ESSAYS ON APPLIED MICROECONOMICS

By

Kyung Ho Song

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Economics — Doctor of Philosophy

2019

ABSTRACT

ESSAYS ON APPLIED MICROECONOMICS

By

Kyung Ho Song

Chapter 1. Education and Neighborhood Premium in Housing Price and its Dynamics

From a newly-assembled panel dataset, this study estimates the implicit prices of better education and neighborhood environments through models in which unobservable housing attributes are allowed to vary over time. The results of this study show that, first, past records of school performances have a long-lasting reputation effect, which has not been considered in the previous literature. Without taking the reputation effect into account, estimation results can be biased. Second, education and neighborhood premiums account for more than 43% of the housing price gaps. This paper also analyzes the dynamics of the housing market by using the panel VAR, a new approach in urban economics. The dynamic model shows more remarkable effects of education and neighborhood environments on housing price than a static model through bilateral – direct and indirect – relations among school performance, neighborhood composition, amenities, and housing price over time. This result supports the idea that education and neighborhood environments have more substantial impacts on housing prices and city segregation in a dynamic framework.

Chapter 2. Impact of Resale Price Maintenance (RPM) Regulation on the Book Market of Korea

The reinforced resale price maintenance (RPM) regulation applied to books took effect from November 2014 in Korea under the policy purpose of preserving cultural variety and knowledge by supporting small publishing companies and bookstores. However, empirical evidence implies that the new regulation should be revised to achieve its policy purpose. After the new regulation, the number of newly-published books, which can be used as the index for cultural variety, was decreased by 7% annually after de-trending. The profit ratio of the six largest online firms was enhanced by more than 5% and the amount of their profit also increased significantly. However, the top 70 publishing companies' sales decreased by more than 10% and the sales of small book stores also decreased. Based on the empirical analyses, this paper suggests that other methods such as direct subsidy should be reviewed instead of

price regulation to achieve the policy's purpose.

Chapter 3. Impact of Resale Price Maintenance (RPM) Regulation on the Used Book Market

A new bill came into effect in November 2014 that further strengthens the existing resale price maintenance (RPM) regulation on the book market in Korea under the policy purpose to foster diversity of culture and knowledge by protecting small publishing companies and offline bookstores. However, after the new regulation took effect, some studies find evidence that is contrary to the intention of the regulation from various data. Most importantly, the gap in the market share between large online firms and small-scale offline bookstores has expanded even more, and the number of new publications has decreased significantly. Another noticeable change since the introduction of the new RPM regulation is the rapid growth of the used book market that has also benefited large online bookstores because the large online firms have dominated the used book market. Theoretically, this study constructs a comprehensive model including the used book market, which has properties of a two-sided market and is linked with the new book market. From the model, this paper explains various effects of new RPM regulations on the used book market.

To my parents,
my wife, Dayeon Lee
and sons, Yejun and Yeonjun

ACKNOWLEDGEMENTS

I am very grateful to my advisor Kyoo il Kim for his constant support, encouragement and helpful feedback. I would also like to thank Jay Pil Choi and Thomas Jeitschko for their invaluable help and advice. This paper also benefited from comments by seminar participants at Michigan State University.

I am deeply indebted to my parents, Samyoung Song and Nasoon Sun, for their enormous support and love. My two sons, Yejun and Yeonjun, became the joy of my life and the source of motivation and encouragement.

Last but by no means least, my wife, Dayeon Lee, has gladly endured a hard time by my side during doctoral program while taking care of two sons with sincere love. Today's accomplishments and the happiness of our family are all thanks to her sacrifice, effort, and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Education and Neighborhood Premium in Housing Price and its Dynamics

1.1 Introduction

The hedonic price regression, using housing market data, has been widely used as the methodology to estimate the implicit prices of local public goods such as school quality, crime rate, and air pollution.[1] Economists, especially, have been trying to estimate households' valuations for better school performance and neighborhood environments, given their relevance to important empirical issues.[2] The primary purposes of estimating the implicit prices of better school quality and neighborhoods can be summarized into two categories. First, estimates illustrate households' valuation for a better performing school and for living around more educated neighbors. Second, estimates allow us to figure out how school quality and neighborhood composition affect micro-geographic patterns of housing prices. Also, estimating households' preferences for schools and neighbors is essential to study school and residential segregation.

In one representative study, Black (1999) examines the parental valuation of elementary education by developing the *boundary discontinuity design* (BDD). The study finds that a 5% increase in elementary school test scores leads to a 2.1% higher housing price. Bayer, Ferreira, and McMillan (2007) embed the BDD in both the sorting model, which considers residential location decision of each household, and the hedonic regression model. The study shows that households are willing to pay less than 1% more for housing when the average test score of the local school increases by 5%. They also find that households prefer to self-segregate based on both race and education.[3]

---

[1] Gibbons and Machin (2008) review the empirical literature on hedonic analysis valuing school quality, better transportation, and lower crime.

[2] Black and Machin (2011), Machin (2011) and Nguyen-Hoang and Yinger (2011) offer comprehensive reviews of this literature. Black and Machin introduce and classify 54 papers studying this topic.

[3] Also, Kane, Riegg, and Staiger (2006) investigate school districts under a court-imposed desegregation order and its effect on housing price. According to their results, 1 SD increases in school quality increases housing price by 10%. Clapp, Nanda, and Ross (2008) show 1 SD increase in students' math scores increases housing price by 1.3∼1.4% while neighborhood characteristic (the portion of Black and Hispanic) becomes more important in the long run. Figlio and Lucas (2004), Fiva and Kirkeboen (2011), Gibbons, Machin, and Silva (2013), and Imberman and Lovenheim (2015) investigated the effect of newly released information such as school report card, school ranking, and school and teacher value-added on the housing price. Boustan (2012) shows the desegregation of public schools in central cities reduced the demand for urban residence, leading urban housing prices and rents to decline by 6% relative to neighboring suburbs.

Reliable estimates for the implicit prices are also required in advance to construct a relevant demand function, which is adopted for counterfactual welfare analyses and policy simulations for various education and school policies. For instance, if a local government considers an increase in expenditure for school quality enhancement in their municipal area, before making the decision, they would compare its planned spending with residents' willingness to pay for better school quality.

However, analyzing the demand for housing is challenging empirical work because a house is a bundle of various types of characteristics. The first type pertains to the physical attributes of a house (size, age, the number of rooms, and type of building). The second attribute concerns local public goods of a community such as education level of the surrounding neighborhood, school quality of the nearby school, transportation systems, crime rate, and air pollution in the area. The third characteristic includes amenities around a house (e.g., supermarkets, shopping centers, sports facilities, hospitals, restaurants, fire and police stations, and etc.), and the fourth type of characteristic is the distance from a house to other locations (e.g., workplaces, schools, and subway stations).

The challenge for economists is that it is impossible to control for all these characteristics in the analysis due to the lack of data, and many of the factors affecting housing price are difficult to measure by numerical value (e.g., curb appeal and the quality of landscaping). Then, such unobservables and omitted variables cause bias problems because they are likely to be correlated with other attributes included in the model. In the case when a cross-sectional dataset is available, instrument variables or the *boundary discontinuity design* approach can be adopted to handle the bias problem. In practice, however, it is hard to find appropriate instrumental variables in the housing market that satisfy exclusion restrictions in the IV framework, and those instruments are not usually free from an ad-hoc selection problem. Exploiting discontinuities at a jurisdictional boundary, the use of BDD has been widely adopted in modern empirical research. To estimate the preference for school quality, the BDD investigates both sides of the narrow area that borders a school zone boundary. In most applications, however, school zone districts are coterminous with administrative districts and geographical discontinuity by a broad road or a river. Also, the possibility of household self-segregation – based on the different school quality – leads to other critical differences (e.g., neighborhood composition) between both sides, while making it hard to obtain a clean variation in school quality.

On the other hand, if a panel dataset with repeated housing transactions is available, we can handle the endogeneity problem by eliminating the time-constant omitted variables from the differenced models. However, another problem in the housing market is that repeated transactions for the same house rarely happen, thereby requiring the use of long-term panel dataset to obtain a sufficient sample size. For this reason, if we use such a long-term panel dataset, the possibility that unobservable information and omitted variables could vary over time precludes the differenced model, solving the bias problem. In particular, when we analyze the housing market of an urban area where many amenities usually change fast compared to rural areas, it is not reasonable to assume that the unobservables and omitted variables are fixed.

The first main methodological feature of this paper, as the solution for the above problems, involves constructing a short-term (two consecutive years) panel dataset. It is more plausible that the unobservables causing the bias problem are fixed over time as the dataset covers a shorter period. From the dataset, this paper analyzes the housing market in Seoul, Korea with assorted amenity information of houses obtained by merging various data sources. The short-term panel dataset – with sufficient sample size – is generated from using a characteristic of the apartments, which is that an apartment building contains many housing units that are almost identical within the same building. The details of this method are explained in Section 1.2.2. Along with constructing a short-term panel dataset, for more robust estimations, this study considers the case that unobservables vary within the two years under the assumption that it follows the first-order Markov process. For the Time-varying Unobservable Model, this paper closely follows the method developed by Bajari, Fruehwirth, Kim, and Timmins (2012).

In the previous literature, housing price hedonic models usually include physical attributes of a house, neighborhood characteristics, other variables of interest according to a topic (e.g., school quality, air quality, and crime rate), and regional- and time-fixed effects to control an omitted variable bias. As noted by Bayer, Ferreira, and McMillan (2007), however, including regional- and time-fixed effects may not be enough to control all omitted variables and unobservable information for an unbiased estimation. According to the study, the inclusion of precise neighborhood socio-demographic information reduces the coefficient of school quality in a hedonic regression by 50%, compared to the model simply including boundary fixed effects. This result implies that it is hard to represent region-specific omitted information and unobservables with regional-fixed effects. For this reason, this study tries to control as much information as possible along with typical regional- and time-fixed effects.[4] The models in this paper also controls the number of similar sized houses within a certain radius from a house to consider the supply side of the housing market.[5] The variables used in this paper are listed in Table 1.1 and Table A1 in the appendix.

Another feature of this study is to adopt a spatial modeling technique using the distance-weight matrices to match a particular house with amenity information. In the housing market literature, amenity information is usually matched with a house by census tract level when the exact location of a house is not identified. However, given that people can benefit from amenities located nearby but in different census areas, using the spatial distance-weight is a more precise way to model the housing markets.[6] To combine the spatial information with hedonic and dynamic models, this study uses the exact locations of houses, schools (elementary, middle, and high school), subway stations, department stores, supermarkets, centroids of census tracts, and air quality monitors by changing the address information into longitude and latitude.

From the newly-assembled panel dataset with the above features, this paper shows the evidence that there exists a long-lasting reputation effect in the implicit price of school quality. When we evaluate a school, it is natural to consider the past records of school performance together with the most recent ones. Let's suppose three schools ("A," "B," and "C") have the same performance records in the most recent year. The

[4] This paper considers regional fixed effect for each of the 11 school zone level, and monthly time fixed effects for each region. As a robustness check, this study also tries generating regional and time fixed effect for 25 district level. The district area is smaller than a school zone. Two or three districts consist of a school zone in Seoul.

[5] Models include the number of similar sized houses within 0.5 km, 1 km and 2 km radius from a house. In this process, the size of house was divided into four categories. i) less than 60 m^2, ii) 60∼100 m^2, iii) 100∼130 m^2, and iv) larger that 130 m^2

[6] More details about distance-weight matrices are explained in Section 1.2.1.

school “A” is a prestigious school which has stable, long-standing good performance records. However, the performance of the school “B” has improved recently and the previous records are not good. Last, the school “C” has no previous records because the school is newly-established. In this case, parents may have a higher willingness to pay to live near the prestigious school than the other two schools, since parents also care about the past records of a school. Even though it is intuitively natural, I could not find any previous research explicitly considering the above reputation effect, which comes from past records of a school. According to the result of this study, people consider the past records of a school up to ten years while placing more weight on recent performances of a school, and the weight fades out as time passes by. The estimated implicit price for one school year result is relatively smaller; however, its accumulated value for long-term good performances of a school becomes more substantial than estimates of prior studies. The previous studies have used the school performance statistics of a specific year without considering previous records, and some studies used a two-year, three-year or four-year average of the school statistics to reduce any year-to-year noise. However, such averaging is different from flexibly considering the above reputation effect; thereby it can cause bias estimation results for the implicit prices of school performances.[7]

The second purpose of estimating households' preferences for schools and neighbors is linked to studies about segregation issues among schools, neighborhoods, and housing prices. Much of the previous literature focuses on residential segregation driven by race. For example, Card, Mas, and Rothstein (2008) find evidence of a tipping point (5∼20% minority share) and study the dynamics of racial segregation within a city.[8] Another category of segregation studies analyzes the effects of neighborhood and school segregation – also by race – on school achievement gaps[9] or school segregation itself from the viewpoint of students' ethnic composition.[10] Card and Rothstein (2007) study

[7] For example, if we use the variable averaging test score results from the past three years, the estimated coefficient of the variable is the same with the result from the model that imposes the restriction – all coefficients for the past three year results are the same when we include each three previous results independently in the model.

[8] About this topic, see also Bajari and Kahn (2005), Bayer, McMillan and Rueben (2004), Bayer, Fang and McMillan (2014), Bayer and McMillan (2012), Easterly (2009), Zhang (2009), Bischoff and Reardon (2014), Reardon and Yun (2001) and Boustan (2010, 2012).

[9] Bifulco and Ladd (2007), Reardon (2011, 2016) and Reardon and Galindo (2009) are also included in this category.

[10] See Reardon, Yun and Eitle (2000), Reardon, Grewal, Kalogrides, and Greenberg (2012), and Cartano and Maheshri (2017).

the effects of school and neighborhood segregation on the relative SAT scores of black students across different metropolitan areas. Lastly, there are a few studies that investigate residential segregation driven by school segregation. Baum-Snow and Lutz (2011) examine the residential location and school choice responses to the desegregation of large urban public school districts. Overall, the common feature of previous literature is a focus on the unilateral effect from one to another ((i) neighborhood segregation $\Rightarrow$ school segregation or school achievement gaps, (ii) school achievement gaps $\Rightarrow$ neighborhood composition, and (iii) school quality or neighborhood composition $\Rightarrow$ housing price) or an analysis of the residential and school segregations driven by race.

Different from previous literature, the contribution of this study is to analyze the bilateral – direct and indirect – dynamic relations among school quality, neighborhood characteristics, housing prices, and other amenities over time, based on the panel VAR (vector autoregression) model and its impulse-response analysis. A coefficient in the hedonic regression is interpreted as an implicit price of a variable and shows the effect on the market price of increasing a particular attribute while holding the other attributes fixed. Hence, we can think of a hedonic model as showing the static picture of a housing market in an equilibrium status. In practice, however, if one of the housing attributes changes, other attributes of the same house could be affected and change over time.

For example, enhanced school qualities may increase housing prices (school quality $\Rightarrow$ housing price) while attracting more educated neighbors around the school (school quality $\Rightarrow$ neighborhood composition) if more educated people prefer a better performing school. Then, the increased neighborhood education level caused by the improved school quality can increase the housing price further (school quality $\Rightarrow$ neighborhood composition $\Rightarrow$ housing price) because a neighborhood composition is another critical housing attribute affecting housing prices. This is the indirect effect of enhanced school qualities on housing price by attracting another good attribute – more educated neighborhoods. It is also likely that a higher neighborhood education level, in the opposite direction, can improve performances of a school (neighborhood composition $\Rightarrow$ school quality) because parental education level may affect their children's academic performance. This is the bilateral relationship between school performances and neighborhood education level (school quality $\Leftrightarrow$ neighborhood composition). Additionally, the increased portion of highly educated people in the neighborhood itself can also attract more educated people in the region (neighborhood composition ⇒ neighborhood composition) due to household self-segregation based on education level. To summarize, there could be bilateral and indirect relationships among important housing attributes over time. Bayer, Ferreira, and McMillan (2007) refer to these effects as 'second-round social multiplier.' To the best of my knowledge, this is the first paper analyzing the above dynamics in the housing market.

Recently, there have been some studies analyzing the dynamics of the housing market. Bayer, Keohane, and Timmins (2009) examine the dynamics of households' residential choice considering moving cost and find that the estimates are three times greater than the marginal willingness to pay, which was estimated by a conventional hedonic model. Bayer, McMillan, Murphy, and Timmins (2016) introduce the dynamic of the neighborhood choice problem through two channels: wealth accumulation and moving costs. In their setting, households are treated as making a sequence of location decisions that maximize the discounted sum of expected per-period utilities. That is, the dynamic analysis in the previous literature is based on the theoretical model considering dynamic decisions such as maximizing the households' lifetime utility from the given utility function.

From the short-term panel regression model and dynamic model, this paper finds evidence that there exists a long-lasting reputation effect in the implicit price of school quality. The second finding comes from the results of price gap decomposition analysis. Similar with other metropolitan cities, residential segregation also has been occurring in Seoul. In Seoul, the Gangnam area is preferred to the other regions for many reasons. Since there is no racial difference in Seoul, the residential segregation is driven by parents' education level, income, and wealth. Regarding neighborhoods' education level, school performance, and average housing price (as a proxy for incomes and wealth of an area), there are substantial differences between Gangnam and the other districts in Seoul. According to the estimation results of this study, more than 43% of the housing price gap between the areas is explained by the differences in education and neighborhood environments.[11] Based on the estimation results, this study simulates

[11] The top university entrance rate (Seoul National University), the portion of BA or above degree holder

an education policy that adjusts school zone boundaries. The government of Seoul is now considering school zone adjustment as an option for immoderate housing price gaps and education performance gaps within the metropolitan area. The simulation predicts the effects of school zone change binding two different areas (one of the best performing areas and a below average area) that are geographically adjacent to each other. In the new school zone, the housing price gaps are estimated to be diminished by 13% in the long-term.

