularly interesting finding is for the case where each platform is split into two independent divisions. We recover a common equilibrium price, charged by all divisions. This common price depends on the product of the group externality parameters at the two sides. This is in sharp contrast to the standard two-sided models (e.g., Armstrong (2006)) where prices differ across the two sides, since price on each side depends only on the group externality parameter of the other side.

While this study focuses on the effects of these two modifications on parameter estimation and equilibrium pricing, future studies can investigate various other topics with these two modifications. These include single-homing vs. multi-homing, price discrimination, merger impacts etc. Various studies have shown that policies that work well in one-sided markets may not have the desirable effects when applied to two-sided markets. The two modifications considered in this paper will add extra complexity.

These changes in the assumption of the model can bring even more complications when a policy is applied in a two-sided market.

Chapter 3

Vertical Relations and Corporate Innovation

3.1 Introduction

Since knowledge is a public good, a few agents conducting research and development (R&D) may benefit other unintentionally (Arrow, 1962). A number of empirical works have supported this claim, indicating positive externalities of innovation activities across countries(Coe and Helpman, 1995; Bernstein and Mohnen, 1998; Keller, 2002), and across firms within the same industry (Bloom et al., 2013; Badinger and Egger, 2016). The study of knowledge spillovers is important because if one's firms research can positively affect other firms' performance, then the social return on R&D investment will be greater than the private return on R&D investment. Therefore, from a welfare perspective, private R&D may be under-invested.

However, what we know less about is the knowledge spillovers across industries. Industries are indeed interconnected. A product in one industry may rely on the technology of various other industries. For example, the quality of a cell phone, depends on the quality of its various components, such as the camera, the display panel, the processor, etc., most of which are not manufactured by the cell phone company itself, but purchased from upstream industries, which leads to an interesting research question: how do R&D activities in upstream/downstream industries affect downstream/upstream industries? For example, if the display panel industry increase its R&D expenditures, how will it affect cell phone companies' decisions and performance?

In this paper, I first establish the vertical relations between industries in intermediate good markets using the industry-level input-output account data, which contains information on intermediate good flows across industries. For each industry, I identify its upstream industries and downstream industries. Nevertheless, most firms operate in multiple industries. Therefore, I use each firm's sales distribution across industries as weights to calculate its weighted average upstream and downstream industries' R&D expenditures.

One novelty about this paper is that I also use patent citations to establish vertical relations between technology classes in the patent system. Production requires not only intermediate goods, but also knowledge, and patent citations are a good indicator of the knowledge flows (Jaffe et al., 1993). For example, Intel cites a lot of Microsoft's patents, because they can improve their chips by learning from how their chips perform in Microsoft operation systems, even though Microsoft is never a supplier of Intel. Similarly, for each patent class, I identify its upstream classes and downstream classes using the citation data of US patents. And for each firm, given its patent portfolio, I also calculate the R&D expenditures of its upstream and downstream industries (technology classes)[1], defined using the patent citation data.

Finally, for each firm, I estimate how innovations in upstream and downstream industries affect its R&D expenditure and market value. Since my independent variables are R&D expenditures in other industries, endogeneity isn't a major issue. However, it might still be possible that innovations in upstream and downstream industries are correlated to some other technology shocks. To address this potential concern, I use R&D capital costs calculated from federal and state R&D tax credit (Wilson, 2009) as an instrumental variable. I discuss the validity of this instrument in Section 4.2.

Using data on a panel of 2,803 public firms in the US from 1976 to 2008, my results show that in intermediate good markets, a 1% increase in research and development (R&D) expenditures in upstream industries will decrease a downstream firm's R&D expenditure by 0.729%, while a 1% increase in R&D of the downstream will reduce a upstream firm's market value by 0.907%. Meanwhile, in the patent system, the upstream to downstream R&D elasticity is estimated to be 0.983, while the upstream to downstream market value elasticity is 1.053.

To my best knowledge, all of the studies in the related literature either focus solely on intra-industry spillovers, or don't explicitly identify the vertical relations between industries.[2]

[1]Industries and technology classes are not the same thing. In this paper, industries are defined using 2002 North America Classification System (NAICS), while patent classes are managed by the US Patent and Trademark Office (USPTO). Indeed, there are some connections between the two systems. Some efforts have been dedicated to assigning each patent class to its NAICS concordance.

[2]For examples, Bloom et al. (2013) identifies the knowledge spillovers from related

The contribution of this paper is threefold. First, instead of studying technology spillovers from *related* industries (usually measured with weighted average R&D stock of firms), this paper identifies the exact vertical relations between industries and study how innovations in upstream (downstream) industries affect firms in downstream (upstream). Second, since both intermediate goods and knowledge are important inputs for production, I distinguish two different types of vertical relations: vertical relations in intermediate good markets and vertical relations in the patent system. Third, I development a novel method to use industrial relations as weights to calculate the R&D expenditures in upstream and downstream for each firm.

The remainder of this paper is organized in as follows. Section 3.2 describes theoretical framework. Section 3.3 presents the data. Section 3.4 discusses the empirical strategy. Section 3.5 presents the results and Section 6 concludes.

3.2 Vertical Relations and Innovation: Channels

A large body of literature has studied vertical relations in the field of industrial organization. However, few studies focus on how innovations in upstream/downstream industries affect firms in downstream/upstream industries. Here I highlight three main channels through which the impacts of innovations transmit vertically across industries: the bargaining power channel, the quality-improving R&D channel, and the cost-reducing R&D channel. I discuss how the mechanism works for firms vertically related in

industries using a symmetric measure of technology similarity.

intermediate good market and the patent system. In the appendix, I also present a model of strategic R&D of vertically related firms.

3.2.1 Intermediate good markets

First, if markets for intermediate inputs are imperfectly competitive, when there is a positive technology shock to upstream industries, it increases the bargaining power of upstream firms when they sell intermediate inputs to downstream firms. Assume intermediate input prices are positively correlated with the bargaining power of upstream firms, then innovations in upstream should raise the production costs in downstream. Similarly, innovations in downstream industries increase the bargaining power of downstream firms, which enables them to lower the prices of intermediate inputs sold by upstream firms. In short, the bargaining power channel suggests that innovations in upstream/downstream industries should negatively affect market values of firms in the corresponding downstream/upstream industries by increasing their production costs.