The third finding comes from the dynamic model. This study finds evidence that positive bilateral dynamic relations among school quality, neighborhood education level, amenities, and housing price exist. That is, those housing attributes turn out to affect each other making the effects of school performance and neighborhood composition more significant by bilateral relationships over time in a dynamic framework. These results imply the impact of school performance and neighborhood composition on housing price and city segregation can be underestimated in a static model if dynamic effects are ignored.

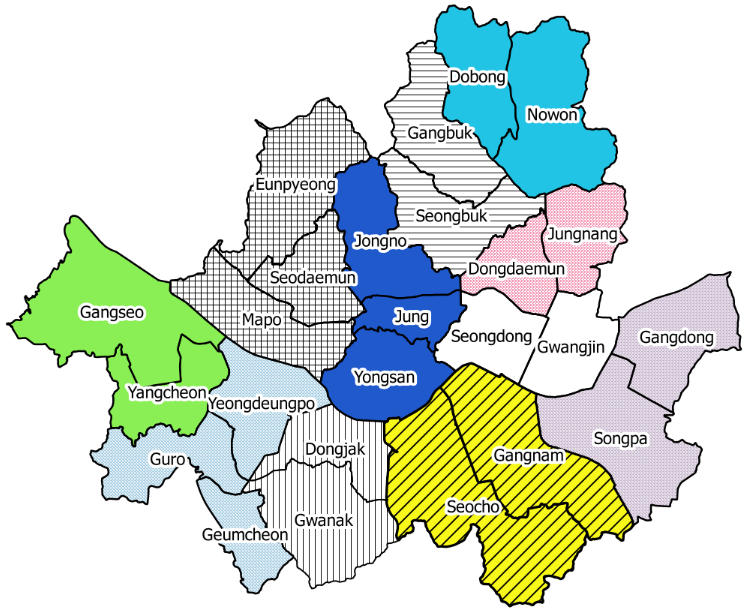
This paper proceeds as follows: Section 1.2 describes the data and explains the panel dataset generating process. Section 1.3 presents results from the models which allow unobservables to vary over time. Section 1.4 shows results from the policy simulation (school zone adjustment). Section 1.5 analyzes dynamics of the housing market by using the panel VAR approach. Section 1.6 concludes.

1.2 Data

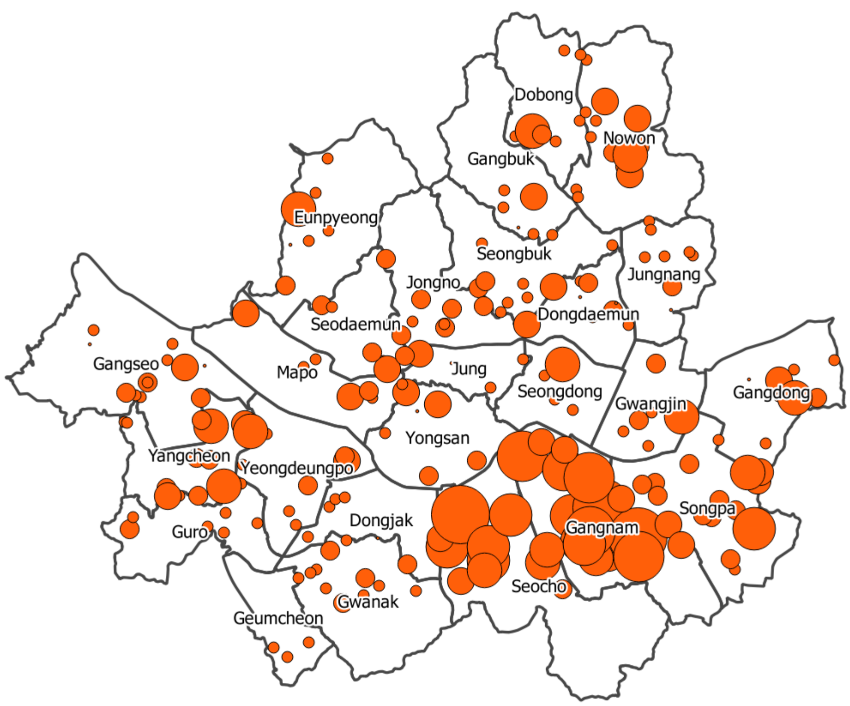
This section describes the main data sources and explains how they are matched with the housing transactions in the Seoul metropolitan area in Section 1.2.1. For the matching process, this paper adopts a spatial modeling technique frequently used in a GIS analysis. Section 1.2.2 shows the process of constructing the panel dataset.

As the spatial summary, Figure 1.1 and 1.2 show the primary statistics of this paper on the map of Seoul. Figure 1.1a shows the school zone map of Seoul. Areas with the same color and pattern represent the same school zone. There are 25 districts

in 40∼49 age group, and private education environment are the three components that explain almost 50% of the housing price gaps in Seoul.

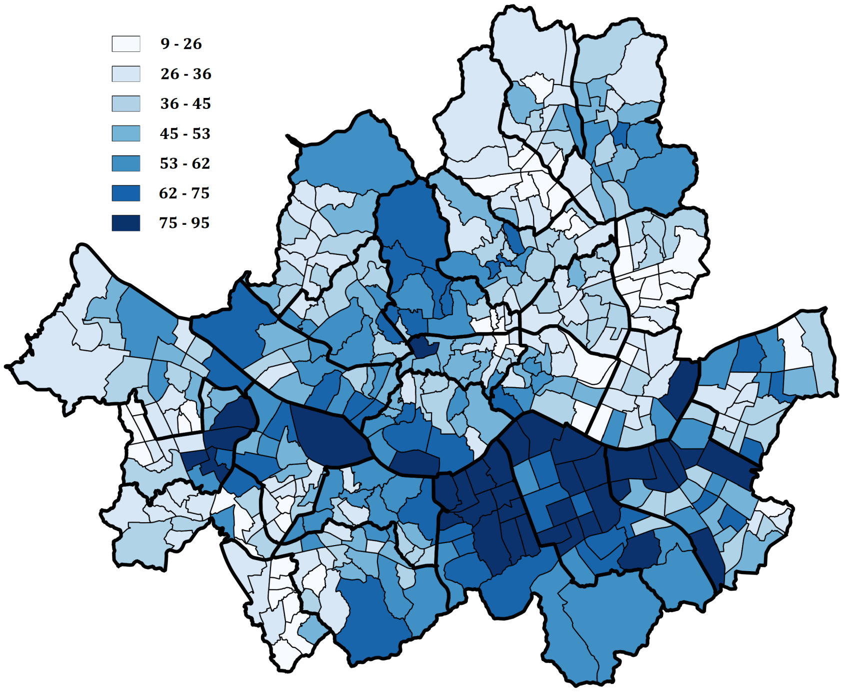


(a) School Zone Map

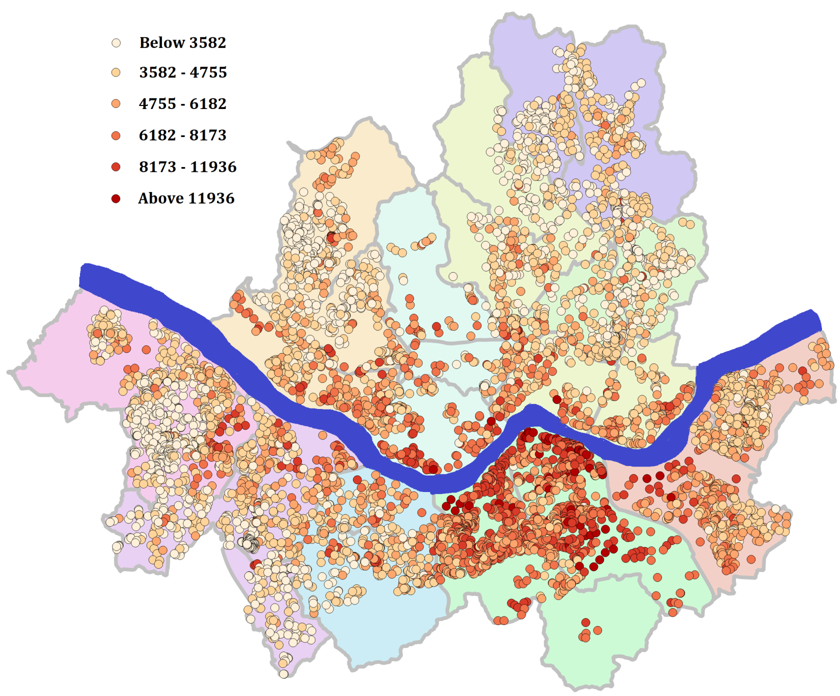


(b) Top University Entrance Rate

Figure 1.1 School Zone and School Performance Gap



(a) University Graduates (%, Age 40∼49)



(b) Housing Price per Size (\$/m^2)

Figure 1.2 Neighborhood Segregation and Housing Price Gap

and 11 school zones in Seoul. The Seocho and Gangnam districts form the Gangnam school zone (the area with oblique lines). In terms of housing price, neighborhood composition, and school performance gaps, the rest of the figures show the substantial differences between the Gangnam school zone and the other areas of Seoul. In Figure 1.1b, each circle represents the location of a high school, and the size of the circle is proportionate to the top university entrance rate of a high school.[12] Figure 1.2a shows the percent of university graduates (BA or above degree holders) in the 40∼49 age group in the 423 census tracts. As the color gets darker, it means a higher portion of well-educated neighbors reside in the area. Figure 1.2b represents the housing price per size (m^2). Each circle identifies the location of a house, and darker points mean more expensive houses. Note that large circles (Figure 1.1b), dark areas (Figure 1.2a), and dark points (Figure 1.2b) are highly concentrated in the Gangnam school zone.

To clearly illustrate the gaps between regions within Seoul, Figure A1 and A2 in the appendix show the locations of top performing high schools (Figure A1b), census tracts where the portion of university graduates is higher than 62% (Figure A2a), and the locations of expensive houses within the top 5% (Figure A2b). Also, to compare both areas, Table 1.1 and A1 show summary statistics of variables for all areas of Seoul, and for Gangnam and the other areas separately. The next paragraphs explain the main data sources where the variables in this paper are drawn from to construct the panel dataset.

A. Housing Transaction and Location Data

The Ministry of Land, Infrastructure, and Transportation Department of Korea provides transaction data on houses.[13] This dataset contains information about the real transaction price, housing type (apartments or single detached house), specific address, size, construction year, and transaction date. In the case of apartment transactions, it also includes the building name or brand of apartments and the floor on which the apartment is located. Table 1.1 – (A.1) describes the summary statistics of the physical attributes of a house (price, floor, age, size, complex, and brand).[14] From the

[12] In this paper, the top university entrance rate means the portion of students who entered Seoul National University (SNU) by high school.

[13] http://rt.molit.go.kr

[14] The details about complex (the type of apartment) and brand of apartments are explained in the data

specific address information, I generate distance variables from a house to some important places and amenities. Table 1.1 – (A.2) shows the distance variables used in this study. The distances from a house to the nearest school (high, middle, and elementary school),[15] subway station, Han-river, department store, and a large supermarket chain are used as distance variables. More detailed explanations about some unique characteristics of the housing market of Seoul, Korea are in the data appendix.

B. Population and Housing Census

The Population and Housing Census data is used to obtain neighborhood demographic information surrounding a house. From the data, I calculate the portion of people by education level and age group, such as the portion of university graduates (BA or above degree holders) in the 40∼49 age group by each census tract. This study uses the portion of BA or above degree holders in the 40∼49 age group as the primary index for neighborhood education level. The first reason is that most people in this age group are school parents who have a stake in the educational environment. Second, due to the rapidly increasing trend in the portion of BA or above degrees in Korea, the BA degree no longer represents a high education level in younger age groups. Also, this paper uses the population of all residents, students, people aged 35∼55, and those over 65 and calculates each portion by census tract. From the census data, I use the number of households by house type and its composition (lives alone, a married couple, and a family with children) by census tract. The above neighborhood characteristics are matched with a house by adopting the distance-weight matrices that use the distances from a house to centroids of census tracts. This paper explains the matching method in Section 1.2.1.

C. Business Census

The Business Census offers information about the number of businesses by 228 industry classifications and the number of employees by the types of employment (full-time,

---

appendix.

[15] I calculated the distances to the third nearest middle and high schools because the school allocation system does not guarantee assignment to the nearest middle and high school from their house. Meanwhile, in the case of an elementary school, there exists one-to-one matching between a house and a school. More details are in Section 1.2.1

Table 1.1 Summary Statistic

A. House Attributes & Distance	All			Gangnam			The Other		
	Average	Median	SD	Average	Median	SD	Average	Median	SD
A.1 Apartments Attributes									
Price ($)	426,655	357,273	282,497	752,714	680,000	409,688	357,002	327,273	182,854
Floor	8.83	8.00	5.93	8.78	7.00	6.65	8.84	8.00	5.77
Age (Year)	16.6	15.0	8.9	19.2	18.0	10.9	16.1	15.0	8.3
Size (m^2)	77.96	83.86	28.58	87.83	84.79	38.53	75.86	82.53	25.47
Complex (0 or 1)	0.03	-	0.18	0.12	-	0.32	0.02	-	0.13
Brand (0 or 1)	0.15	-	0.36	0.14	-	0.35	0.15	-	0.36
A.2 Distance to (km)									
High School_1	0.73	0.64	0.44	0.63	0.56	0.41	0.75	0.67	0.45
High School_2	1.14	1.07	0.49	1.07	0.97	0.51	1.15	1.09	0.49
High School_3	1.49	1.43	0.55	1.41	1.34	0.53	1.51	1.44	0.55
Middle School_1	0.49	0.44	0.29	0.49	0.43	0.31	0.49	0.44	0.28
Middle School_2	0.84	0.78	0.37	0.85	0.76	0.40	0.84	0.78	0.36
Middle School_3	1.13	1.07	0.44	1.11	1.03	0.53	1.13	1.07	0.42
Elementary School	0.33	0.30	0.16	0.35	0.33	0.19	0.32	0.29	0.16
Subway	0.60	0.52	0.38	0.50	0.46	0.28	0.62	0.53	0.40
Han-river	4.47	3.51	3.70	2.43	2.62	1.59	4.91	3.97	3.88
Department Store	2.39	2.15	1.43	1.87	1.60	1.18	2.50	2.25	1.45
Supermarket	1.35	1.19	0.83	1.72	1.59	0.92	1.27	1.11	0.79
A.3 New House Supply									
Radius : 0.5 km	14.32	-	116.35	38.86	-	238.20	9.27	-	66.99
Radius : 1 km	37.32	-	176.84	80.77	-	326.31	28.37	-	123.73
Radius : 2 km	131.96	-	333.97	223.41	-	497.03	113.12	-	285.55
Obs.		199,673			35,146			164,527	

B. School Performance	All			Gangnam			The Other		
	Average	Median	SD	Average	Median	SD	Average	Median	SD
B.1 High School									
Top Univ Entrance (persons)	3.51	2.00	4.55	9.40	8.00	6.55	2.12	1.00	2.32
Top Univ. Entrance Rate (%)	0.81	0.54	1.00	2.04	1.76	1.46	0.53	0.36	0.53
N-SAT	100.8	101.3	19.2	109.8	114.4	15.5	98.6	99.2	19.4
Students per School	384	384	105	442	445	79	370	355	106
Students per Class	31.1	31.4	4.3	33.4	33.7	4.1	30.6	31.2	4.1
Students per Teacher	14.3	14.4	2.2	15.6	15.7	2.2	14.0	14.1	2.1
Full-time Teacher (%)	85.1	86.1	6.6	83.4	84.8	8.7	85.6	86.3	5.9
Obs.		220			42			178	
B.2 Middle School									
Top High School (persons)	5.8	4.7	4.2	7.7	7.3	4.2	5.4	4.3	4.1
– Entrance Rate (%)	2.02	1.90	1.03	2.51	2.44	1.08	1.92	1.77	0.99
Private High School (persons)	23.9	18.0	19.6	36.5	29.0	24.1	21.3	16.3	17.4
– Entrance Rate (%)	8.57	7.10	6.29	12.11	10.60	6.98	7.83	6.16	5.87
Obs.		379			66			313	

C. Demographics	All			Gangnam			The Other		
	Average	Median	SD	Average	Median	SD	Average	Median	SD
BA or Above 40s (%)	46.5	43.7	18.3	68.1	68.4	18.5	42.3	41.4	14.9
BA or Above 50s (%)	32.8	27.8	18.5	56.8	56.6	21.2	28.2	24.9	13.8
Population	22,616	21,867	9,045	23,300	21,611	7,497	22,394	21,867	9,193
Average Age	40.8	40.8	2.1	39.7	39.4	1.6	41.1	41.0	2.1
Portion of Student (%)	13.3	12.9	3.7	15.1	14.6	4.2	13.0	12.6	3.5
Portion of 35∼55 Age (%)	32.5	32.7	2.8	33.9	33.8	2.5	32.3	32.6	2.7
Portion of 65 Age Above (%)	13.0	13.0	2.6	11.1	10.6	2.3	13.3	13.3	2.5
Full-time Employee	8,174	3,524	15,024	16,935	9,252	19,659	6,717	2,986	13,673
Part-time Employee	1,559	914	2,003	3,316	2,731	3,205	1,257	830	1,543
Self-Employee	1,794	1,452	1,632	1,917	1,725	1,327	1,784	1,451	1,686
Obs.		423			66			357	

part-time, and self-employed) for each business by census tract level. There are 423 census tracts in Seoul, and the average area is $1.4km^2$ ($0.54mile^2$). Given that housing prices are affected by the amenities surrounding the house, among the 228 industry classifications, this paper selected several classifications that could affect the housing price. Table A1 shows the list of selected variables from the Business Census data, and the employment information of businesses is presented in Table 1.1 – (C). The information assembled from the Business Census is also matched with a house by using the distance-weight matrices as described in Section 1.2.1.