Second, innovations in upstream/downstream industries might increase the quality of final products. Assume production costs don't go up accordingly, an increase in the quality of goods should increase the demand for the final products, which will in turn lead to an increase in demand for intermediate inputs. To summarize, the quality-improving R&D channel suggests that innovations in upstream/downstream should increase the market values of downstream/upstream firms by increase the demand for their products.

Third, if innovations reduce the costs of production in upstream, it will lower the price of intermediate inputs, which should subsequently benefit the downstream firms. If innovations reduce production costs of downstream firms, it will increase the supply of the final products, which will subsequently lead to an increase in quantity sold, which will increase the demand for intermediate inputs. In short, the cost-reducing R&D channel suggests that cost-reducing innovations in upstream/downstream industries should benefit downstream/upstream firms.

3.2.2 The patent system

How innovations affect firms vertically related in the patent system works a little differently, since knowledge is a public good. If upstream firms are able to charge a license fee to downstream firms for using upstream technology, then all the three channels discussed in the previous section should still work, because in this case knowledge is no longer a public good, similar to intermediate goods. However, if downstream firms can use upstream technology for free, then we have a different story. First of all, the bargaining channel story doesn't hold anymore, since there are no negotiations between upstream and downstream firms. Second, the quality-improving channel partly holds, since downstream benefits from the improvement in the upstream technology. However, innovations in the downstream should have no impacts on upstream because upstream cannot benefit from the improvement in downstream technology, since downstream doesn't purchase technology from upstream. Third, the cost-reducing channel doesn't work because there are no transactions between upstream and downstream and neither can benefit from the cost reductions of the other side.

3.3 Data

3.3.1 Input-Output Data

I use the Annual Input-Output (I-O) Accounts from the Bureau of Economic Analysis (BEA) for the information on the flow of intermediate goods across industries. The data is from 1998 to 2011[3]. Industries are defined according to the 2002 North American Industrial Classification System (NAICS). I calculate the sum of flows of intermediate goods at industry-pair level.

3.3.2 Patent Data

I combine three widely used patent datasets. The main patent dataset is by Li et al. (2014). It includes all patents granted by the United States Patent and Trademark Office (USPTO) from 1975 to 2014. I supplement it with a second dataset by Kogan et al. (2012), who calculates the private economic values of patents of public firms in US granted from 1926 to 2010 using stock price changes on the day the patent was granted. They also identify patent assignees' Compustat PERMNO ID, which will be later used to merge with Compustat firm data.

The USPTO has a highly detailed classification schedule. Each patent belongs to a technology class and subclass within a class. Up to this date, there are more than 450 classes and 150,000 subclasses, which is too many for any type of empirical analysis. Hall et al. (2001) regroup patents classifications into 36 categories that are more friendly for analysis. I merge the three patent datasets using the patent identifier number. The final com-

[3]The table I use is "1998-2001 Summary Make Anuual I-O Table before Redefinitions"

bined patent dataset consists of all utility patent applied by public firms in the US from 1976 to 2008.[4]

3.3.3 Firm Data

Firm data is Compustat data from Wharton Research Data Services (WRDS). I calculate each firm's R&D stock with its R&D expenditure using a perpetual inventory method with a 15% depreciation rate. In Section 5, I also use alternative depreciate rates (10% and 20%) for sensitivity analysis. Firms' market value is the sum of the values of common stock, preferred stock and total debt net of current assets. I use the link variable by Kogan et al. (2012), Compustat PERMNO ID, to merge the firm data with the patent data. My final dataset is composed of 2,803 firms spanning the periods from 1976 to 2008.

3.3.4 Summary Statistics

Table 3.1 presents summary statistics. On average, a firm in my dataset spends \$67,851/year on R&D, and we observe a large standard deviation (\$395,789) relative to mean, which means R&D expenditures vary a lot across firms and years. The average market value of a firm is \$101 million. Total assets have a mean of \$732,850 and a standard deviation of \$3,798,717.

[4]For my study, it is more interesting to look into the application date rather than the grant date of patents, because the former is a better proxy for the time when the firm conducts the corresponding innovation activities. On average, it takes two year for a patent to be granted after the inventor files the application. I drop patents granted in year 2009 and 2010 because it is very likely that a patent applied in 2009 is not granted prior to the end of 2010, which means it may not show up in the sample.

Table 3.1: Summary statistics

Variable	N	Mean	Std. Dev.	Min	Max
R&D expenditure	87,165	67.851	395.7878	0	12,183
Market value	86,589	101.283	811.9765	-78	109,746
Total assets	86,326	732.85	3,798.717	0	146,171,100
Number of employees	84,916	8.946	33.99	0	876.8

R&D expenditure, total assets are in thousands of 2009 US dollars. Market value is in millions of 2009 US dollars. Number of employees are in thousands.

3.3.5 Vertical relations in intermediate good markets

I use the Annual Input-Output (I-O) Accounts from Bureau of Economic Analysis (BEA) to calculate the vertical relations between industries. I group industries using the two-digit North America Industrial Classification System (NAICS). There are a total of 22 industries. For each industry s, $pbuy_{sr}(t = 1, 2, ..., 22$ and $r \neq s)$[5] is the share of market values of input industry s buys from industry r. $pbuy_{sr}$ can range from 0 to 1, with 1 meaning industry s purchasing all the inputs from industry t and 0 meaning industry s purchases nothing from industry r. Similarly, $psell_{sr}$ is the share of input industry s sells to industry r ($r = 1, 2, ..., 22$ and $r \neq s$). $pbuy_{sr}$ and $psell_{sr}$ are indicators of vertical relations. For example if most of Industry A's purchase of intermediate goods is from industry B, $pbuy_{AB}$, which reflects that B is an upstream industry of A.