D. Education Information

As variables of interest, this study uses the top university (Seoul National University) entrance rates and the results of Nationwide Scholastic Achievement Test (N-SAT) by a high school and entrance rates for private and top high schools by each middle school.[16] In addition, school characteristics such as the number of students per class and per teacher, as well as the portion of full-time teaching staff are controlled in the analysis. In most previous literature analyzing the implicit price of school performance, researchers usually control the characteristics of a school along with the test scores, such as the racial composition in a school (e.g., White, Black, Hispanic, and Asian) and the portion of international students. However, racial diversity does not exist as such in Korea,[17] thereby it offers a better environment to reveal the households' willingness to pay for better school performances. Another critical factor in the educational environments is the distance between a house and a school. This paper uses the latitude and longitude information of houses and schools and calculates the exact distances from houses to schools. This paper also uses the number of private academies by census tract from the Business Census as a proxy for the private education environment. There were 13,149 private academies preparing various school exams

[16] In the data appendix, this paper explains further information about these statistics. In this study, I use "top high school" instead of its official name "special-purpose high school." In general, only the top 2∼3% ranked students in each middle school enter special-purpose high schools.

[17] The percent of other race and foreigner is negligible in Seoul. According to the statistics published by Seoul Metropolitan Office of Education (https://kess.kedi.re.kr), the portion of international students was 0.42% (elementary), 0.15% (middle school) and 0.09% (high school) among all students in Seoul. Moreover, most of them attend separate international schools for foreigners. Thus, the portion of international students in general schools is close to zero.

as well as university entrance exams in Seoul in 2015. According to a recent survey (The Private Education Expenses Survey) by Statistics Korea, 75% of students in Seoul (more than a million) attended private academies after school.[18] As a result, private academies also become another important factor affecting the educational environment.

E. Air Quality and House Supply

This paper additionally assembles information about the air quality around a house. Many recent studies show that households are willing to pay to avoid air pollution. According to Bajari, Fruehwirth, Kim, and Timmins (2012), a home buyer would be willing to pay 0.36% of the housing price to avoid 1 $\mu g/m^3$ increase in PM10 (particles less than 10 micrometers in diameter) concentration.[19] This paper uses the average annual PM10 concentration ($\mu g/m^3$) statistic, which has recently been classified as Category 1 carcinogen by the World Health Organization (WHO) and is also known to cause various health problems such as asthma, bronchial pneumonia, and heart attacks. The PM10 concentration is measured at 25 monitors in Seoul.[20] Additionally note that this study also controls for the number of new housing units to consider the supply side of the housing market. To be specific, each model controls for the number of similar sized apartment units within a certain distance (0.5 km, 1 km, and 2 km) from a particular house.

1.2.1 Data Matching

For the estimation of the implicit prices of housing attributes, all the assembled data should be matched with an individual house. For the matching process, unlike most of the previous literature, this study uses spatial weight matrices used in a recent GIS analysis. When the location of a house is identified by census tract level resolution, researchers usually match the amenity information to a house based on the census tract level (this study defines this method as the simple matching method). Then, identical

[18] Also, according to the survey, on average, a student spends $ 5,180 for the private education in a year.

[19] Sieg, Smith, Banzhaf, and Walsh (2004), Chay and Greenstone (2005), Bayer, Keohane, and Timmins (2009), Tra (2010), Grainger (2012), Bento, Freedman, and Lang (2015), and Hamilton and Phaneuf (2015) examine this topic.

[20] The Korea Environment Corporation publishes this data and it is available at https://www.airkorea.or.kr

information is matched with all the houses located in the same census tract. However, if the specific location of a house is identified, better methods can be considered for the matching technique.

To illustrate the problem of simple matching, we can suppose the following example. If a house is located in "A" census tract but is close to the boundary line of "B" census tract, then the amenities in "B" could also affect the housing price in "A." However, this reality – people can take advantage of amenities near their house irrelevant from census boundary – is ignored when the simple matching method is adopted.[21] To solve the problem of interrupting the space created artificially by census tract boundaries, this study adopts various distance-weight matrices for the matching between surrounding amenity information and a house. This method uses amenity information by the census tract level while merging the information with a house by considering distances. That is, amenities near a house have larger weights than those far away from the house. The neighborhood and amenity information (variables in Table 1.1 – (C) and Table A1 – (D)) is matched with a house using this method.

This paper adopts different distance-weight matrices for the matching between school information and a house to consider school zone and the student allocation systems of Seoul. Note that one-to-one mapping does not exist between a residential location and a school in Seoul, since there are multiple middle and high schools in the same school zone.[22] The proximity of a residence to a school and students' preference are primarily considered in the student allocation process; however, there is no guarantee that students are assigned to the school nearest to their houses. To consider the above properties, this paper uses the distance-weight method for matching between a school and a house. By using this method, a school near a house has a larger weight than those far away from the house in the data matching process. There are several reasons for justifying the use of this method. First, even though uncertainties exist in the student allocation process, it is more likely students are allocated to the nearest

[21] I could find some extreme cases that reveal this problem. For instance, according to the Business Census 2015, there were 817 private academies in the Dachi census area while there were 37 in the Yeoksam census area. The distance between the centroid of both areas is less than 1 km. Accordingly, people in the Yeoksam area can benefit from facilities located in the Daech area, and those facilities may also affect the housing price in the Yeoksam area.

[22] In the case of an elementary school, however, there exists one-to-one matching according to the address of a house.

school since the allocation system considers the distances from a house to schools. Second, when considering students' commute cost, the performance of a school far away from a house should be discounted compared to a nearby school with the same performance statistics. Third, when households decide where to live, they consider both the possibility of being allocated to the school where they want their child to attend and the ease of commute. Thus, they tend to live near a school where they want to apply, or to be assigned while considering the performances of schools near the house. This paper argues the school information is appropriately matched with a house while considering the properties of the school zone and student allocation systems by using the distance-weight matching method. More details about school zone and the student allocation systems along with the matching methods are explained in the data appendix.

For the weights, this study adopts the inverse-squared distance-weight matrices using the distances from a house to i) centroids of census tracts (for the matching of neighborhood demographics and amenity information), ii) schools (for the matching of school information)[23], and iii) air quality measuring monitors. For robustness checks, this study also tries other weight matrices with some cut-off radius distances for the matching of neighborhood and amenity information surrounding a house, and the N-nearest information matching methods for school statistics.[24]

1.2.2 Panel Dataset Generating Process

As discussed in the introduction, the problem of housing panel dataset is that it inevitably covers the long-term to have a sufficient sample size because repeated transactions for the same house are rare events. Using the long-term panel dataset, however, makes the panel regression (using mean-differenced or first-differenced models) less reliable because the analysis is based on the unreasonable assumption that unobservables or omitted attributes of a house – which cause the bias estimation – stay fixed for the

[23] There are two types of weight matrices (one for high schools and another weight matrix for middle schools). The distance weight matrices are changed over time due to a new establishment of a school.

[24] This paper additionally tried 1 km and 2 km radius cut-offs from an individual house for generating the distance-weight matrix. For the matching of school statistics, the three-nearest school matching method is adopted for distance-weights along with the overall weights using statistics of all schools within the same school zone.

long-period that the panel dataset covers.

To overcome the above problem, this study generates a short-term panel dataset by using the apartments' characteristic that an apartment building includes multiple and almost identical housing units in the same building. Given this property, it is reasonable to assume a group of equal sized housing units in the same apartment complex as identical houses. In this study, if apartments have the same address, size, construction year, and are located in the same floor group, they are treated as the same housing unit in the panel dataset.[25] A critical point for using this method to generate a short-term panel dataset is that it makes much more frequent repeated transactions in a given period than using the repeated transactions of exactly the same house, since multiple housing units are included in the same group.

As a possible problem of this method, there could be some differences within an identical group of apartments owing to renovations and interiors. However, the same issues happen even when we use the panel dataset using the repeated transactions of the same single detached house because such renovations and interiors are also mainly hidden information for researchers. On the other hand, there are some advantages of using this method. First, apartments have more restrictions than the single detached house for renovations because a unit is a part of the entire building. For instance, structural renovations such as the extension of house size or making more rooms are not possible due to safety restrictions. Also, enhancement in curb appeal could be another critical change in the unobservable attribute of a single detached house; however, this is not the case with apartments. Put differently, apartments have less possibilities for some fundamental changes in unobservable attributes that make the panel analysis less reliable compared to a single detached house.

This study also uses different numbers of transactions for each group of apartments as the sample weight, and the different weights act as an additional control to mitigate the possible differences in unobservable attributes within a group of apartments. For example, "A" group of apartments has 30 transactions while "B" group of apartments has three transactions within a year. Then, the weight of observation "A"

[25] A floor is divided into five groups. Group 1 includes the first floor and lower. Group 2: floors 2∼5, Group 3: floors 6∼10, Group 4: floors 11∼20, Group 5: floors higher than 20. Also, this paper also tried other group compositions; however, there were no meaningful differences in the results.

is ten times larger than the weight of observation "B" in the panel regression analysis. Since some unobservables – such as renovation, interiors, and bargaining power between a buyer and a seller – can be different within the same group of apartment transactions, it is reasonable to assume that an observation generated from more frequent transactions reveals a more accurate price of the group of apartments. In this respect, the different number of transactions are used as weights to consider the accurate price of a group of apartments.[26]

It is important to note the limitation of this method to generate a panel dataset. This method can be justified only when the apartment transaction samples represent the whole housing market of the area. If the portion of apartments among entire houses is low, and if the implicit price of housing attributes vary depending on the house type, there may be a sample bias problem. For instance, we could expect that a household with children would have a higher willingness to pay for better educational environments than a household without children. According to the Housing and Population Census 2015, due to the high population density, less than 15% of households live in a single detached house in Seoul and almost 90% of housing transactions come from apartments.[27] Also, the portion of families with children among apartment residences (74%) is similar to that of single-detached house residences (67%). Based on the above statistics, this study assumes that apartments can represent the entire housing market of Seoul.

1.3 Hedonic Price Regressions

In this section, this paper uses a hedonic price regression to investigate education and neighborhood premiums in housing prices. For the analysis, the following baseline model is used to estimate the implicit prices for educational and neighborhood environments:

$$log(price)_{ijt} = W_{it}X_{it}\beta_t + Z_i\phi_t + Dist_{ist}\delta_t + D_{jt} + \xi_{it} \tag{1}$$

[26] In the process for generating the price of a group of apartments in a specific year, this study controls monthly-regional fixed effect to consider monthly price fluctuation by each region. For more technical details, refer to the data appendix.

[27] The population density of Seoul (16,204/km^2) is higher than New York City (10,725/km^2)

where $log(price)_{ijt}$ is the log of the transaction price of house i in a school zone j at time t. X_{it} includes time-varying variables such as amenities, neighborhood demographics, various education statistics, air pollution observations around a house i, and the number of new housing units within a certain radius from a house i at time t. W_{it} is the spatial weight matrix and it has different forms for amenity information (W_{it}^{a}), school statistics ($W_{it}^{s,m}$ (middle school), $W_{it}^{s,h}$ (high school)), and air pollution levels (W_{it}^{p}). Z_i includes time-constant variables such as physical attributes of a house (size, floor, age, brand, and type of house) and distances to a fixed location.[28] $Dist_{ist}$ includes various distances from house i to location s at time t.[29] D_{jt} represents the school zone fixed-effect (11 school zones by monthly level), thereby this study flexibly allows the regional fixed-effect to vary separately over time by each school zone. The role of school zone fixed-effect is to control housing price fluctuations by macro factors – speculation and asset investment decision by forecasting future prices – that are irrelevant to changes in housing attributes. Lastly, ξ_{it} denotes the omitted or unobservable attributes of the house i at time t.

By using two consecutive years of data and additionally assuming that the unobservable attributes of a house and the implicit prices for time-varying variables do not change over two years ($\beta_{t'} = \beta_t = \beta, \;\; \delta_{t'} = \delta_t = \delta, \;\; \xi_{it'} = \xi_{it}, (t' > t)$), the equation (1) can be re-written as the following model by first differencing:

House Fixed-Effect Model[30]

$$\Delta log(price)_{ij} = (\Delta W_i X_i)\,\beta + (\Delta Dist_{is})\,\delta + Z_i\,(\phi_{t'} - \phi_t) + \Delta D_j \tag{2}$$

This study also flexibly allows the unobservables to vary within the two consecutive years given the assumption that it follows the first-order Markov process:

$$\xi_{it'} = \gamma\,\xi_{it} + \eta_{it'} \quad (t' > t) \tag{3}$$

[28] Straight distance from a house i to the Han-river.

[29] $s \in$ {the nearest school (high, middle, and elementary school), subway station, department store, supermarket}

[30] Note that the coefficient of time constant variable (Z_i) shows the estimated change in the implicit price: $\phi_{t'} - \phi_t$

here, $\gamma\xi_{it}$ is the expected value of the unobservable attributes at time t', and $\eta_{it'}$ is the stochastic innovation in the unobservable attributes. Note that this study assumes $E[\eta_{it'}|I_t]= 0$, where I_t denotes the available information at time t. This assumption is based on the efficiency house market by Case and Shiller (1989) implying that it is impossible to earn excess returns by using the available information at time t in the housing market. Then, equation (1) can be re-written as follows by using the Markov process assumption. For the Time-varying Unobservable Model, this paper closely follows the method developed by Bajari, Fruehwirth, Kim, and Timmins (2012).

Time-varying Unobservable Model

$$\begin{aligned}
log(price)_{ijt'} &= W_{it'}X_{it'}\beta + Z_i\phi_{t'} + Dist_{ist'}\delta + D_{jt'} + \xi_{it'} \qquad (4)\\
&= W_{it'}X_{it'}\beta + Z_i\phi_{t'} + Dist_{ist'}\delta + D_{jt'} + \gamma\xi_{it} + \eta_{it'}\\
&= W_{it'}X_{it'}\beta + Z_i\phi_{t'} + Dist_{ist'}\delta + D_{jt'}\\
&\qquad + \gamma[log(price)_{ijt} - W_{it}X_{it}\beta - Z_i\phi_t - Dist_{ist}\delta - D_{jt}] + \eta_{it'}\\
&= (D_{jt'} - \gamma D_{jt}) + \gamma log(price)_{ijt} + [W_{it'}X_{it'} - \gamma W_{it}X_{it}]\beta\\
&\qquad + [Dist_{ist'} - \gamma Dist_{ist}]\delta_t + Z_i(\phi_{t'} - \gamma\phi_t) + \eta_{it'}
\end{aligned}$$

This paper also considers the case that $X_{it'}$ could be correlated with $\eta_{it'}$ and use two-stage nonlinear least squares (2SNLS) to recover parameters in equation (4) under the above assumptions. In the first stage, all the exogenous variables and predetermined variables are used as instruments to replace $X_{it'}$ with its projected value. Intuitively, the above Time-varying Unobservable Model uses the housing price information of the previous period to impute the value of unobservable attributes of the house. For example, if the housing price of i is unusually high after controlling the variables (X_i, Z_i, $Dist_{is}$, D_j) in the model, it means the value of unobservable attributes of the house ξ_i – observed by the buyer of the house – is high. From this, the value of the unobservable attributes is used as additional information to analyze the housing price of the next period.

By using the above models, this paper analyzes the long-term reputation effect of school performance on household implicit price in Section 1.3.1, and Section 1.3.2

examines the implicit price for better educational environments and shows how much of housing price gaps within a city are explained by the differences in educational environments.

1.3.1 School Reputation Effect

People naturally consider both the past records of a school and the most recent information when assessing the quality of a school. In this respect, however, I could not find any previous studies that explicitly consider the 'reputation effect,' which comes from past records of a school. To the best of my knowledge, all previous research analyzing the willingness to pay for a better educational quality use a one-year statistic or averaged statistics over a short-term (e.g., averaged over 2∼4 years), not to consider the reputation effect, but to reduce any year-to-year noise in the school quality variable.[31]

In this section, I show that there exists a long-lasting reputation effect in willingness to pay for the better performance of a school, and also that estimating an implicit price by using an one-year or averaged statistics over the short-term can be biased. Suppose a model uses a three-year averaged test score of a school as the variable of interest. In this case, the model implicitly assumes the implicit prices are all the same for the test scores of the most recent three years.[32] In practice, however, people can consider the history of school performance longer than three years while placing more weight on the recent results than the past results.[33] As such, the estimated model should include longer past records than three years without restriction on coefficients for unbiased estimations.

Table 1.2 shows various results about the school reputation effect. All models (column (1)∼(12)) include all the other variables listed in the summary Table 1.1 and A1 in addition to the top university entrance rate for the past ten years represented in

[31] 2-year average: Bayer, Ferreira, and McMillan (2007). 3-year average: Gibbons and Machin (2003), Kane, Riegg, and Staiger (2006), Clapp, Nanda, and Ross (2008). 4-year average: Reback (2005), and most other research use merely one-year statistics.