3.3.6 Vertical relations in the patent system

I use patent citations to calculate vertical relations between technology classes. The logic behind is that I consider knowledge as an input of in-

[5]In general, flows of intermediate goods can happen within the same industry. However, since this paper focus on cross-industry spillovers, I require $r \neq s$).

novation activities. However, it is almost impossible to actually trace the knowledge that was used in any sort of research. A good predictor of knowledge flows is patent citations (Jaffe et al., 1993). The US patent laws require inventors to report any known prior arts by citing the relevant patents. Similarly, if technology class A cites a lot from technology class B, then B might as well be an upstream technology class of A.

There are a total of 36 technology classes in my data. For each technology class p, $citing_{pq}$ ($q = 1, 2, ..., 36$ and $q \neq p$) is the share of patent citations made from technology class p to q. $citing_{pq}$ ranges from 0 to 1, and greater $citing_{pq}$ indicates a higher technological dependence of industry p on q. Similarly, $cited_{pq}$ is the share of patent citations technology class p receives from q. $citing_{pq}$ and $cited_{pq}$ are going to be used as weights to pin down vertical relations (discussed in Section 3.3.7).

3.3.7 Positions in intermediate good markets and the patent system

For each firm i, I calculate its position in the intermediate good markets by calculating $sales\_pct_{is}$, the share of firm i's sales in industry s. Sales data is from Compustat segment. Similarly, for firm i's position in the patent system, $patent_{ip}$ is the share of market values of patents in technology class p.

3.3.8 Constructing $InterUp_{it}, InterDown_{it}, PatentUp_{it}$, and $PatentDown_{it}$

For each firm i in year t, I first calculate the R&D expenditures in its upstream industries and downstream industries in both intermediate good markets and the patent system. Define $indRD_{rt}$ as the total R&D expenditures of industry r in year t.

$$Inter\_R\&D\_Up_{it} = \Sigma_s \Sigma_r sales\_pct_{is} pbuy_{sr} RD_{rt}$$

$$Inter\_R\&D\_Down_{it} = \Sigma_s \Sigma_r sales\_pct_{is} psell_{sr} RD_{rt}$$

$$Patent\_R\&D\_Up_{it} = \Sigma_p \Sigma_q patent_{ip} citing_{pq} RD_{qt}$$

$$Patentt\_R\&D\_Down_{it} = \Sigma_p \Sigma_q patent_{ip} cited_{pq} RD_{qt}$$

In short, for each firm, its upstream/downstream R&D expenditure is the weighted average of the weighted average of upstream/downstream industries, where the first weights are the firm's position in intermediate good market/the patent system, and the second weights are the vertical relations between industries.

3.4 Empirical Strategy

3.4.1 Equations for estimation

The general equation of my analysis takes the form:

$$ln(Y_{it}) = \beta_1 ln(InterUpStk_{it}) + \beta_2 ln(InterDownStk_{it}) + \beta_3 ln(PatentUpStk_{it}) + \beta_4 ln(PatentDownStk_{it}) + \beta_5 X_{it} + \lambda_i + \lambda_t + \epsilon_{it}$$

where Y_{it} is the dependent variable of interest of firm i in year t. X_{it} is a vector of controls, γ_i is firm fixed effect and γ_t is the year fixed effect. $InterUpStk_{it}$, $InterDownStk_{it}$, $PatentUpStk_{it}$ and $PatentDownStk_{it}$ are the R&D stock in the upstream industry and downstream industry, respectively, both in intermediate good markets and the patent system, calculated using a perpetual inventory method. In particular,

$$ln(InterUpStk_{it}) = (1-\delta)ln(InterUpStk_{it-1}) + ln(InterUp_{it})$$

$$ln(InterDownStk_{it}) = (1-\delta)ln(InterDownStk_{it-1}) + ln(InterDown_{it})$$

$$ln(PatentUpStk_{it}) = (1-\delta)ln(PatentUpStk_{it-1}) + ln(PatentUp_{it})$$

$$ln(PatentDownStk_{it}) = (1-\delta)ln(PatentDownStk_{it-1}) + ln(PatentDown_{it})$$

where $\delta = 0.15$ is the depreciate rate.

In the first set of equations for estimation, I let Y_{it} be the R&D intensity (R&D expenditure divided by sales). Since it usually takes time for firms to respond to shocks from other industries and formulate new R&D strategies, I lag all the independent variables by one year in this set of equations. In the second set of equations, I let Y_{it} be the log of market values.

3.4.2 Instrumental variable approach

A potential of this estimation technique is that it might be susceptible to endogenous R&D shocks. If new research opportunities arise, it may lead to an increase in focal firm's R&D and related industries' R&D. To address this issue, I use the costs of R&D capital calculated from federal and state R&D tax credit (Wilson, 2009) as instruments. One might be concerned that R&D tax credit policy may also be endogenous to innovation shocks. However, the existing literature agrees that there is no evidence that economic conditions predicted R&D policy. In my estimation, I use costs of R&D capital to predict each firm's R&D expenditure, and use the predicted R&D expenditures to construct $InterUp_{it}, InterDown_{it}, PatentUp_{it}$, and $PatentDown_{it}$.

For R&D tax credit to be a valid instrument, it has to satisfy two requirements: 1) the instrument must be correlated with the endogenous explanatory variable. 2) the instrument might be exogenous. I discuss them separately here.

The F-statistic of the first stage regression of an instrumental variable approach is commonly used to test the relevance of the instrument. A rule of thumb is that the F-statistic should be larger than 10. In Table 3.2, I present the first stage result. The F-statistic is 15.345.

This is no statistical tool to test the exogeneity of an instrument. Here I discuss it based on a survey of the literature. Studies find evidence that state R&D tax credit policies are partly driven by some macroeconomic conditions, but local economic variables seem to be not important (Chirinko and Wilson, 2008). This is partly because there are randomly long time delays in passing state R&D tax credit policies and R&D tax credits are generally small so that they are usually not affected by state budget conditions.

Table 3.2: First stage regression

Variable	(1)
R&D capital cost	-0.224***
	(0.0586)
Industrial sales	0.050***
	(0.016)
F-statistic	15.345
Number of observations	21,299

Standard errors in parentheses.* ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). The dependent variable is R&D expenditures at firm-year level. Firm fixed effects and year fixed effects are used. Standard errors are clustered at firm level.