[32] Let's suppose the true model is: $Log(Price)_t = \beta_t S_t + \beta_{t-1} S_{t-1} + \beta_{t-2} S_{t-2} + \delta X_t$ (S_t denotes test score of a school at time t, and X_t contains other control variables), and the fitted model is: $Log(Price)_t = \alpha \cdot AS_t + \delta X_t$ (here, $AS_t = \dfrac{S_t + S_{t-1} + S_{t-2}}{3}$, the three-year averaged test score) Then, $\alpha = \beta * 3$ under the restriction that $\beta_t = \beta_{t-1} = \beta_{t-2} = \beta$

[33] In the equation, if $\beta_t > \beta_{t-2}$ (people think the recent record (S_t) is more important than the past record (S_{t-3})), and if the true model includes until S_{t-i} ($i > 3$), the estimated α should be biased.

Table 1.2 School Reputation Effect

Dependent Variable = Log(Price)

Panel A : Including all previous (10 years) results of schools

	Cross Section		House Fixed		Time-varying	
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Top Univ. Entrance Rate (t)	8.370***	7.232***	2.068***	1.528***	2.099***	1.503***
	(0.597)	(0.675)	(0.314)	(0.363)	(0.331)	(0.273)
$(t-1)$	0.0311	-1.601*	3.270***	1.928***	3.693***	2.159***
	(0.621)	(0.748)	(0.301)	(0.322)	(0.318)	(0.349)
$(t-2)$	-2.999***	-4.138***	2.040***	1.908***	2.577***	2.021***
	(0.628)	(0.747)	(0.265)	(0.363)	(0.252)	(0.259)
$(t-3)$	3.040***	0.663	2.080***	1.596***	2.661***	1.812***
	(0.530)	(0.637)	(0.240)	(0.326)	(0.274)	(0.314)
$(t-4)$	5.312***	7.772***	1.229***	1.728***	1.856***	1.63***
	(0.518)	(0.594)	(0.219)	(0.281)	(0.255)	(0.359)
$(t-5)$	2.954***	6.173***	1.316***	1.297***	1.690***	1.674***
	(0.605)	(0.645)	(0.315)	(0.340)	(0.318)	(0.292)
$(t-6)$	-2.156***	-3.393***	1.239***	1.180**	1.499***	1.016***
	(0.556)	(0.617)	(0.245)	(0.372)	(0.237)	(0.345)
$(t-7)$	-4.193***	-4.587***	0.479*	0.801**	0.814***	0.727***
	(0.534)	(0.616)	(0.194)	(0.246)	(0.198)	(0.224)
$(t-8)$	-2.965***	-3.028***	0.635**	0.157	0.590**	0.331
	(0.428)	(0.507)	(0.207)	(0.276)	(0.223)	(0.263)
$(t-9)$	2.525***	2.396***	0.786***	0.662**	0.861***	1.069***
	(0.502)	(0.552)	(0.177)	(0.226)	(0.184)	(0.215)
Sum($t \sim t-7$)	**9.920*** **	**7.487*** **	**13.720*** **	**11.965*** **	**16.888*** **	**12.541*** **
	(1.089)	(1.093)	(1.473)	(1.754)	(1.567)	(1.643)
Sum($t \sim t-9$)	**10.360*** **	**8.120*** **	**15.141*** **	**12.784*** **	**18.339*** **	**13.942*** **
	(1.345)	(1.343)	(1.632)	(1.971)	(1.761)	(1.874)
γ					0.961***	0.976***
					(0.003)	(0.002)
W/ All lags		**Yes**		**Yes**		**Yes**
Obs.	144,492	144,492	72,246	72,246	72,246	72,246
Adj R	0.872	0.886	0.183	0.230	0.991	0.993

Panel B : Including previous 3-year averaged results of schools

	Cross Section		House Fixed		Time-varying	
Variable	(7)	(8)	(9)	(10)	(11)	(12)
Top Univ. Entrance 3-Year Average	**9.358*** **	**8.366*** **	**2.259*** **	**0.832**	**2.562*** **	**0.914**
	(0.964)	(0.981)	(0.518)	(0.688)	(0.824)	(0.589)
W/ All lags		**Yes**		**Yes**		**Yes**
Obs.	144,492	144,492	72,246	72,246	72,246	72,246
Adj R^2	0.869	0.884	0.174	0.225	0.987	0.993

Standard errors clustered at IDs are in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: All coefficients are reported as β*100. All models include fixed effects (11 school zones by monthly level) and the other control variables listed in Table 1.1 and A1 in the model.

Table 1.2. Columns (3), (4), (9), and (11) represent the results from the House Fixed-Effect Model using the equation (2), and columns (5), (6), (11), and (12) show the estimation results from the equation (4), the Time-varying Unobservable Model. Each column in Panel A is the result when the model includes the past records of a school respectively until ten years ago, while Panel B shows the results when each model includes only the three-year averaged statistics of a school instead of ten variables represented in the Panel A. As other explanatory variables, regression models use the same variables for the comparison between Panel A and B. All models in columns (1) to (12) also control monthly time fixed effects by 11 school zones.

According to the results, in the cross-section model ((1) and (2)), positive and negative coefficients are mixed, and it is hard to find any patterns among the implicit prices of the past records. However, in the results from the House Fixed-Effect Model ((3) and (4)) and the Time-varying Unobservable Model ((5) and (6)), all coefficients show a reasonable positive sign, and the size of the coefficients becomes smaller as time passes. These results can be interpreted as people evaluating a school with higher weight placed on recent results, while also taking into account the past records. It is important to note that the implicit price of the top university entrance rate for each year is relatively small, but the accumulated value for eight years and ten years becomes over 12.5% of a housing price according to the most conservative results (column (6)). This result implies people do not respond instantly to a transient, one-time performance record of a school; however, they have a large amount of willingness to pay for long-term, stably accumulated results. That is, the reputation of a school has a significant effect on people's willingness to pay for better school performance.

A comparison of the cross-section model with the other two models suggests that although the cross-section model controls for the monthly time fixed effects by each 11 school zones – and the adjusted R^2 almost reaches 0.9 – the model still suffers from the endogeneity problem. If we look at the results from the House Fixed-Effect Model and the Time-varying Unobservable Model (column (3)$\sim$(6)), the model in columns (3) and (5) includes the top university entrance rate variables for the last ten years and other explanatory variables only at time t. However, one could expect significant correlations between the past records of top university entrance rate and the past amenity information around a house to cause a biased estimation. To control this problem, the models (2), (4), and (6) in Panel A, and (8), (10), and (12) in Panel B include the equivalent time lengths (from t to $t-9$) of other explanatory variables in the model in addition to the top university entrance rate variables of the past ten years. If we compare the results of column (4) with (3), and (6) with (5), the cumulative value of coefficients for eight years decreases 13% (House Fixed-Effect Model, from 13.72 to 11.97) and 26% (Time-varying Unobservable Model, from 16.89 to 12.54) respectively. This result implies that positive correlations exist between school performance and good amenities around a house in the past records.

Note that results from both models – House Fixed-Effect Model and Time-varying Unobservable Model – are similar. The estimated γ in equation (4), which comes from the Markov process assumption (equation (3)), is 0.976 in the model (6). First, this result means that the unobservables and omitted attributes of a house stay at almost the same level during the two-year period, given the Markov process assumption about unobservables. Second, it also verifies this paper's expectation that it is more likely that unobservables stay at the same level if a housing panel covers the short-term. Note that if the γ is equal to one, the Time-varying Unobservable Model is converted into the House Fixed-Effect Model while justifying the use of a mean- or first-differenced model. Because the estimated γ is close to one, both models presented in Table 1.3 have similar results. However, the estimated γ – statistically different from one[34] – also provides a reason to use the Time-varying Unobservable Model instead of the House Fixed-Effect Model for unbiased estimations and more robust results.

Panel B shows the results when the model includes the recent three-year average top university entrance rate instead of its records for the last ten years respectively. When we use a cross-sectional model, if most of the schools' previous records have been stable – a statistically high correlation among the past records of a school – then the estimation bias may not be severe even though it does not consider the reputation effect. However, a cross-sectional model is not free from omitted variable bias. On the other hand, if we use a differenced model to control time-invariant omitted variables while using the short-term averaged school performance statistics to estimate its implicit

[34] The estimate is statistically different from 1 even under 0.1% significance level.

price, the bias can become severe because of the differencing. The intuitive reason is that the differencing used to delete time-invariant omitted variables can also eliminate the reputation effects that come from all previous school performance results. For instance, if a recent three-year averaged test score is used in the differenced model, the differencing deletes a large portion of the reputation effects coming from the test score more than four years ago. If we compare the results in the House Fixed-Effect Model (column (3) and (9)), the estimated coefficient decreases 84% (from 13.72 to 2.26) and 85% (from 16.89 to 2.56) in the Time-varying Unobservable Model (column (5) and (11)). In case of the *boundary discontinuity design* (BDD) based on a cross-sectional model – which controls omitted variable bias by including boundary fixed-effects and other controls in the region – the bias can be trivial, since the correlation among past performance results of a school is usually high. Even in this case, however, this study suggests considering a long-term performance history of a school to obtain more accurate estimates for the implicit prices associated with better school performance.

As robustness checks, this study uses the regional fixed-effect for smaller area (25-district) instead of 11 school zone level fixed-effect (Table A2, Table A5, and Table A6). Also, I try different distance-weight matrices for the matching of amenity information and school statistics with a house. To be specific, this study adopts the distance-weight matrices with 1 *km* and 2 *km* radius cut-offs for the matching between amenity information and a house instead of using the matrices without radius cut-offs (columns (1) and (2) in Table A3 and Table A4).[35] Also, for the matching between school statistics and a house, this paper tries another matching matrix using the information of the nearest three schools (three schools for each middle and high school) instead of using all schools in the school zone where the house is located (columns (3) and (4) in Table A3 and Table A4). As a distance-weight, the same metric – inverse squared distance – is used. According to the results, the primary results, long-lasting reputation effects and more weight on recent results, are maintained although there are differences in the estimated values of accumulated coefficients.[36]

[35] I do not report the results from the model using the 2 *km* radius cut-off distance-weight matrix because the results are quite similar to the presented results from using a 1 *km* radius cut-off matrix.

[36] Columns (5) and (6) in Table A3 and Table A4 use both a 1 *km* radius cut-off for amenity information matching and the nearest three schools statistics for school information matching with a house.

When the Time-varying Unobservable Model with the three-nearest school matching method (column (4) in Table A4) is adopted, the estimated coefficient decreases to 10.058 compared to the result in column (6) in Table 1.2. Note that the coefficient decreases by 20% (from 12.541 to 10.058), however, the three-nearest school matching method also increases the gaps in top university entrance rate between the Gangnam and the other areas in Seoul by 19%. As a result, the difference in the school statistics matching methods does not have a significant effect on housing price gaps caused by the differences in top university entrance rate results. This paper explains the details about the housing price gap decomposition in the next section.

1.3.2 Education Environment and City Segregation

This section analyzes the implicit prices of better educational environments and shows how much the housing price gaps within a city are explained by the differences in educational environments.

Table 1.3 shows the estimated implicit prices and calculated willingness to pay for better educational environments in addition to the top university entrance rate from the model (4) (House Fixed-Effect Model) and the model (6) (Time-Varying Unobservable Model) in Table 1.2. Note that given the estimated γ is close to one, the results from both models are similar. In discussing the results, we focus on the Time-varying Unobservable Model. The first column shows the estimated coefficients, and the second column represents the calculated willingness to pay for a one-unit increase of each variable per year. For the calculation, the mean housing price ($ 438,656) is used and annualized at the rate of 7%.

According to the results, an average household has the highest willingness to pay for the one-unit (1%p) higher top university entrance rate among the presented variables, and distances from a house to schools also have substantial negative implicit prices per unit (1 km). Next, the results of WTP for the one standard deviation increase show more realistic figures to compare relative importance among variables. The top university entrance rate is still the most critical attribute among all of the educational environment variables in the model. To live around a school which has one-SD (0.71%p) higher entrance rate to a top university, an average household is willing

Table 1.3 Willingness to Pay for Educational Environments

	Dependent Variable = log(price)					
	House Fixed-Effect Model			Time Varying Unobservable		
VARIABLE	Coefficient	WTP ($) (1 unit)	WTP ($) (1 SD)	Coefficient	WTP ($) (1 unit)	WTP ($) (1 SD)
Top Univ Entrance Rate	11.965***	3,903	2,781	12.541***	**4,103**	**2,924**
N-SAT Score	0.017	5	50	0.053**	16	159
Top High School Entrance Rate	0.035	11	10	0.062	19	17
Private High School Entrance Rate	0.050	15	73	0.061	19	89
Neighborhood % BA or Above	0.167	51	590	0.338**	104	**1,200**
Private Academy	0.114	35	757	0.103**	32	**683**
Student per Class	-0.844***	- 258	- 596	-0.625***	- 191	- 442
Student per Teacher	-0.423***	- 130	- 174	-1.233***	- 376	- 504
% Regual Teacher	0.011	3	13	0.073	23	89
Distance to Secondary School				-2.85**	**- 859**	- 377
Distance Elementary School				-5.32**	**- 1,560**	- 247
PM10 Air Pollution	-0.833***	- 255	- 292	-0.918***	- 281	- 321

All coefficients are report as β*100.
Willingness to pay is annualized at rate of 7% for mean house price of $ 438,656
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

to pay \$ 2,924 per a year. The portion of the highly educated in a neighborhood, the number of private academies (as a proxy for a better environment for private educations after school) around a house, and distances from a house to schools also become critical attributes affecting housing prices.

Recently, rapid air quality deterioration in Seoul has caused many people to become interested in PM10 (particles less than 10 micrometers in diameter) concentration. Note that, according to the estimation result, a home buyer would be willing to pay 0.9% of the housing price to avoid $1\mu g/m^3$ increase in PM10 concentration. This estimated value is higher than the estimates from Bayer, Fruehwirth, Kim, and Timmins (2012). They estimate the (negative) implicit price of PM10 as 0.36∼0.58% of a housing price. To estimate the price of air pollution, they analyze the housing market of California's Bay Area. The critical difference is that the average annual PM10 concentration in Seoul ($47.5\mu g/m^3$) during the sample period is much higher than California's Bay Area ($23\mu g/m^3$) in their study. Due to the increasing marginal disutility of non-goods such as air pollution, the higher average level of air pollution could increase households' willingness to pay to avoid the one unit increase in air pollution.

In contrast to the previous literature, in this study, the test score (N-SAT) of a school only has a marginal effect on housing prices. However, when the top university entrance rate variables were excluded from a model, the N-SAT score became the most critical variable. Given the meaning of a coefficient in a hedonic price regression – the effect of a one unit increase of a variable on the market price while holding the other attributes fixed – the above results can be interpreted as a higher N-SAT score not being meaningful if the top university entrance rate is the same. Also note that, as with the performance of a middle school, the estimated implicit price of top high school entrance rate is not significant, and the private high school entrance rate only has a marginal implicit price.

Next, we investigate the city segregation and the effect of educational environments on housing price gaps within a city. As a geographical summary, see Figure 1.1 and 1.2. The figure shows the overall segregation status in school performances, neighborhood composition, and housing price on the map of Seoul. Note that the

Table 1.4 Housing Price Gap Decomposition

Variable	Mean value of Variable: Gangnam	Mean value of Variable: The other	Coefficient	($) Price Gap Decomposition	% of Difference(Mean)
Top Univ Entrance Rate	2.02	0.60	12.541***	**101,269**	**29.40%**
N-SAT Score	109.45	100.07	0.053**	2,833	0.82%
Top High School Entrance Rate	2.71	2.09	0.062	220	0.06%
Private High School Entrance Rate	12.83	8.53	0.061	1,487	0.43%
Neighborhood % BA or Above	70.01	43.44	0.338**	**51,100**	**14.83%**
Private Academy	50.61	32.74	0.103**	**10,441**	**3.03%**
Student per Class	33.76	31.40	-0.625***	**- 8,362**	**-2.43%**
Student per Teacher	15.85	14.47	-1.233***	**- 9,707**	**-2.82%**
% Full-time Teacher	84.44	86.36	0.073	- 802	-0.23%
Distance to Secondary School	1.01	1.13	-2.85**	1,820	0.53%
Distance Elementary School	0.33	0.32	-5.32**	- 443	-0.13%
Mean Housing Price ($900ft^2$)	$757,076	$412,610	**SUM**	**149,855**	**43.5%**

All coefficients are report as β*100.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

better performing schools, well-educated people, and more expensive houses are highly concentrated in the Gangnam school zone area.

Numerically, the first two columns in Table 1.4 show the differences in average statistics between the two areas. The Gangnam area has 1.4%p higher top university entrance rate and 9%p higher N-SAT score than the other area. In neighborhood composition, there is a 25%p gap in the portion of BA or above degree holders in the 40∼49 age group between the two areas. Also note that, on average, there are 18 more private academies around a house in the Gangnam school zone; however, there are also more students per class and teacher in Gangnam school zone compared to the other area. Last, there is little difference in distances from a house to schools.[37]

[37] The summary Table 1.1 and Table A1 compare all variables between the Gangnam and the other area.

Table 1.4 also shows the housing price gap decomposition resulting from differences in educational environments between the Gangnam school zone and other areas in Seoul. For the calculation of the price gap decomposition, this study uses the estimated coefficients from the model (6) in Table 1.2 and uses the mean housing price of mean-size apartments ($900ft^2$) in each region. The right two columns show the number of price gaps explained by each attribute and the ratio of it to the total average housing price gap.[38] Note that the difference in top university entrance rates explains almost 30% of the housing price gap. The neighborhood composition and the number of private academies also account for 14.83% and 3.03% of the housing price gap respectively. In summation, almost 50% of the housing price gap comes from differences in the above three components between the two areas. On the other hand, more students per class and teacher in the Gangnam school zone diminish the housing price gap by 5.25%. Other attributes have relatively marginal effects on the price gap, and overall, the educational and neighborhood environments account for 43.5% of the housing price gap.

If the Time-varying Unobservable Model uses the three-nearest school matching method for the matching of school statistics (column (4) in Table A4), the estimated coefficient of the top university entrance rate decreases by 20% (from 12.541 to 10.058). However, note that the gaps in the top university entrance rate between the Gangnam and the other areas increase by 19% from 1.42 (= 2.02 – 0.60) to 1.69 (= 2.2 – 0.51) when the three-nearest school matching is adopted. Thus, the top university entrance rate still explains 28.27% of the mean housing price gap between the two areas under the different matching method. As a robustness check, this paper also adopts the regional fixed-effect by smaller areas (25 district level) instead of 11 school zone level for the housing price gap decomposition. Table A7 and A8 show the willingness to pay for better educational environments and the results of the housing price gap decomposition when models use a 25-district-level fixed effect. According to the results, the estimated implicit prices of primary variables decrease; however, the differences in educational environments still explain the significant portion (34%) of the housing

[38] Because the dependent variable is log(price), we calculate the portion of the price gap explained by each variable as follows: $\frac{\beta * (X_G - X_{NG})}{log(PRICE_G) - log(PRICE_{NG})}$, here the subscript G denotes the Gangnam school zone and NG means the other areas)

price gaps between the Gangnam and the other areas.

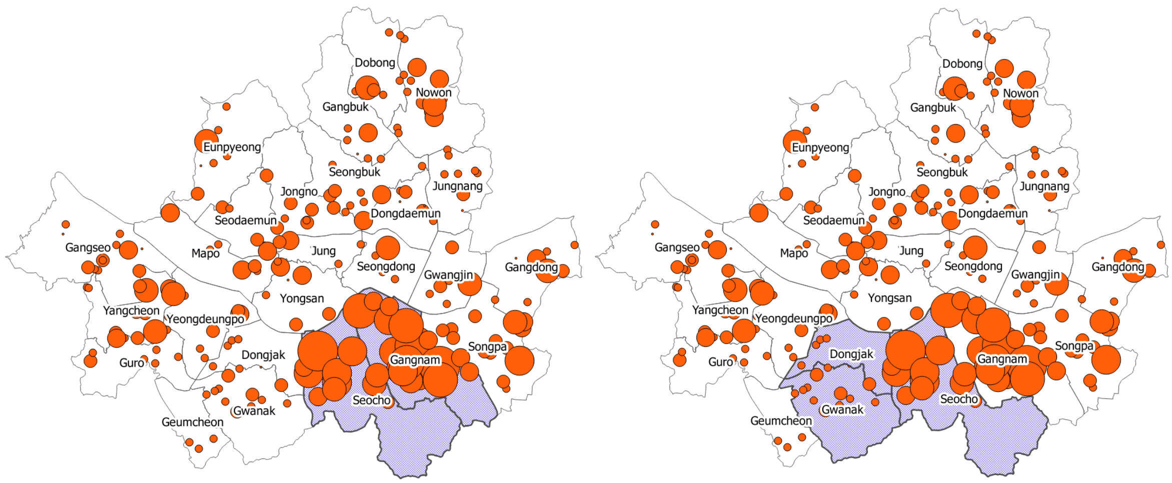
1.4 Policy Simulation

As discussed in the above section, educational environments have a significant effect on housing prices based on households' substantial willingness to pay for them. Accordingly, educational policies such as school desegregation by changing school catchment areas and student allocation systems inevitably affect housing prices in the area where the new policy is applied. Hence, it is impossible and undesirable to separate out an urban housing policy from such educational policies. Indeed, the government of Seoul is currently considering various educational policies such as school zone adjustment and changes in student allocation methods to mitigate severe segregations in both housing prices and school performance gaps within the city.

In this regard, we can refer to many precedents and studies on this topic. In most cases, the purpose of the educational policies was not originally intended to affect housing prices; however, it turned out to significantly affect housing prices in the region. Baum-Snow and Lutz (2011) analyze the public-school desegregation policy in US cities and its effect on residential location choice by race. Machin and Salvanes (2016) investigate the change in the student allocation system of Oslo, Norway in 1997 and show its effect on housing prices. Boustan (2012) examines the changes in housing price gaps in US metropolitan areas with and without public school desegregation policies. Ries and Somerville (2010) use school zone adjustment in Vancouver as a quasi-experiment to measure housing price capitalization of school quality.

In this section, as the policy simulation, this paper changes the school zone boundaries and forecasts what would happen in the new school zone area. In the case of educational policy, such as school zone and student allocation changes, there should be households who support the new policy and those who oppose it. Therefore, it might be meaningful if we can predict the result of the vote for the new policy and calculate households' willingness to pay for the policy change. The expected number of people in favor of the new policy and their estimated willingness to pay can provide a valuable criterion for evaluating the superiority of policies.

Figure 1.3 presents the school zone change simulation. The Gangnam school zone



(a) Current **(b) Simulation**

Figure 1.3 School Zone Adjustment

(shaded areas in the left figure) currently includes the Gangnam and Seocho districts. This study changes the school zone boundaries as in Figure 1.3b, which includes the Seocho, Dongjak, and Gwanak districts in the new school zone. That is, the new school zone policy is designed to tie up adjacent areas with significant gaps in the top university entrance rate.[39] Schools in the Dongjak and Gwanak districts have lower top university entrance rates compared to the schools in the Seocho district. School zone adjustment will affect the matching between a house and schools and the distances from a house to schools in the same school zone immediately after the change. For example, suppose house "A" is located in the Seocho district and the nearest high school is "B," however, the school is located in the Dongjak district, which is located in the other school zone. In this situation, the house "A" is not affected by the performance of high school "B" or the distance to get there under the current school zone system. However, after the school zone is adjusted like the above simulation, the high school "B" has the largest effect on house "A" of all the high schools.

In the long-term, different student composition – students coming from different regions – can affect the performance of schools and consequently affect the neighborhood composition due to a household self-segregation according to school performance. For these reasons, it is expected that the performance of schools and the neighborhood

[39] In this case, the Gangnam district can form another school zone with other districts such as Seongdong or Gwangjin.

composition could vary in the new school zone over time.

Accordingly, the simulation considers two possible scenarios based on time. The first scenario considers the short-term situation immediately after the policy takes effect. In the short-term, the effects come from the differences in matching a house to a school under the new school zone while assuming that school performance statistics and neighborhood composition stay at the same level. The second scenario allows the possibility that school statistics and neighborhood composition can vary after the new school zone system takes effect in the region. This scenario can be considered as a long-term situation. For the analysis of the long-term, we need virtual school performance statistics and neighborhood composition after the new regime. To generate those numbers, this paper uses the following method.

First, I calculate the average standard deviation of school performance (the top university entrance rate) and neighborhood composition (% BA or above degree holders in the 40∼49 age group) within the same school zone from the current statistics. Next, virtual values for each school and census block in the new school zone are generated based on the assumption that standard deviation of school performance and neighborhood composition in the new school zone will converge to the average standard deviation of school zones in Seoul. To be specific, the standard deviation of the top university entrance rate is 0.8% in the new school zone area; whereas, the average standard deviation within the same school zone (calculated from 11 school zones) is 0.4%. In neighborhood demographics, the standard deviation in the portion of BA or above degree holders in the 40∼49 age group is 19% in the new school zone, and that value from the average of 11 school zones is 14%. This study generates the virtual top university entrance rate and the portion of highly educated in a neighborhood by using the initial and target standard deviations and the Z-value in the standard normal distribution.[40]

This study also considers the sample weights by using the total housing population data of the region to increase the accuracy of the policy simulation for the voting experiment and estimating overall households' WTP in the region. As discussed in the

[40] That is, this paper assumes the standard deviations within the new school zone – 0.8% (top university entrance rate) and 19% (neighborhood composition) – will converge to 0.4% and 14% respectively in the long-term.

previous Section 1.2.2, transaction data of apartments is used to construct the short-term panel dataset. However, the samples from the transactions are a part of the total number of apartments in the policy area, and there are also other types of houses in the area. Given that the portion of actual transactions among all of the housing units can be different depending on the region, one might expect that the transaction data could not represent the whole households in the region due to sample bias.

To overcome this problem, I use Housing Census data and compare the number of real transactions to the total number of apartment units by census tract level then generate sample weights to complement the real transaction samples. Another problem with the policy simulation is that other types of houses are in the region, such as single detached houses. Note that all households should be included in the analysis to precisely forecast the result of the vote for the new school zone policy. For this, the simulation model also includes the transaction data of all types of houses along with apartments and generates the sample weights by using the method as for apartments to fix the sample bias. The results, which do not consider the above sample weights, are presented in Table A9 and A10 in the appendix. In conclusion, there is no significant difference in outcomes between the two methods.

Figure 1.4 shows the kernel density estimation of households' willingness to pay for the school zone adjustment in the region.[41] The upper figure shows the WTP as the percent of housing prices and the lower figure shows it as the amount of money. The line and dashed graphs represent the short-term and long-term situations respectively. In the short-term, again note that the effect comes from the different matching between a house and a school as the school zone changes. Households in the lower score zone area (Dongjak and Gwanak districts) benefit from the new matching to schools with better performances, and the opposite happens in the higher score zone area (Seocho district).[42] In the long-term, the gaps in WTP between two regions becomes broader than in the initial short-term situation, since the simulation assumes that school per-

[41] The Epanechnikov kernel function is adopted for the estimation.

[42] Additionally, note that the short-term effect also comes from the different distances from houses to schools and the number of student per class and teacher. The higher score zone area benefits from the lower number of student per class and teacher. However, its effect is much smaller than the effect from the top university entrance rate of a school, since there are small differences in the number of students per class and teacher between the two regions.

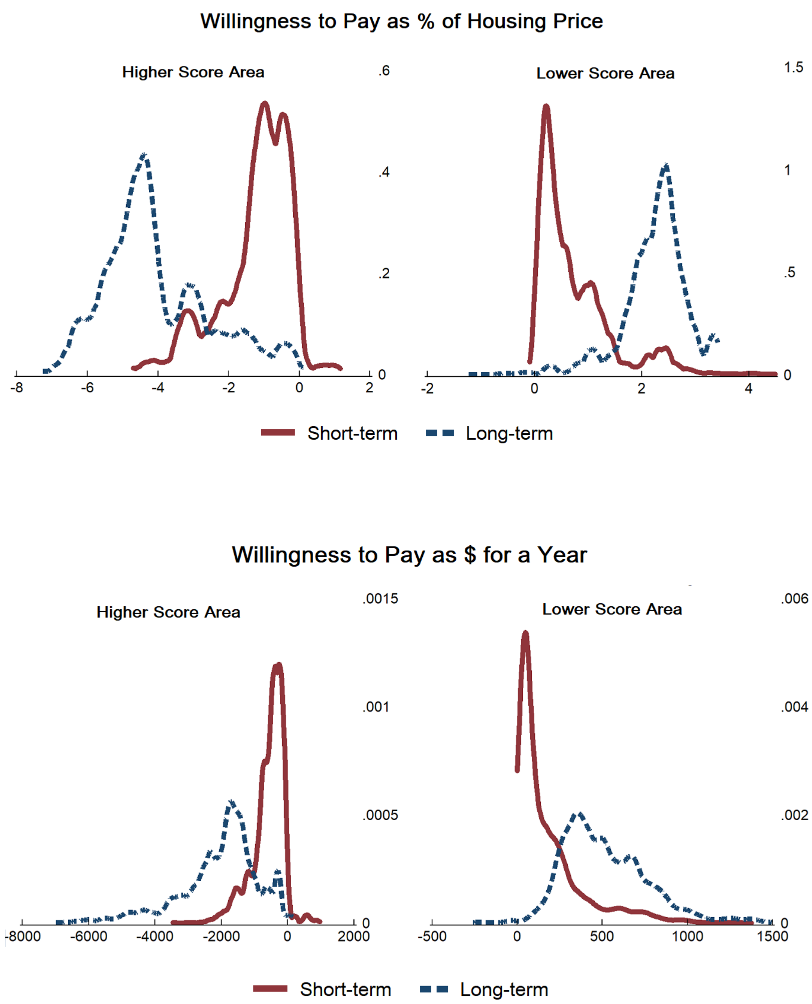


Figure 1.4 School Zone Adjustment Simulation

formance (top university entrance rate) and neighborhood composition (% BA or above in the 40∼49 age group) will be homogenized within the new school zone over time. Even though there are no significant differences in the absolute value of WTP as the percentage of a housing price, there are more substantial gaps in WTP in the amount of money between the two areas because the average housing price in the higher score zone area is more than double that of the lower score zone area.

Tables 1.5 and 1.6 show the summary statistics for the policy simulation results. Overall, regarding the voting experiment, median WTP as a percent and money value show positive numbers, which means there would be more households supporting the new school zone in both the short-term and long-term. However, the sum of all house-

Table 1.5 WTP for the School Zone Adjustment

Area	Time	WTP % Median	WTP % Mean	WTP($) / Year Median	WTP($) / Year Mean	Obs
All	Short-term	0.18	0.24	55	10	183,504
	Long-term	2.77	– 0.11	468	– 677	
Lower Score	Short-term	0.42	0.61	89	155	112,458
	Long-term	3.50	3.39	697	754	
Higher Score	Short-term	– 0.37	– 0.52	– 193	– 220	71,046
	Long-term	– 5.24	– 5.63	– 2,488	– 2,942	

Table 1.6 The Changes in Experimental Area

Area	Time	Housing Price ($1,000) Median	Housing Price ($1,000) Mean	Housing Price ($1,000) SD	Top Univ. Entrance(%) Median	Top Univ. Entrance(%) Mean	Top Univ. Entrance(%) SD	BA or Above(%) Median	BA or Above(%) Mean	BA or Above(%) SD	Obs
All	Initial	401.9	485.0	286.1	0.61	1.20	0.97	52.81	56.08	13.13	183,504
	Short-term	402.7	485.2	284.7	0.75	1.20	0.90				
	Long-term	406.9	475.4	262.9	0.86	1.18	0.69	53.48	55.81	10.25	
Lower Score	Initial	304.0	330.6	135.7	0.44	0.51	0.28	48.39	47.24	6.48	112,458
	Short-term	305.9	332.8	137.0	0.49	0.59	0.34				
	Long-term	314.0	341.4	138.9	0.69	0.77	0.24	50.06	48.94	5.26	
Higher Score	Initial	696.9	729.5	291.2	2.24	2.29	0.59	70.36	70.06	7.71	71,046
	Short-term	693.1	726.3	291.1	2.12	2.15	0.64				
	Long-term	657.1	687.4	273.2	1.66	1.84	0.66	66.84	66.68	5.88	

holds' WTP in the long-term becomes negative (the mean WTP is – 0.11% and $ – 677) since the negative WTP in the higher score area ($ – 2,942) is almost four times larger than the positive WTP in the lower score area ($ 754). Table 1.6 shows the changes in housing prices, top university entrance rates, and neighborhood compositions in the new school zone. The standard deviation of the top university entrance rate and neighborhood composition decreases in the long-run by the assumption of the second scenario, and consequently, housing price gaps within the area are diminished. Note that the standard deviation of housing price decreases by 8.1% in the long-term compared to the initial status. The mean housing price and the standard deviation of housing prices increase in the long-term by 3.27% and 2.35% respectively in the lower score area and decrease by 5.77% (mean housing price) and 6.18% (standard deviation) in the higher score area. As a result, the housing price gap in median housing prices decreases by 12.67% (from $ 392,900 to $ 343,100). In conclusion, the results of the policy simulation show that the school desegregation policy, through school zone adjustment, significantly reduces the housing price gaps within the city.

1.5 Dynamic Model

In this section, I investigate the dynamics of the housing market. The analysis in the above sections using a hedonic price regression shows the implicit prices of housing attributes in a static equilibrium. Note that the estimated implicit price is interpreted as the effect of an attribute on market price while holding the other attributes fixed. However, there could be bilateral relationships among housing attributes over time. For instance, we might expect that better school performance can attract more educated people – which is another critical factor among households' implicit prices for housing attributes – around the school (school quality ⇒ neighborhood composition). Given that parents' education level could also affect the children's academic performance, the higher education level of the neighborhood – attracted by improved school performance – would additionally enhance school performance over time (neighborhood composition ⇒ school quality).