3.5 Results

In Table 3.3, I present results of the strategic R&D estimation, where the dependent variable is the log of R&D intensity and R&D intensity is calculated by dividing R&D expenditure by sales. In Column (1), I don't control for industrial sales. Results show that innovations in downstream industries in intermediate good market are negatively correlated with a downstream firm's R&D expenditure. In the patent system, R&D activities in upstream are positively associated with a downstream firm's R&D, while downstream industries' R&D expenditures are negatively correlated with an upstream firm's R&D.

In Column (2), I additionally control for industrial sales. The signs of coefficients on $ln(InterDown), ln(PatentUp)$ and $ln(PatentDown)$ don't change, but their magnitudes increase a little bit. The coefficient on $ln(InterUp)$

Table 3.3: Strategic R&D estimation

Variable	(1)	(2)	(3)
	OLS	OLS	IV
$ln(InterUp)$	-0.146	-0.363*	-0.729***
	(0.161)	(0.195)	(0.153)
$ln(InterDown)$	-0.402**	-0.551***	0.018
	(0.176)	(0.196)	(0.131)
$ln(PatentUp)$	0.841***	1.373***	0.983***
	(0.191)	(0.234)	(0.166)
$ln(PatentDown)$	-0.485***	-0.653***	-0.433***
	(0.142)	(0.167)	(0.144)
			F-test
			15.345
Number of observations	26,280	21,299	21,299

Standard errors in parentheses.* ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). The dependent variable is R&D intensity (R&D expenditure/Sales) at firm-year level. Column (2) and (3) include industrial sales as a control variable. Column (3) also uses costs of R&D capital (calculated from federal and state R&D tax credit) as an instrument. Standard errors are clustered at firm level.

is negative and becomes marginally significant at 10% level.

Column (3) displays the instrumental variable approach results. It shows that in intermediate good markets, a 1% increase in R&D expenditures in upstream industries will decrease a downstream firm's R&D expenditures by 0.729%. Meanwhile, in the patent system, a 1% increase in R&D expenditures in upstream raises a downstream firm's R&D expenditure by 0.983%, and on the reverse side, a 1% increase in R&D expenditures in downstream lowers an upstream firm's R&D by 0.433%.

Table 3.4 presents results where the dependent variable is the log of market values of firms. In Column (1), I don't control for industrial sales, while in Column (2) I additionally include industrial sales as an independent variable. Column (3) is the instrumental variable approach. We can see that the coefficient on $ln(InterUp)$ is insignificant in all columns. The coefficient on $ln(InterDown$ are consistently negative and significant in all specifications. In general, a 1% increase in R&D investment in the intermediate good downstream industries lead to a -0.907% decrease in an upstream firm's market value. In the patent system, a 1% increase in R&D investment in the technology upstream industries lead to a 1.053% increase in the market value of a downstream firm. Note that the instrumental variable approach also makes the coefficient on $ln(PatentDown)$ marginally significant.

3.5.1 Sensitivity Analysis

In this section I fist allow alternative depreciation rates. In Table 3.5 and 3.6, I use $\delta = 0.1$ and $\delta = 0.2$, where 3.5 presents results with log R&D

Table 3.4: Market value equation

Variable	(1)	(2)	(3)
$ln(InterUp)$	-0.306	-0.353	-0.263
	(0.245)	(0.256)	(0.243)
$ln(InterDown)$	-0.474*	-0.882***	-0.907***
	(0.267)	(0.287)	(0.212)
$ln(PatentUp)$	0.849***	1.443***	1.053***
	(0.322)	(0.349)	(0.275)
$ln(PatentDown)$	0.097	0.081	0.350*
	(0.217)	(0.249)	(0.200)
Number of observations	28,546	23,327	23,327

Standard errors in parentheses.* ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). The dependent variable is market value at firm-year level. Column (2) and (3) include industrial sales as a control variable. Column (3) also uses costs of R&D capital (calculated from federal and state R&D tax credit) as an instrument. Standard errors are clustered at firm level.

as the dependent variable and 3.6 presents results with log market value as the dependent variable. To make it easy for comparison, Column (1) of Table 3.5 is identical to column (3) of Table 3.3 and Column (1) of Table 3.6 is the same as Column (3) of Table 3.4. We can see that using different depreciation rates of doesn't change the sign of the coefficients. My results are robust to different depreciation rates.

Table 3.5: Sensitivity analysis: alternative depreciation rate (strategic R&D)

Depreciate rate	15%	20%	10%
Variable	(1)	(2)	(3)
$ln(InterUp)$	-0.729***	-0.683***	-0.774***
	(0.153)	(0.145)	(0.162)
$ln(InterDown)$	0.018	0.007	0.026
	(0.131)	(0.125)	(0.138)
$ln(PatentUp)$	0.983***	0.931***	1.045***
	(0.166)	(0.158)	(0.176)
$ln(PatentDown)$	-0.433***	-0.421***	-0.454***
	(0.144)	(0.136)	(0.153)
Number of observations	21,299	21,299	21,299

Standard errors in parentheses.* ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). The dependent variable is R&D expenditures at firm-year level. Each column uses a different depreciate rate to calculate R&D stocks in upstream and downstream industries. All columns control for industrial sales and also use costs of R&D capital (calculated from federal and state R&D tax credit) as instruments. Standard errors are clustered at firm level.