As a consequence, the better school performance indirectly increases the housing price by attracting a more educated neighborhood (school quality⇒ neighborhood composition ⇒ housing price) along with its own direct effect on the housing price (school ⇒ housing price). Also, we can think of the second and the third round bilateral indirect effects, such as the echo effect (school quality ⇒ neighborhood composition ⇒ school quality ⇒ neighborhood composition). As such, in a dynamic framework, it is likely that a particular attribute can play a more important role if it attracts and increases other good attributes of a house. This study uses the panel VAR (vector autoregression) model as follows to consider the dynamics, which show direct and indirect effects from the bilateral relationships between housing attributes.

$$Y_{it} = \sum_{p=1}^{T} \Gamma_p Y_{it-p} + \Theta X_{it} + D_{jt} + \xi_i + \varepsilon_{it} \tag{5}$$

Here, Y_{it} represents the vector of endogenous variables in the VAR system, and Γ_p and Θ are coefficient matrices. The Y_{it} includes 19 variables (log of housing price; top university entrance rate; percent of BA or above degree holders in the 40∼49 age group; log of population; log of employment (full-time and part-time employment), log of self-employed; the number of student per class and per teacher; and the number of private academies, libraries, supermarkets, culture related stores, shopping stores, clinics, restaurants, sport facilities, banks, and grocery stores around a house). X_{it} denotes time-variant exogenous variables. In the case of some variables, it is hard to think that there exists bilateral relationships among variables. For example, the distance to the nearest school or subway station are more likely to act as exogenous variables. Those variables can affect the other attributes of a house; however, it is hard to think of a case where other attributes affect those distances. The vector X_{it} includes distances from a house to the nearest subway station, department store, supermarket, and schools (elementary, middle, and high school). Also, as in the above hedonic price regression analysis, the model takes into account D_{jt}, the fixed effects over time (t) in each region (j).[43] ξ_i represents a vector of the house-specific unobserved heterogeneity. Note that the model uses the first-differenced variables to control the endogeneity caused by correlations between ξ_i and Y_{it-1}, while eliminating the unit root in the variables.[44] Last ε_{it} denotes a vector of shock.

From the model selection criteria, the equation (5) is estimated for T = 3 using the system GMM estimator developed by Arellano and Bover (1995) and Blundell and Bond (1998), which is applied to the panel VAR by Binder, Hsiao, and Pesaran (2005). The system GMM also handles the remaining bias problem caused by the first-differenced error term (Nickell's bias, Nickell (1981)) by using the lagged variables as instruments. This estimator is consistent when the panel dataset has a large number of observations i in a given time dimension. For the estimation, the dataset is constructed using the same method as the hedonic price regressions in the previous sections; however, it covers a more extended period (2006∼2015) than the hedonic model for the panel VAR analysis.[45] This study also tries the structural panel VAR model with the smaller set of endogenous variables while considering the contemporaneous relations among endogenous variables in the appendix A.3.

The panel VAR system includes 19 endogenous variables, so the system generates 361 impulse-response functions to show the relations among endogenous variables

[43] This paper considers the monthly time fixed effect by 11 school zones, which is the same with the hedonic regression analysis in the previous sections.

[44] After the variable transformation, this study checks the panel VAR system stability condition, and the system strongly satisfies the condition.

[45] The house real transaction data does not exist before 2006.

over time. Among them, this paper focuses on the impulse-response functions related to school performance, neighborhood composition, and housing price. Note that the results of the panel VAR cannot be directly compared with the estimation results of the hedonic price regression because, first, each panel VAR equation includes the lags of dependent variables while the hedonic regression does not. Second, owing to the data limitation, the panel VAR model does not include all variables used in hedonic price regressions. Accordingly, in this section, the analysis focuses on showing the dynamic properties of the housing market rather than comparing the estimated numbers with hedonic regressions.

(a) BA : House Price (%)

(b) Top Univ : House Price (%)

(c) BA : Top Univ (%p)

(d) Top Univ : BA (%p)

Figure 1.5 (Cumulative) IRF
– (Impulse) : (Response) –

Note: "BA" denotes the percent of BA or above degree holders in the 40∼49 age group. "Top Univ" means the top university (Seoul National University) entrance rate of a high school.

Figure 1.5 shows the cumulative impulse response functions among three variables of interest: school performance, neighborhood composition, and housing price.

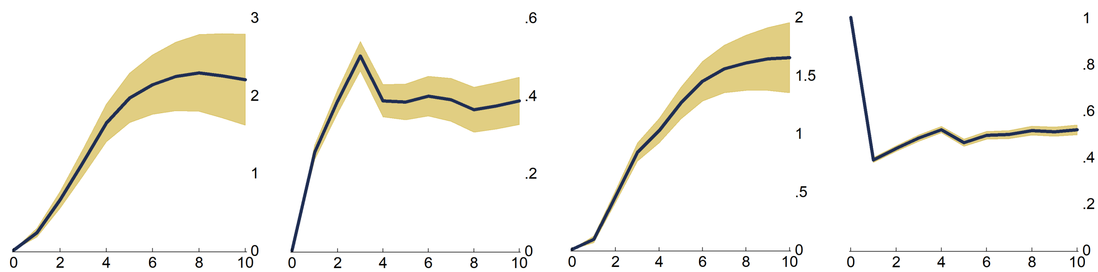
The left side of the colon is an impulse variable and the right side denotes a response variable. The shock is a one-unit increase, and the line represents the estimated impulse response function. The shaded area is the 95% confidence interval generated by the 500 Monte Carlo draws. It is important to note that there exists a long-term effect from two variables (top university entrance rate and the portion of BA or above degree holders in the the 40∼49 age group) on housing price over time as Figures 1.5a and 1.5b demonstrate. The one-time shock at the initial time significantly affects housing price even up to ten years later. Also, Figures 1.5c and 1.5d show that there are significant bilateral long-term effects between school performance and neighborhood composition.

Figures 1.5a shows that 1%p increase in the percent of BA or above degree holders in the 40∼49 age group increases the housing price by 11% in 10 years (in standard deviation, a 1 SD increasing shock from the "BA" variable leads to the 1.8 SD increase in the housing price in ten years). Its effect on housing price at time t+1, only including the direct effect on the housing price, is estimated as 2.4%; however, its indirect effect through the bilateral relationships among variables consistently increases housing price for ten years. Meanwhile, according to Figure 1.5b, 1%p higher top university entrance rate increases a housing price by 1.7% and 4.6% at time t+1 and t+10 respectively (in standard deviation, a 1 SD impulse shock from "Top Univ" causes a 0.1 SD increase of the housing price). Figures 1.5c and 1.5d represent the bilateral relationships between top university entrance rate and the neighborhood composition. Importantly, they increase and attract each other. The higher portion of educated parents increases the primary school performance (top university entrance rate), and the better school performance consistently attracts more educated neighborhoods around a school.

Specifically, a 1 SD (2%p) increase in the portion of BA or above degree holders in the 40∼49 age group in neighborhood enhances the top university entrance rate of schools in the region by 0.96 SD (0.28%p) for ten years after eliminating the regional time trend.[46] In the opposite direction, a 1 SD (0.29%p) higher top university entrance rate attracts more educated people to a neighborhood by 0.06 SD (0.11%p) in ten years. In the case of the impulse response from school performance to the neighborhood composition as shown in Figure 1.5d, the overall long-term effect (0.38%p, at time

[46] The panel VAR model includes the regional time fixed effects (11 school zone – monthly level)

t+10) – including direct and indirect effects among endogenous variables – becomes significantly larger than its short-term direct effect (0.09%p, at time t+1). On the other hand, Figure 1.5c demonstrates that the 95% confidence interval of the overall long-term effect [0.089, 0.175] at time t+10 includes the size of its direct effect (0.105) at time t+1. In other words, the long-term effect of school performance on the portion of the well-educated in a neighborhood is more significant than the opposite direction (from neighborhood composition to school performance). Whereas regarding the size of the effect, the neighborhood composition has a more substantial effect on school performance (1 SD impulse to the response of 0.96 SD) than the opposite direction from school performance to neighborhood composition (1 SD impulse to the response of 0.06 SD).

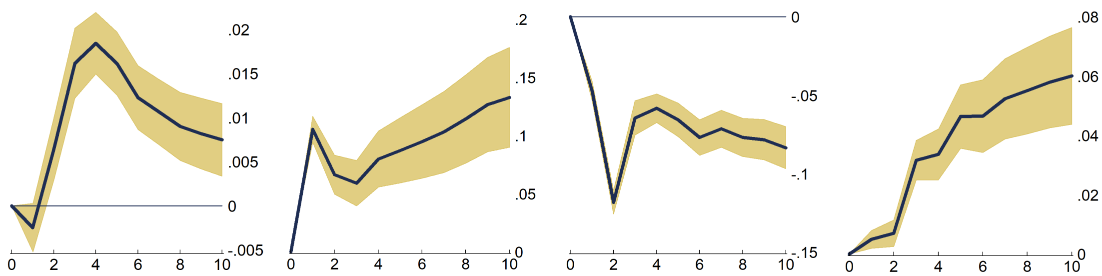


(a) Culture **(b) Library** **(c) Population (%)** **(d) Top Univ (%p)**

Figure 1.6 (Cumulative) IRF
– Impulse = Top Univ (1%p) –

Figures 1.6, 1.7, 1.8, and 1.9 show the more specific relationships among endogenous variables regarding school performance and neighborhood composition. Figure 1.6 shows the cumulative response functions from the one-unit (1%p) increase in shock of the top university entrance rate. Enhanced school performance increases the number of some amenities such as cultural stores (book, music and sporting goods stores) and libraries. Also, it increases the population growth while attracting more educated people to a neighborhood in the region. Figure 1.6d shows the cumulative response function of the top university entrance rate from the same variable's shock. Note that the one-time increase in the top university entrance rate of a school tends to converge to the original level, which shows the mean-reversion property; however, it does not decrease to zero and half of the initial shock is maintained.

Figure 1.7 shows what statistically induces better school performance. The increasing number of private academies (as a proxy for a private educational environment) and cultural stores such as book-stores enhance school performance in the region. Along with a higher portion of educated people (% BA or above in the 40∼49 age group) in a neighborhood, the improvement in student-class ratio also increases the top university entrance rate. All those effects are significant and have time-lagged long-term effects.

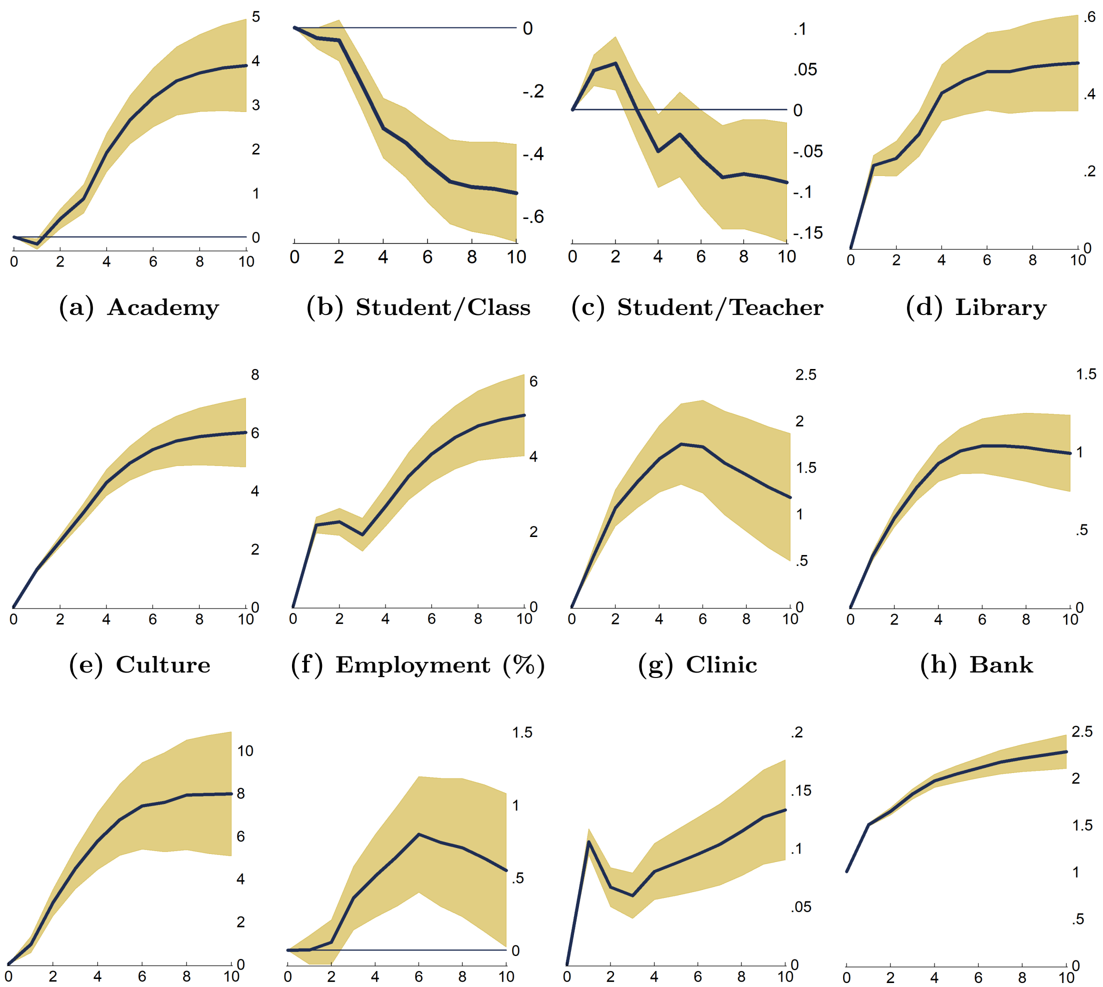


(a) Academy **(b) BA (1%p)** **(c) Student/Class** **(d) Culture**

Figure 1.7 (Cumulative) IRF
– Response = Top Univ (%p) –

Figure 1.8 presents the response functions corresponding to the impulse shock (1%p) of neighborhood composition (% BA or above degree holders in the 40∼49 age group). The results imply that the higher portion of educated people in a neighborhood attracts many amenities such as private academies, libraries, culture stores, clinics, banks, and restaurants. If we focus on the educational environment more specifically, a 1 SD (2.05%p) increase in the portion of BA or above degree holders in the parents' age group results in a 1.78 SD (7.9) increase in the number of private academies. Interestingly, the increasing portion of the highly educated in a neighborhood leads to improvements in the ratio of students per class and teacher; however, in the case of students per teacher, its effect is somewhat insignificant due to the relatively wide confidence interval. We can interpret these results as more educated parents searching for and moving to an area where those ratios are expected to be enhanced shortly because of a new school opening in the region.[47] The increased neighborhood education level also tends to create more jobs in the region, and the above results are meaningful

[47] The plan for establishing a new school usually becomes available to the public a couple of years before it opens.



(a) Academy (b) Student/Class (c) Student/Teacher (d) Library

(e) Culture (f) Employment (%) (g) Clinic (h) Bank

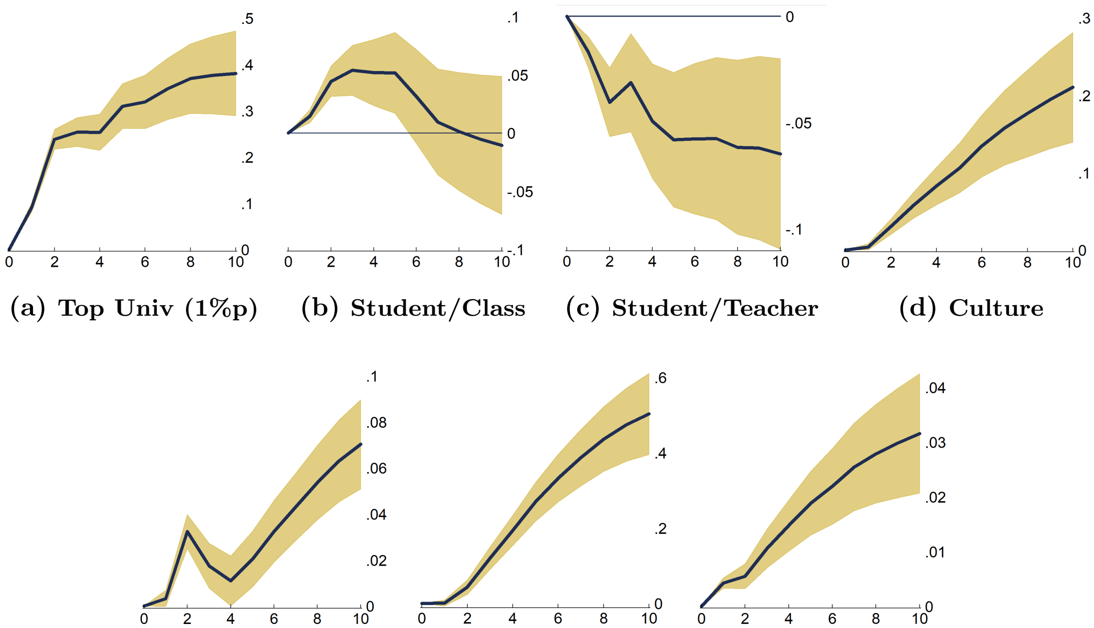
(i) Restaurant (j) Supermarket (k) Top Univ (%p) (l) BA (%p)

Figure 1.8 (Cumulative) IRF
– Impulse = % BA or Above (1%p) –

in that the model also controls for population growth. Given that there is a positive correlation between education level and income, the increasing number of highly educated people who are likely to have more disposable income could have more effect on amenities and businesses around the region compared to a population growth which does not change the neighborhood education level.