All the previous results use the total R&D expenditures in the upstream/downstream industries. One potential concern is that the size of the upstream/downstream industries might affect downstream/upstream

Table 3.6: Sensitivity analysis: alternative depreciation rate (market value)

Depreciate rate	15%	20%	10%
Variable	(1)	(2)	(3)
$ln(InterUp)$	-0.263	-0.253	-0.270
	(0.243)	(0.229)	(0.162)
$ln(InterDown)$	-0.907***	-0.870***	-0.954***
	(0.212)	(0.203)	(0.223)
$ln(PatentUp)$	1.053***	1.021***	1.098***
	(0.275)	(0.260)	(0.294)
$ln(PatentDown)$	0.349*	0.348*	0.350*
	(0.200)	(0.191)	(0.213)
Number of observations	23,327	23,327	23,327

Standard errors in parentheses.* ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). The dependent variable is market values at firm-year level. Each column uses a different depreciate rate to calculate R&D stocks in upstream and downstream industries. All columns control for industrial sales and also use costs of R&D capital (calculated from federal and state R&D tax credit) as instruments. Standard errors are clustered at firm level.

firms. In this section, I address this issue by using the average R&D expenditures (over firms) in upstream/downstream industries. Table 3.7 presents the results. The dependent variable in Column (1) and (2) is log of R&D intensity (R&D expenditure/Sales) at firm-year level, and in Column (3) and (4) the dependent variable is log of market value. All columns include industrial sales as a control variable, and use costs of R&D capital (calculated from federal and state R&D tax credit) as an instrument.Column (2) and (4) use average R&D expenditures in the upstream/downstream industries. If we compare the results in this table with those in Table 3.3 and Table 3.4, we can see the magnitudes and significance of the coefficients don't change much.

3.6 Conclusion

A firm's performance is not only affected by its own R&D, but also by other firms' research efforts. Although many studies have found empirical evidence that technology spills over across countries and within industries, few studies focus on inter-industrial spillovers. This paper studies how innovations transmit vertically across industries. I first develop a method to establish vertical relations between industries using sales data in intermediate good markets and patent citation data. Then given each firm's position in intermediate good markets and the patent system, I calculate its upstream and downstream industries' R&D expenditures.

Using data on 2,803 public firms in the US from 1976 to 2008, I estimate the impacts of innovations in the upstream/downstream industries on downstream/upstream firms. To deal with potential endogeneity problems,

Table 3.7: Average industrial R&D expenditures

Variable	(1)	(2)	(3)	(4)
$ln(InterUp)$	-0.729***	-0.844***	-0.263	0.533*
	(0.153)	(0.175)	(0.243)	(0.278)
$ln(InterDown)$	0.018	-0.500***	-0.907***	-1.550***
	(0.131)	(0.182)	(0.212)	(0.309)
$ln(PatentUp)$	0.983***	1.503***	1.053***	1.010***
	(0.166)	(0.234)	(0.275)	(0.361)
$ln(PatentDown)$	-0.433***	-0.335*	0.350*	0.251
	(0.144)	(0.181)	(0.200)	(0.266)
Number of observations	21,299	21,299	23,327	23,327

Standard errors in parentheses.* ($p < 0.05$), ** ($p < 0.01$), *** ($p < 0.001$). The dependent variable in column (1) and (2) is log of R&D intensity (R&D expenditure/Sales) at firm-year level, and in column (3) and (4) the dependent variable is log of market value. All columns include industrial sales as a control variable, and use costs of R&D capital (calculated from federal and state R&D tax credit) as an instrument.Column (1) and (3) use total R&D expenditure in the upstream/downstream industries. Column (2) and (4) use average R&D expenditures in the upstream/downstream industries. Standard errors are clustered at firm level.

I use costs of R&D capital calculated from state R&D tax credit as an instrumental variable. My results show that in intermediate good markets, a 1% increase in research and development (R&D) expenditures in upstream industries will decrease a downstream firm's R&D expenditure by 0.729%, while a 1% increase in R&D of the downstream will reduce a upstream firm's market value by 0.907%. Meanwhile, in the patent system, the upstream to downstream R&D and market value elasticities are estimated to be 0.983 and 1.053.

Future research should first focus on the theoretical foundations of vertical spillovers. In this paper, I highlighted three possible channels through which technology could transmit vertically across industries, but we still need a complete model to answer these questions. Second, due to constraint of data, I am unable to explore how market structures might play a role in vertical spillovers. It could be interesting to see how technology spillovers depend on the type of the market.

References

Agrawal, Ajay, Alberto Galasso, and Alexander Oettl (2016), "Roads and innovation." *Review of Economics and Statistics*.

Almeida, Paul and Bruce Kogut (1997), "The exploration of technological diversity and geographic localization in innovation: start-up firms in the semiconductor industry." *Small Business Economics*, 9, 21–31.

Almeida, Paul and Bruce Kogut (1999), "Localization of knowledge and the mobility of engineers in regional networks." *Management science*, 45, 905–917.

Anderson, Simon P and Stephen Coate (2005), "Market provision of broadcasting: A welfare analysis." *The review of Economic studies*, 72, 947–972.

Anderson, Simon P and Bruno Jullien (2015), "The advertising-financed business model in two-sided media markets." In *Handbook of Media Economics*, volume 1, 41–90, Elsevier.

Argentesi, Elena and Lapo Filistrucchi (2007), "Estimating market power in a two-sided market: The case of newspapers." *Journal of Applied Econometrics*, 22, 1247–1266.

Armstrong, Mark (2006), "Competition in two-sided markets." *The RAND Journal of Economics*, 37, 668–691.

Arrow, Kenneth (1962), "Economic welfare and the allocation of resources for invention." In *The rate and direction of inventive activity: Economic and social factors*, 609–626, Princeton University Press.

Arvey, Richard D (2009), "Why face-to-face business meetings matter." *White Paper for the Hilton Group*.

Athey, Susan, Emilio Calvano, and Joshua Gans (2013), "The impact of the internet on advertising markets for news media." Technical report, National Bureau of Economic Research.

Badinger, Harald and Peter Egger (2016), "Productivity spillovers across countries and industries: New evidence from OECD countries." *Oxford Bulletin of Economics and Statistics*, 78, 501–521.

Bernstein, Jeffrey I and Pierre Mohnen (1998), "International R&D spillovers between us and japanese R&D intensive sectors." *Journal of international economics*, 44, 315–338.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen (2013), "Identifying technology spillovers and product market rivalry." *Econometrica*, 81, 1347–1393.

Breitzman, Anthony (2013), "Patent trends among small and large innovative firms during the 2007-2009 recession." *US Small Business Administration. Available at¡ http://www. sba. gov/sites/default/files/rs411tot. pdf.*

Caillaud, Bernard and Bruno Jullien (2003), "Chicken & egg: Competition among intermediation service providers." *RAND journal of Economics*, 309–328.