Lastly, Figure 1.8l shows the response function of neighborhood composition from the same variable's shock, and we can find the critical difference between the corresponding impulse response function of the top university entrance in Figure 1.6d. In the case of top university entrance rate, it shows the mean reversion trend. That is, if the top university entrance rate increases in a school, then it tends to lower back to its original level in the future. On the other hand, neighborhood composition shows strong persistency, implying that if the portion of highly educated in a neighborhood increases in a region, it tends to increase more in the future. Note that a 1%p increase in the portion of BA or above degree holders (at time t) induces a 2.28%p increase after ten years. Because of this different property between school performance and neighborhood composition – mean reversion and persistency – neighborhood composition plays a more critical role than school performance in a dynamic framework based on more fluent and significant relationships with other endogenous variables.

Next, the results in Figure 1.9 show what attracts the more educated parents to the region. Importantly, as I discuss in the above main panel VAR results, better school performance attracts more educated parents to the region. If we look into the impulses from two ratios (students per class (Figure 1.9b) and teacher (Figure 1.9c)), in the case of the number of students per class, the result is not significant, which is different from the result of the opposite impulse-response relationship shown in Figure 1.8b. On the other hand, the enhanced ratio of students per teacher attracts the highly educated to a neighborhood even though its confidence interval is relatively wide, and the result is similar to the impulse response function of the opposite direction as shown in Figure 1.8c. If we examine the amenities attracting more educated parents, the increasing number of culture stores (book, stationery, music, and sporting goods stores), sports facilities, commercial banks, and restaurants tends to attract more educated parents to the region.



(a) **Top Univ (1%p)** (b) **Student/Class** (c) **Student/Teacher** (d) **Culture**

(e) **Sport** (f) **Bank** (g) **Restaurant**

Figure 1.9 (Cumulative) IRF
– Response = % BA or Above (%p) –

Figures A3 and A4 in the appendix show the other impulse response functions related to the housing price. Figure A3 presents the responses of housing price from shocks of other variables. The increase in the number of banks, culture stores, sports facilities, and restaurants has a significant long-term effect on the housing price. Note that those variables coincide with the factors that attract more educated people to a neighborhood and those that are increased by the higher portion of the educated in a neighborhood. Thereby, we could expect that those amenities have a long-term positive effect on housing prices due to the bilateral relationship with neighborhood composition over time. Figure A4 shows the response functions corresponding to the shock from a housing price, which demonstrates the definite difference from the hedonic model. Note that the increasing housing price usually crowds out many amenities including the portion of highly educated people in a neighborhood and the top university entrance rate. However, given that the size of the effects is quite small for all variables, we could conclude that the housing price has a strong property as an endogenous variable.

To summarize, this study verifies the significant bilateral relationships between the critical attributes of a house. Importantly note that neighborhood composition (the percent of BA or above degree holders in the 40$\sim$49 age group) leads the key dynamics of housing attributes while attracting good amenities and enhancing school performance. The school performance also has an important role in the dynamics; however, the strong mean-reversion trend weakens the effect whereas the persistence in neighborhood compositions strengthens the dynamic effect.

1.6 Conclusion

This study shows that past records of school performance have a long-lasting reputation effect from the hedonic price regression model using a short-term panel dataset that overcomes the problems of the housing market panel dataset. The results imply that people consider the past records of school performance as well as the most recent results and that households have different weights on previous results. That is, people place more weight on recent results of a school, and its weight fades out as time passes by.

This study examines households' willingness to pay for better educational environments and a more educated neighborhood and shows that they account for more than 43% of housing price gaps in Seoul. Based on the estimates of hedonic analysis, this paper tries a policy simulation that changes school zone boundaries. According to the simulation results, the school zone adjustment has significant effects on housing prices and neighborhood compositions. The simulation model also predicts that the policy – changing school zone boundaries – would help to mitigate huge gaps in housing prices and school performance between the Gangnam and the other areas in Seoul.

This study also investigates the dynamics of the housing market by using the panel VAR model. Unlike the hedonic analysis showing the effect of a variable on the market price of a house while holding the other attributes fixed, the panel VAR model allows bilateral relationships among housing attributes and examines the direct and indirect long-term effects among them. According to the results, the overall effects of school performance in the long-term becomes more significant than its effects in the short-term. Also, neighborhood composition shows a more remarkable effect in the dynamic framework compared to the static hedonic model. Because of its persistence, which is different from the mean reversion property of school performance, neighborhood composition plays a critical role in the long-term based on fluent and significant relationships with other housing attributes. To summarize, this paper verifies households' significant willingness to pay for better educational and neighborhood environments and also finds evidence that they play a more critical role if we consider their dynamic effects on housing attributes over time.

APPENDIX

APPENDIX A

A.1 Data Appendix

A.1.1 Housing Environment in Seoul

Seoul Metropolitan Area

The population of Seoul is over 10 million, and its population density (16,204/km^2) is higher than New York City (10,725/km^2). Due to the high population density, most people live in apartments, and less than 15% of people live in a single detached house. 1.2 million people commute from outside the city, and on average, 7 million people use the subway each day to avoid traffic jams. Therefore, the distance from a house to the nearest subway station becomes a critical attribute affecting housing prices.[48] In 2015, there were over 22 department stores and 59 large supermarket chains in Seoul, and people frequently visited them for groceries and shopping. Distance to the nearest department store and supermarket also turns out to be an important factor in housing price. The Han-river runs through the middle of Seoul. In Seoul, it has a roughly 1km (0.62miles) river width on average, and the houses alongside the river offer an expansive river view. Two expressways and many parks located along both sides of the river offer good transportation environments and recreation areas for nearby residences. For these reasons, there could be a premium in the price of a house near the Han-river. This paper generates three dummy variables according to the distance to the Han-river (0∼0.5km, 0.5∼1km, and 1∼1.5km from the Han-river).

Brand of Apartments and High-rise Residential Complexes

By using the name or brand of the apartment, I generate a brand dummy variable if the apartment has a major brand name.[49] In Korea, almost every apartment has a brand name, similar to cars, and the brand name represents the construction firm of

[48] There were 285 subway stations in Seoul in 2015

[49] I selecte major apartment brands using BSTI (Brand Stock Top Index, www.brandstock.co.kr) information. It publishes the brand ranking of various products and firms. I make the brand dummy variable 1 if the apartment brand is Ramian(Samsung), Xi(GS), I'PARK(Hyundai), Hill-state(Hyundai), We've(Doosan), Prugio(Daewoo), The Sharp(Posco), Lotte Castle(Lotte), E-Pyunhan Sesang(Daerim), or SK-VIEW(SK), and 0 if otherwise. (in parenthesis is the construction firm of the brand).

the apartment building. According to a recent survey,[50] the brand of apartments was the first consideration for people given apartments are located in the same place. That is because the brand shows the construction company, and people prefer to have their apartments constructed by famous conglomerate firms such as Samsung, Hyundai, GS, or Doosan. Additionally, the top-ranked brand apartments usually have better conditions (e.g., better landscape and curb appeal, better location, spacious parking lots, security, and more amenities), which are hard to measure by numerical values and are usually unobservable attributes to an econometrician. Consequently, controlling the brand effect can help capture one of the significant unobservable attributes of apartments.

In addition to the brand of apartments, there were 84 high-rise residential complexes in Seoul in 2015. In the complex building, lots of amenities are located on the lower level and apartments units on high floors offer an expansive view. To consider the effect from different types of apartments, this paper also generates a dummy variable for high-rise residential complexes.

A.1.2 Education and Neighborhood Statistics

Top University Entrance Rate

This paper uses the top university entrance rate of a high school as the primary variable of interest representing the performance of a school, which is different from previous studies that use various test score results as a proxy for school performance.[51] According to the results of this paper, both the top university entrance rate and the test score (N-SAT) are critical factors regarding household willingness to pay. However, the effect of top university entrance statistics turns out to overwhelm that of test score results in households' implicit price as discussed in Section 1.3.2.

That being said, there are some problems in using the university entrance statistics in the analysis. First of all, the university entrance statistics by a high school are usually not publicly available. Second, the universities considered prestigious can be different from region to region. For example, in a vast country like the US, a prestigious

[50] http://www.rcast.co.kr/

[51] To the best of my knowledge, I could not find any previous literature using prestigious university entrance rate as a variable of interest representing the performance of a school.

school which the top students enter could be different from the west side to the east side. However, such variations do not exist in small countries like Korea, and moreover, this study focuses only on the housing market of Seoul, Korea. As the primary variable, this paper uses the portion of students who entered Seoul National University (SNU) by high school. The prestige university ranking is stable in Korean society, and Seoul National University (SNU) is the top-ranked university. It would be better if the data about the entrance statistics of other prestige universities were also available to be used together to generate performance statistics of high schools. Unfortunately, this data is not available to the public, and the entrance statistics of SNU by high school becomes available only due to a member of the National Assembly who requested the data for parliamentary inspection of Seoul National University. Given data limitations, this paper assumes there would be a strong positive correlation between the number of students who enter SNU and other prestigious universities when considering the strong university hierarchy in Korea. Under this assumption, the top university entrance rate can be a good proxy to evaluate the university entrance results of a high school, which is one of the most critical performance results.

Nationwide Scholastic Achievement Test

The results of the N-SAT[52] are available from the school information system of Korea[53] by each school. This site offers information about the number of students taking the test and test results for each exam subject (Korean, Math, and English). The results are provided after dividing students into three categories (above normal, normal, and below normal) and showing each percentage by subject in a high school. Based on the data, I calculated the average "Above Normal" ratio of three subjects and divided it with the total average "Above Normal" ratio of all high schools in Seoul. For example, N-SAT score "120" means that there are 20% more "Above Normal" category students in the high school compared to the average high school in Seoul. The average test score is an imperfect quality measure. Nonetheless, it has the advantage of being easily mea-

[52] This test is different from the SAT which is used as the university entrance exam. The purpose of this test is to measure the average academic levels of students and schools. However, this test score is free from the selection bias problem which can occur in scores of an exam that students can choose whether to take the exam because almost all students take the exam and there is no choice for students.

[53] http://www.schoolinfo.go.kr

sured by numbers, observed by both parents and researchers, and compared with other schools; as a result, it has been used in most analyses that attempt to measure the implicit price for school quality.

Types of High Schools and the Matching Process of School Performance Statistics

There are three types of high schools in Seoul: public, private, and special-purpose high schools. Since private and special-purpose high schools adopt an open enrollment system, students can apply for admission to those schools regardless of residence location while public high schools use the allocation system following school catchment area. Private high schools usually have better performance results in top university entrance rate and N-SAT scores compared to nearby public schools. Note that also, in general, only the top 2∼3% ranked students in each middle school enter special-purpose high schools, and consequently, the test score results and top university entrance rates of special-purpose high schools are outstanding. In this study, I use "top high school" instead of its official name "special-purpose high school." The bottom line is that the three types of high schools and different enrollment systems create obstacles for analyzing implicit prices of better educational environments.

To overcome the above problem, this study tried various methods as follows. The first method used the high school entrance rate for three types of high schools by each middle school. For the first step, I matched the high school entrance rate of middle schools to a house by distance-weight matrices considering the middle school enrollment system,[54] and used the matched statistics (different entrance rate for each type of high school by a house) as a weight to combine performance statistics of different types of high schools.

Next, I matched high school performance statistics with a house by each high school type. In the process, this paper used a different method to reflect the different enrollment systems (open enrollment and catchment area). For the matching of private and top high schools, which adopt open enrollment systems, with a house, this paper used distance-weight matrices for each type of high school, and each matrix is generated

[54] Different from the high school enrollment system, middle schools only use the school zone system following catchment area.

by calculating the distances from a house to all private and top high schools in Seoul. Even though the location of a house is irrelevant in applying to those schools, in practice, the benefit from and probability of attending nearby schools becomes higher as the distance to the school gets closer because of daily commute costs. Moreover, households generally select a house near a school which they want their child to attend when they choose the location of the house.

For the performance statistics of public high schools, this paper also used distance-weight matrices. However, for this step, it also considered the school zone and student allocation system of public schools along with the distance from houses to public schools.[55] As the last step, I combined the matched statistics by high school type by using the weight among private, public, and top high schools calculated during the first step.

For example, suppose that the matched high school entrance rate statistics of a house "A" are as follows: 80% (public high school), 17% (private high school), and 3% (top high school), which means 80%, 17%, and 3% of students respectively enter public, private, and top high schools from middle schools around the house "A." Next, we have high school performance statistics for each type of high school matched with the house "A" generated during the second step. Last, we can combine the performance statistics from three different high school types by the weight – 0.8, 0.17, and 0.03.

However, the critical problem of the above method is that people could hold different levels of importance among different types of high school. For example, suppose that on average 17% and 80% of students enter private and public high schools in the region. Then, in this case, the performance of public high schools should have a more significant effect than private high schools when people evaluate school performance in the region. However, if households over-estimate the probability of attending the private high school than the actual rate, then the actual weight of the household, which is based on their subjective probability and affecting the willingness to pay, could be entirely different from the actual entrance rate of each type of high school.

This paper found strong evidence that households' implicit weight among different types of schools could be substantially different from the actual probability of

[55] I further explain school zone and the students' allocation system of public high schools in the next section.

entering each type of high school. I used the same model (6) in Table 1.2, and included the last ten years' records of top university entrance rate of public, private, and top high school respectively. According to the results, the implicit price of top high school performance was not significant. Given that the entrance rate for the top high school is around 2∼3%, the result seems reasonable. However, households' willingness to pay for the better performance of private high schools was more than twice than that of public high schools. Note that 85% of students enter public high school whereas only 12% of students are admitted to private high schools in Seoul. Hence, the severe level of discrepancy between high school entrance rate statistics and estimated implicit price of school performance by high school type implies that households' subjective weight among different types of schools affecting their WTP for better school performances does not reflect the actual high school entrance rate.

There could be a few reasons for the discrepancy. First, households can overestimate the probability of attending the private high school. The willingness to pay for better school performance is revealed when people buy a house. In most practices, household movements are usually made long before children go to high school.[56] Accordingly, what they anticipated at the time of moving and actual high school enrollment a few years later can be somewhat different. Second, psychological factors such as the bandwagon effect can be another reason. Private high schools are evenly distributed in Seoul, and each high school represents the school performance of the region because they are, in general, the best performing school in each region. Thus, if people consider the best performing high school representing the area when evaluating the school performances of a certain region, the performance of a private high school representing each area can have a larger effect than a public school. To summarize, the first method – which uses the high school entrance rate for three types of high schools as the weight for combining the high school performances from three different types – can be inappropriate for matching the high school performances to a house.

Based on the above problems and findings, this paper used the following method. First, I excluded the performance statistics of top high schools for the matching process based on the fact that only 2∼3% of top-ranked students in a middle school enter

[56] According to a survey, most households with students move before their child enters a secondary school and they tend to stay while the child goes to school.

top high schools. Also, the estimation result showing no significant implicit price of performance of top high schools supports the exclusion of top high schools. Second, this paper did not distinguish between private and public high schools for the matching process; however, it considered the school zone system. For example, suppose that 20 public schools are located in school zone "A" , and there are 22 private schools in Seoul. Again note that public high schools adopt a school zone system while private high schools use open enrollment. For the matching between the school performances and a house located in school zone "A", the distance weight matrix calculates distances from the house to both 20 public high schools in the same school zone and 22 private schools in Seoul. That is, the 22 private high schools are involved in the matching process of all houses in Seoul while considering distances, whereas public schools are included in the distance matrix according to the school zone and the location of a house.

School Zone and Student Allocation System

To analyze the housing market with the hedonic price regression, we need to match the education-related statistics with a house. If a one-to-one mapping is possible between a school and a house according to a clearly defined school zone, the matching is simple. However, the school zone system is rather complex in Seoul. The first problem is that there are multiple public high schools in the same school zone. Also, the student allocation system considers the ease of commute when assigning students to a high school, but there is no guarantee that students are allocated to the nearest school from their houses. Moreover, the education authority of Seoul uses a three-step application system when allocating middle school students to high schools.[57]

However, even though the multi-step student allocation process includes an open enrollment system during the first step, according to the recent publication data,[58] more than 96% of students have attended a school in the school zone where they reside.

[57] At the first stage, 20% of students among the total capacity of each high school are enrolled by student applications, and the application is not restricted by the school zone and the location of house. That is, in the first stage, 20% of student enrollment is the same as the open enrollment system. Next, 40% of the total capacities are made up of the applications of students who live in the school zone. Lastly, the remaining 40% of students are assigned from the wide school zone (two or three adjacent school zones consist of a wide school zone). During the process, if there are more applicants than the number of capacity of a school at each step, the lottery system is used to select students. However, since even narrow school zones include more than 20 high schools on average, we cannot match a house with a high school by one-to-one.

[58] The Seoul Metropolitan Office of Education offers this data (http://statistics.sen.go.kr/)

There can be several reasons for this. First of all, daily commute costs to attend high schools in another school zone could be substantial. Secondly, even though students have the choice to apply for a school located in the other school zone, the possibility of actually attending the school can be pretty low. In general, the schools that a parent and a student want to attend usually coincide with other people's choices. The similar preferences among people can lower the allocation possibility of students from other school zones. For instance, students can apply to a popular high school in their school zone at every application step, whereas a student from other school zones can only be an applicant during the first step, which accounts for only 20% of the total allocations. Therefore, this study assumes that all students are allocated to a public high school within their school zone, ignoring the less than 4% of students who attend a public high school in other school zones.