Chirinko, Robert S and Daniel J Wilson (2008), "State investment tax incentives: A zero-sum game?" *Journal of Public Economics*, 92, 2362–2384.

Choi, Jay Pil (2006), "Broadcast competition and advertising with free entry: Subscription vs. free-to-air." *Information Economics and Policy*, 18, 181–196.

Coe, David T and Elhanan Helpman (1995), "International R&D spillovers." *European economic review*, 39, 859–887.

Giroud, Xavier (2013), "Proximity and investment: Evidence from plant-level data*." *Quarterly Journal of Economics*, 128.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg (2001), "The nber patent citation data file: Lessons, insights and methodological tools." Technical report, National Bureau of Economic Research.

Hausman, Jerry, Bronwyn H Hall, Zvi Griliches, et al. (1984), "Econometric models for count data with an application to the patents-r&d relationship." *Econometrica*, 52, 909–38.

Hicks, Diana, Tony Breitzman, Dominic Olivastro, and Kimberly Hamilton (2001), "The changing composition of innovative activity in the usa portrait based on patent analysis." *Research Policy*, 30, 681–703.

Jaffe, Adam, Manuel Trajtenberg, and Michael Fogarty (2000), “Knowledge spillovers and patent citations: Evidence from a survey of inventors.” *The American Economic Review*, 90, 215–218.

Jaffe, Adam B (1988), “Demand and supply influences in R&D intensity and productivity growth.” *The Review of Economics and Statistics*, 431–437.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson (1993), “Geographic localization of knowledge spillovers as evidenced by patent citations.” *Quarterly Journal of Economics*, 577–598.

Kaiser, Ulrich and Minjae Song (2009), “Do media consumers really dislike advertising? an empirical assessment of the role of advertising in print media markets.” *International Journal of Industrial Organization*, 27, 292–301.

Kaiser, Ulrich and Julian Wright (2006), “Price structure in two-sided markets: Evidence from the magazine industry.” *International Journal of Industrial Organization*, 24, 1–28.

Keller, Wolfgang (2002), “Trade and the transmission of technology.” *Journal of Economic growth*, 7, 5–24.

Kind, Hans Jarle, Tore Nilssen, and Lars S rgard (2007), “Competition for viewers and advertisers in a tv oligopoly.” *Journal of Media Economics*, 20, 211–233.

Krugman, Paul R (1991), *Geography and trade*. MIT press.

Li, Guan-Cheng, Ronald Lai, Alexander DAmour, David M Doolin, Ye Sun, Vetle I Torvik, Z Yu Amy, and Lee Fleming (2014), “Disambiguation and co-authorship networks of the us patent inventor database (1975–2010).” *Research Policy*, 43, 941–955.

Marshall, Alfred (1890), *Principles of political economy*. Maxmillan, New York.

Maurseth, Per Botolf and Bart Verspagen (2002), “Knowledge spillovers in europe: a patent citations analysis.” *The Scandinavian Journal of Economics*, 104, 531–545.

Mehta, Aditi, Marc Rysman, and Tim Simcoe (2010), “Identifying the age profile of patent citations: New estimates of knowledge diffusion.” *Journal of Applied Econometrics*, 25, 1179–1204.

Nelson, Richard R. (1959), "The Simple Economics of Basic Scientific Research." *Journal of Political Economy*, 67, 297.

Reisinger, Markus (2012), "Platform competition for advertisers and users in media markets." *International Journal of Industrial Organization*, 30, 243–252.

Rysman, Marc (2004), "Competition between networks: A study of the market for yellow pages." *The Review of Economic Studies*, 71, 483–512.

Silva, JMC Santos and Silvana Tenreyro (2006), "The log of gravity." *The Review of Economics and statistics*, 88, 641–658.

Sonn, Jung Won and Michael Storper (2008), "The increasing importance of geographical proximity in knowledge production: an analysis of us patent citations, 1975–1997." *Environment and Planning A*, 40, 1020–1039.

Thompson, Peter and Melanie Fox-Kean (2005), "Patent citations and the geography of knowledge spillovers: A reassessment." *American Economic Review*, 450–460.

Wagner, Stefan, Karin Hoisl, and Grid Thoma (2014), "Overcoming localization of knowledgethe role of professional service firms." *Strategic management journal*, 35, 1671–1688.

Weber, Sylvain, Martin Péclat, et al. (2016), "A simple command to calculate travel distance and travel time." Technical report, IRENE Institute of Economic Research.

Wilbur, Kenneth C (2008), "A two-sided, empirical model of television advertising and viewing markets." *Marketing science*, 27, 356–378.

Wilson, Daniel J (2009), "Beggar thy neighbor? the in-state, out-of-state, and aggregate effects of R&D tax credits." *The Review of Economics and Statistics*, 91, 431–436.

Wooldridge, Jeffrey M (1999), "Distribution-free estimation of some nonlinear panel data models." *Journal of Econometrics*, 90, 77–97.

Appendices

A Calculating driving time

I use the Stata command *georoute* written by Weber et al. (2016). It utilizes HERE's map services. To acquire the driving time between two locations, one can use either addresses or coordinates of the two points. While obtaining airport addresses is not an issue, it is not straightforward to assign an address or a coordinate to an MSA. I use the average coordinates weighted by number of patents to calculate the "coordinate" for an MSA. For example, suppose MSA j has 7 patents in my sample, and 4 of them are applied at the coordinate (lat_1, lon_1) and 3 are applied at (lat_2, lon_2), then the coordinate of MSA i is $((4lat_1 + 3lat_2)/7, (4lon_1 + 3lon_2)/7)$.

B Reductions in travel time as a treatment

I consider reductions in travel time as treatments. Specifically, I divide the treatments into several groups, each corresponding to a certain amount of travel time reduction. $\boldsymbol{trt_{ijt}} = (trt0\_15_{ijt}, trt15\_30_{ijt}, trt30\_60_{ijt}, trt60\_120_{ijt}, trt\_120_{ijt})$. Each element of the vector equals one if the reduction falls in its range. For example, $trt30\_60_{ijt} = 1$ if the reduction is between 30 and 60 minutes, and $trt\_120_{ijt} = 1$ if travel time is reduced by more than 120 minutes.

In many cases, an airline exits a market shortly after its entry once it finds the market is not profitable. It is hard for business travelers to benefit from those temporary travel time reductions. To make sure the reduction in travel time is stable, I first construct two variable.

$$atra\_before_{ijt} = (travel_{ijt}+travel_{ij,t-1}+travel_{ij,t-2}+travel_{ij,t-3}+travel_{ij,t-4})/5$$

$$atra\_after_{ijt} = (travel_{ijt}+travel_{ij,t+1}+travel_{ij,t+2}+travel_{ij,t+3}+travel_{ij,t+4})/5$$

$atra\_before_{ijt}$ is the five-year average travel time prior to year t and $atra\_after_{ijt}$ is the five-year average travel time after year t. Each element of $\boldsymbol{trt_{ijt}}$ equals 1 if $travel_{ijt}$ is reduced by certain amount compared to $travel_{ij,t-1}$, and $atra\_after_{ijt}$ is reduced by certain amount compared to $atra\_before_{ij,t-1}$. Finally, $trt_{ijt} = 1$ if $trt_{ij,t-1} = 1$.

C Proof of Chapter 2 Proposition 2

Since we have a continuum of agents of mass 1 on either side, we have $N_i^j = n_i^j$ and $n_1^j = x^j$, $i = 1, 2$, $j = a, r$. Then equations (2.9) and (2.11) also provide the expressions for marginal advertiser and marginal reader. Setting $\varepsilon_i^j = 0$, we have

$$x^a = \frac{1}{2} + \rho(N_1^r - N_2^r) + \frac{\rho^2}{2}\left[(N_1^r)^2 - (N_2^r)^2\right] - (a_1 - a_2),$$

$$x^r = \frac{1}{2} + \gamma(N_1^a - N_2^a) - \gamma \cdot \rho \cdot (N_1^r \cdot N_1^a - N_2^r \cdot N_2^a) - (p_1 - p_2).$$

Substituting the x^j expressions into the following

$$N_1^j = x^j, \quad N_2^j = 1 - x^j, \quad j = a, r,$$

and solve for N_i^j, we can obtain

$$N_1^a = \frac{1}{2} - \frac{\gamma(\rho+2)(a_1 - a_2) + \rho(\rho+2)(p_1 - p_2)}{\gamma\rho^3 - 3\gamma\rho + 1},$$

$$N_1^r = \frac{1}{2} + \frac{\gamma(\rho-2)(a_1 - a_2) - (p_1 - p_2)}{\gamma\rho^3 - 3\gamma\rho + 1}.$$

Platform i's problem is :

$$\max_{p_i, a_i} \quad \pi_i = p_i \cdot N_i^r + a_i \cdot N_i^a.$$

Solving firms' FOCs, the equilibrium prices are

$$p_i = \frac{1}{2} - \rho + \frac{1}{2}\rho(\gamma - \rho), \quad a_i = \frac{1}{2} - \gamma + \frac{1}{2}\rho\gamma,$$

and each platform earns a profit of

$$\pi_i = \frac{1}{2}\gamma\rho - \frac{1}{2}\gamma + \frac{1}{2} - \frac{1}{2}\rho - \frac{1}{4}\rho^2 \quad i = 1, 2.$$

■

D Proof of Chapter 2 Proposition 3

(i) Equilibrium prices at the benchmark model (with superscript 'b') and strategic agents model (with superscript 's') are,

$$p_i^b = \frac{1}{2} - \rho, \quad a_i^b = \frac{1}{2} - \gamma,$$

$$p_i^s = \frac{1}{2} - \rho + \frac{1}{2}\rho(\gamma - \rho), \quad a_i^s = \frac{1}{2} - \gamma + \frac{1}{2}\rho\gamma.$$

The difference $p_i^b - p_i^s = -\frac{1}{2}\rho(\gamma - \rho) > 0$ since γ is negative and ρ is positive. Similarly, $a_i^b - a_i^s = -\frac{1}{2}\rho\gamma > 0$. Combined, equilibrium prices on both sides are lower under strategic agents relative to the benchmark case.

(ii) With lower prices on both sides, advertisers and readers must be better off at the cost of platforms. ■

E Proof of Chapter 2 Proposition 4

There is no change on the final agents relative to the benchmark case. Therefore, the agents' demand functions are the same,

$$n_1^r = \frac{1}{2} + \frac{2\gamma(a_1 - a_2) + (p_1 - p_2)}{4\gamma\rho - 1}$$

$$n_1^a = \frac{1}{2} + \frac{2\rho(p_1 - p_2) + (a_1 - a_2)}{4\gamma\rho - 1}$$

Let π_i^{ret} denote the profit of retailer $i = 1, 2$, and let π_i^{man} denote platform (manufacturer) i's profit. The profit maximization problems are

$$\max_{p_1} \quad \pi_1^{ret} = (p_1 - w_1) \cdot n_1^r, \quad \max_{p_2} \quad \pi_2^{ret} = (p_2 - w_2) \cdot (1 - n_1^r),$$

$$\max_{a_1} \quad \pi_1^{man} = w_1 \cdot n_1^r + a_1 \cdot n_1^a, \quad \max_{a_2} \quad \pi_2^{man} = w_2 \cdot (1 - n_1^r) + a_2 \cdot (1 - n_1^a).$$

Solving the FOCs, we can obtain the optimal retail prices as functions of w_1 and w_2.[6] We also verify that

$$p_1 - p_2 = -\frac{(8\gamma^2 + 3)(w_1 - w_2)}{16\gamma\rho - 9}.$$

Next, we substitute these retail prices into platforms' profit maximization problems,

$$\max_{w_1} \quad \pi_1^{man} = w_1 \cdot n^r + a_1 \cdot n^a,$$

$$\max_{w_2} \quad \pi_2^{man} = w_2 \cdot (1 - n^r) + a_2 \cdot (1 - n^a).$$

Taking derivatives and then imposing symmetry ($w_1 = w_2$), we can obtain

$$w_1 = w_2 = \frac{9 - 4\rho - 6\gamma - 16\rho\gamma}{16\gamma^2 + 6}.$$

Substituting the wholesale prices into retail prices, we have

$$p_i = \frac{9 - 4\rho - 6\gamma - 16\rho\gamma}{16\gamma^2 + 6} + \left(\frac{1}{2} - 2 \cdot \rho \cdot \gamma\right) = w_i + \left(\frac{1}{2} - 2 \cdot \rho \cdot \gamma\right),$$

$$a_i = -2 \cdot \gamma \cdot \frac{9 - 4\rho - 6\gamma - 16\rho\gamma}{16\gamma^2 + 6} + \left(\frac{1}{2} - 2 \cdot \rho \cdot \gamma\right) = -2 \cdot \gamma \cdot w_i + \left(\frac{1}{2} - 2 \cdot \rho \cdot \gamma\right).$$

[6] They are lengthy and skipped. A maple file containing the results is available upon request.

$\blacksquare$

F A model of strategic R&D and vertical relations

I develop a theoretical framework for strategic R&D of vertically related firms. Firms $U_1, U_2, ..., U_N$ are in the upstream and firms $D_1, D_2, ...D_M$ are in the downstream. For simplicity, I assume $M = N$ and each downstream firm D_i has only one supplier U_i. The latter can be true when the cost of switching to another supplier is extremely high.

There are three stages of game. In the first stage, both upstream and downstream firms engage in R&D activities that will affect the quality of products and marginal cost of production. In the second stage, the upstream firm U_i produces an input at marginal cost c_U, and sells it to the downstream firm D_i at price w. In the third stage, downstream firm D_i use to input and produce the final product at marginal cost c_D and sell the final product to consumers at price p. Suppose the demand function is given by $D = \alpha + \beta s_U + s_D - p$, where s_U and s_D are the quality of products sold by the upstream firm and downstream firm, respectively, and $\beta > 0$, then the profit function for firm D_i in the third stage is

$$\pi_D = (p - w - c_D)(\alpha + \beta s_U + s_D - p).$$

Solving the first order condition yields

$$p = \frac{1}{2}(\alpha + \beta s_U + s_D + w + c_D) \tag{1}$$

Anticipating p, the profit function for firm U_i in the second stage is

$$\pi_U = (w - c_U)(\alpha + \beta s_U + s_D - p) \tag{2}$$

Substitute (1) into (2) and solve the first condition $\partial \pi_U / \partial w$, and we have

$$w^* = \frac{1}{2}(\alpha + \beta s_U + s_D - c_D + c_U) \tag{3}$$

Substitute (3) for w in (1), we have

$$p^* = \frac{3}{4}(\alpha + \beta s_U) + \frac{1}{4}(c_D + c_U) \tag{4}$$

The optimal choice of w^* and p^* are functions of quality of products and marginal cost of production, which are determined by R&D in the first stage. Moreover, the profit functions for both firms are given by

$$\pi_D = \frac{1}{16}(\alpha + \beta s_U + s_D - c_D - c_U)^2 \tag{5}$$

$$\pi_U = \frac{1}{8}(\alpha + \beta s_U + s_D - c_D - c_U)^2 \tag{6}$$

In the next section, I will show how firm $U_i (or D_i)$ will respond when there is an exogenous increase in firm $D'_i (or U'_i)$ R&D expenditures in two different cases: 1) R&D increases quality (s); 2) R&D reduces marginal cost (c).

Suppose R&D improve product quality for both upstream and downstream firms.[7] The profit function for downstream firm D_i in the first stage is

$$\pi_D = \frac{1}{16}(\alpha + \beta s_U + s_D - c_D - c_U)^2 - r_D \tag{7}$$

where r_D stands for the R&D spending of the downstream firm. Since R&D is quality-improving, let $s_U = \phi(r_U)$ and $s_D = \lambda(r_D)$ and assume $\phi^{'}(\cdot) > 0, \lambda^{'}(\cdot) > 0, \phi^{"}(\cdot) < 0$, and $\lambda^{"}(\cdot) < 0$. The concave feature of the functions ensure that there is diminishing marginal return on quality-improving R&D. First order condition yields

$$\frac{1}{8}(\alpha + \beta\phi(r_U) + \lambda(r_D) - c_D - c_U)\lambda^{'}(r_D) - 1 = 0 \tag{8}$$

Comparative statics show that

$$\frac{\partial r_D^*}{\partial r_U} = \frac{\beta\lambda^{'}(r_D)\phi^{'}(r_U)}{-\lambda^{''}(r_D)\Delta - (\lambda^{'}(r_D))^2} \tag{9}$$

where $\Delta = \alpha + \beta s_U + s_D - c_D - c_U$. In the appendix I prove that $\Delta > 0$. Since $\beta\lambda^{'}(r_D)\phi^{'}(r_U)$ and $(\lambda^{'}(r_D))^2$ are always positive, the sign of $\partial r_D^*/\partial r_U$ is ambiguous and depends on $\lambda^{''}(r_D)$. $\partial r_D^*/\partial r_U$ is positive when $-\lambda^{''}(r_D)\Delta - (\lambda^{'}(r_D))^2 > 0$, negative when $-\lambda^{''}(r_D)\Delta - (\lambda^{'}(r_D))^2 < 0$

The profit function for firm U_i in the first stage is

$$\pi_D = \frac{1}{8}(\alpha + \beta s_U + s_D - c_D - c_U)^2 - r_U \tag{10}$$

[7]Another type of R&D is cost-reducing. In the Appendix I show it generates similar results

Similarly, total-differentiating the first condition yields

$$\frac{\partial r_U^*}{\partial r_D} = \frac{\lambda^{'}(r_D)\phi^{'}(r_U)}{-\phi^{''}(r_U)\Delta - \beta(\phi^{'}(r_U))^2} \tag{11}$$

A little further analysis shows that $\partial r_U^*/\partial r_D$ is positive when $-\phi^{''}(r_U)\Delta - \beta(\phi^{'}(r_U))^2 > 0$, and negative when $-\phi^{''}(r_U)\Delta - \beta(\phi^{'}(r_U))^2 < 0$