Neighborhood Statistics

This paper primarily uses the portion of BA or above degree holders in the 40∼49 age group in a neighborhood as a proxy for neighborhood education level in a region. The data is drawn from the Population Census which contains details of the education level of people and is published every five years. However, for the panel VAR analysis, we need every year of information about neighborhood education levels by age group. This paper estimates the neighborhood characteristic between the five years by using interval regressions. For this process, this paper adopts the method used by Bayer, Ferreira, and McMillan (2007). They use the Census data that represents housing prices as 26 categorical bands.[59] From the categorical value, they estimate the housing price – the dependent variable of their analysis – by using interval regressions, which restrict the estimated housing price so it lies in the interval. In this paper, I use interval regressions based on the generalized Tobit and estimate the neighborhood education level of a region in a year when the Population Census is not available. This paper uses the value of two adjacent Population Census dataset as lower and upper boundaries. For example, if we estimate the portion of BA or above degree holders in the 40∼49 age group in a census tract in 2013, this paper uses the portions of that in 2010 and 2015 as

[59] For example, the Housing Census shows the interval in which the housing price is included such as \$ 100,000 ∼ 200,000.

the lower boundary and upper boundary respectively. The neighborhood information is estimated separately for each of the 25 districts in Seoul. For the estimation, this paper uses the demographic information that is published every year by census tracts such as the changes of a population by age group (the changes of population in the 35∼39, 40∼44, 45∼49, and 50∼54 age group in a region). This paper also uses neighborhood education levels of different age groups (the portion of BA or above degree holders in the 35∼39 and 50∼54 age groups) and the changes in the number of employees (the number of full-time and part-time workers and the number of self-employed) by census tracts.

A.1.3 Panel Dataset Generating Process

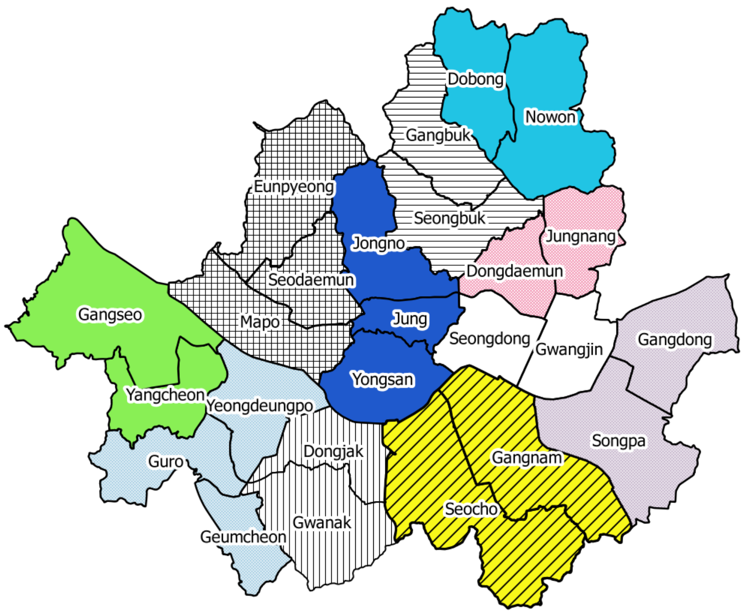
This section explains how the two consecutive years panel dataset was generated for the analysis in this paper. Because of the same group definition as described in section 2.2, it is likely that multiple transactions of the same housing unit happen within a year, and in some cases, even in a month. From the multiple transactions within a year, this study generated the price of the year of a group of houses as follows. First, this paper calculated the average monthly price of a group of apartments from the multiple transactions. In the case of a month, it is not common to have multiple transactions within the same group. Second, from the monthly average price data, this paper constructed monthly panel dataset (unbalanced) within a year. Third, by using the monthly panel dataset, the monthly fixed effects by 25 district regions are estimated to eliminate irrelevant macro fluctuations from the analysis. Fourth, I converted the average monthly price into the price of the base month using the estimated monthly price index.

As a robustness check for the same floor group definition, this paper also tried other group definitions. To generate more observations in the panel dataset, this paper put a group of apartments in the same group if their address, size, and construction year are all the same after eliminating the different floor effect. That is, this paper converted the housing prices located on the various floors in the same building into the virtual prices assuming all houses are located on the same floor. To estimate and eliminate the floor effect, I constructed the monthly panel dataset with the same method that is used to estimate the monthly price index by region. After that, the floor group fixed effect is estimated from the panel dataset and used to change the housing price into the counterfactual price of the same floor. Next, this paper also converted the housing price – with the floor effect eliminated – into the same monthly price, again using the monthly price index by region.

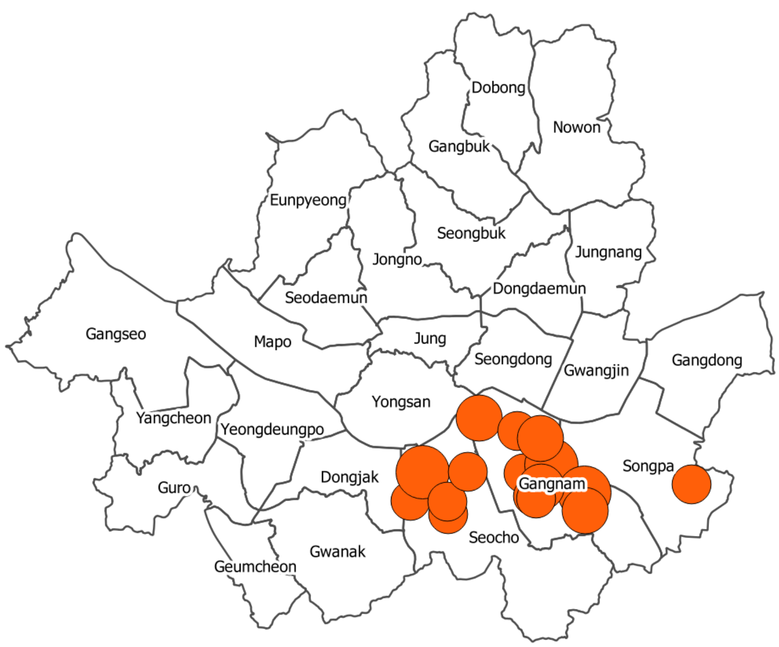
A.2 Table & Figure Appendix

Table A1 Summary Statistic

D. Amenities	**All**			**Gangnam**			**The Other**		
	Average	Median	SD	Average	Median	SD	Average	Median	SD
Manufacturing	147	66	327	72	68	55	162	67	353
Supermarket	43.9	38.0	28.6	45.7	37.0	32.5	43.8	38.0	28.1
Grocery	40.8	31.0	62.6	34.4	32.0	20.4	42.0	31.0	67.5
Cloth	77.3	37.0	147.3	93.6	45.0	141.5	73.4	36.0	147.1
Culture	14.3	11.0	13.1	18.9	15.0	12.8	13.5	11.0	13.0
Restaurant	194.3	150.0	172.0	259.4	197.0	235.5	183.8	144.0	156.8
Bank	7.9	5.0	9.6	13.0	9.0	11.7	7.0	5.0	8.9
Research	4.9	1.0	14.1	9.4	4.5	12.2	4.1	1.0	14.3
Law and Tax	24.1	4.0	99.0	82.3	19.5	232.6	13.3	3.0	29.3
Company HQ	16.2	3.0	49.2	54.5	19.0	94.5	10.0	3.0	33.2
Gov_office	1.7	1.0	1.5	1.6	1.0	1.0	1.7	1.0	1.6
Police, Fire Office	1.4	1.0	1.3	1.2	1.0	1.4	1.4	1.0	1.2
Elementary School	3.5	3.0	2.1	3.2	3.0	1.9	3.5	3.0	2.2
Middle and High School	1.7	1.0	1.8	1.8	1.0	1.6	1.7	1.0	1.8
Private Academy	31.1	22.0	37.3	55.1	36.0	61.0	27.1	21.0	30.3
Clinic	37.3	29.0	37.9	69.7	43.0	75.7	31.7	28.0	23.0
Library	4.2	3.0	5.7	6.1	5.0	4.9	3.9	3.0	5.8
Sport	15.4	13.0	12.6	26.3	25.0	20.3	13.6	12.0	9.7
Sewage	0.4	-	0.7	0.3	-	0.7	0.4	-	0.7
Trash Collection	0.4	-	0.8	0.4	-	0.7	0.5	-	0.8
Trash Disposal	0.1	-	0.4	0.2	-	0.5	0.1	-	0.4
Hotel, Motel	8.4	3.0	15.4	5.5	1.0	11.2	9.0	3.0	16.0
Alchols	88.8	63.0	98.3	110.2	78.0	117.0	85.7	63.0	94.8
Welfare Accommodate	2.2	1.0	2.6	1.6	1.0	1.9	2.3	2.0	2.6
Obs.		423			66			357	
PM10	47.5	47.4	2.1	47.8	48.0	1.4	47.5	47.3	2.2
Obs.		40			5			35	

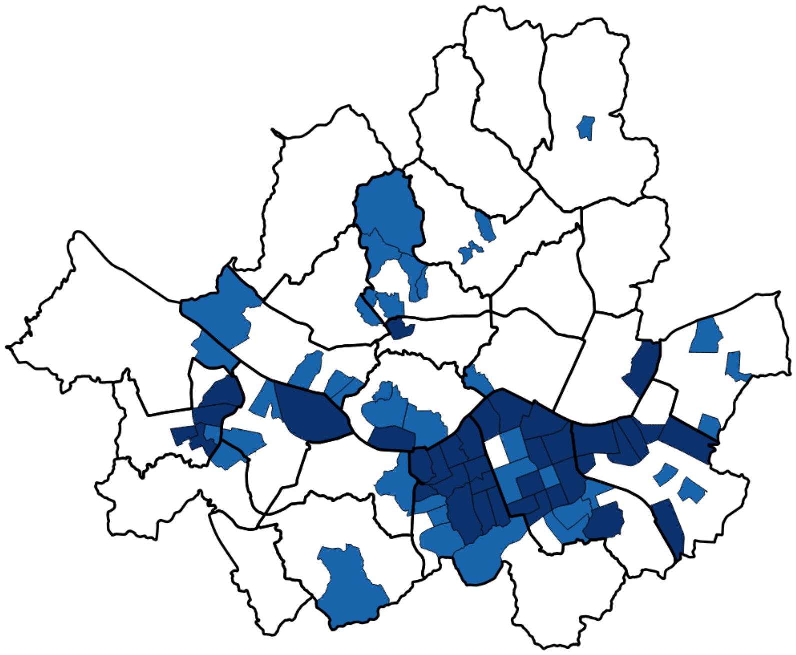


(a) School Zone Map



(b) Top University Entrance Rate (Above 2%)

Figure A1 School Zone, Top Performing High School



(a) Portion of University Graduates (Age 40∼49, Above 62%)

(b) Housing Price per Size (Top 5%)

Figure A2 Neighborhood Segregation and Housing Price

Table A2 School Reputation Effect

– 25 District Fixed Effect –

Dependent Variable = Log(Price)						
Panel A : Including all previous (10 years) results of schools						
	Cross Section		**House Fixed**		**Time-varying**	
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Top Univ. Entrance Rate (t)	2.769***	3.938***	1.083***	1.381***	1.343***	1.489***
	(0.618)	(0.729)	(0.320)	(0.373)	(0.322)	(0.358)
$(t-1)$	4.140***	6.309***	2.716***	1.834***	2.976***	1.595***
	(0.550)	(0.704)	(0.305)	(0.335)	(0.328)	(0.366)
$(t-2)$	-0.865	-1.393	2.048***	1.505***	2.119***	1.237***
	(0.634)	(0.748)	(0.313)	(0.373)	(0.334)	(0.284)
$(t-3)$	2.088**	4.815***	1.702***	0.982***	2.009***	1.191***
	(0.689)	(0.633)	(0.259)	(0.263)	(0.268)	(0.277)
$(t-4)$	0.0687	-0.950	1.393***	1.075***	1.758***	1.077***
	(0.630)	(0.718)	(0.257)	(0.300)	(0.289)	(0.324)
$(t-5)$	3.277***	-1.162	1.329***	0.925**	1.876***	0.607***
	(0.641)	(0.740)	(0.279)	(0.343)	(0.269)	(0.238)
$(t-6)$	-1.797**	-1.699**	0.720***	0.904**	0.946***	0.763**
	(0.563)	(0.622)	(0.213)	(0.344)	(0.215)	(0.312)
$(t-7)$	-4.271***	-5.491***	0.728**	0.786*	0.840***	0.465
	(0.578)	(0.646)	(0.235)	(0.387)	(0.251)	(0.358)
$(t-8)$	-2.653***	-3.257***	0.822***	-0.648*	0.751***	-0.576**
	(0.451)	(0.520)	(0.180)	(0.303)	(0.187)	(0.286)
$(t-9)$	4.672***	3.763***	0.393	0.153	0.120	0.543**
	(0.526)	(0.564)	(0.226)	(0.238)	(0.240)	(0.226)
Sum($t \sim t-7$)	**5.408*** **	**4.366*** **	**11.718*** **	**9.392*** **	**13.866*** **	**8.425*** **
	(1.408)	(1.432)	(1.514)	(1.834)	(1.619)	(1.993)
Sum($t \sim t-9$)	**7.428*** **	**4.872*** **	**12.933*** **	**8.897*** **	**14.737*** **	**8.392*** **
	(1.207)	(1.239)	(1.695)	(2.088)	(1.832)	(2.019)
W/ All lags		**Yes**		**Yes**		**Yes**
Obs.	144,492	144,492	72,246	72,246	72,246	72,246
Adj R^2	0.887	0.890	0.197	0.237	0.991	0.993
Panel B : Including previous 3-year averaged results of schools						
	Cross Section		**House Fixed**		**Time-varying**	
Variable	(7)	(8)	(9)	(10)	(11)	(12)
Top Univ. Entrance 3-Year Average	**8.510*** **	**5.697*** **	**2.369*** **	**0.537**	**2.108*** **	**0.842**
	(1.094)	(0.378)	(0.536)	(0.702)	(0.756)	(0.609)
W/ All lags		**Yes**		**Yes**		**Yes**
Obs.	144,492	144,492	72,246	72,246	72,246	72,246
Adj R^2	0.878	0.889	0.188	0.232	0.989	0.993

Standard errors clustered at IDs are in parentheses
*** p < 0.01, ** p < 0.05, * p < 0.1

Note: All coefficients are reported as β*100. All models include fixed effects (25-district by monthly level) and the other variables listed in Table 1.1 and A1 in the model.

Table A3 Reputation Effect (House Fixed-Effect)

– 11 School Zone –

Dependent Variable = Log(Price)						
Variable	**Distance-Weight Radius : 1km**		**Nearest 3 Schools**		**Radius : 1km Nearest 3 schools**	
	(1)	(2)	(3)	(4)	(5)	(6)
Top Univ. Entrance Rate (t)	1.991***	1.942***	1.175***	0.902***	1.038***	1.149***
	(0.328)	(0.359)	(0.214)	(0.233)	(0.227)	(0.261)
$(t-1)$	3.117***	2.602***	2.113***	1.622***	1.932***	1.732***
	(0.307)	(0.310)	(0.191)	(0.226)	(0.196)	(0.251)
$(t-2)$	2.049***	2.457***	1.511***	1.361***	1.485***	1.754***
	(0.273)	(0.361)	(0.173)	(0.240)	(0.179)	(0.218)
$(t-3)$	1.936***	1.612***	1.688***	1.093***	1.607***	0.983***
	(0.241)	(0.391)	(0.169)	(0.252)	(0.170)	(0.216)
$(t-4)$	1.104***	1.595***	0.697***	1.094***	0.585***	1.044***
	(0.217)	(0.378)	(0.151)	(0.198)	(0.150)	(0.265)
$(t-5)$	1.218***	1.372***	0.632**	0.710***	0.538*	0.779***
	(0.324)	(0.260)	(0.215)	(0.194)	(0.225)	(0.172)
$(t-6)$	1.147***	1.052**	0.427**	0.749**	0.396**	0.536*
	(0.242)	(0.392)	(0.130)	(0.266)	(0.135)	(0.271)
$(t-7)$	0.448*	0.626	0.380*	0.499**	0.324*	0.311
	(0.197)	(0.407)	(0.150)	(0.167)	(0.149)	(0.279)
$(t-8)$	0.619**	0.780**	0.188	-0.0967	0.215	0.580**
	(0.200)	(0.289)	(0.143)	(0.193)	(0.138)	(0.206)
$(t-9)$	0.563**	0.377	0.512***	0.473**	0.338**	0.404*
	(0.177)	(0.239)	(0.117)	(0.152)	(0.118)	(0.167)
Sum($t \sim t-7$)	**13.010*** **	**13.259*** **	**8.623*** **	**8.030*** **	**7.904*** **	**8.289*** **
	(1.505)	(1.936)	(0.955)	(1.165)	(0.995)	(1.276)
Sum($t \sim t-9$)	**14.192*** **	**14.416*** **	**9.323*** **	**8.407*** **	**8.456*** **	**9.273*** **
	(1.654)	(2.161)	(1.061)	(1.306)	(1.096)	(1.435)
W/ All lags		**Yes**		**Yes**		**Yes**
Obs.	72,246	72,246	72,246	72,246	72,246	72,246
Adj R^2	0.194	0.258	0.185	0.235	0.194	0.256

Standard errors clustered at IDs are in parentheses
*** p < 0.01, ** p < 0.05, * p < 0.1

Note: All coefficients are reported as β*100. All models include fixed effects (11 school zone by monthly level) and the other variables listed in Table 1.1 and A1 in the model.

Table A4 Reputation Effect (Time-varying)

– 11 School Zone –

	Dependent Variable = Log(Price)					
Variable	**Distance-Weight Radius : 1km**		**Nearest 3 Schools**		**Radius : 1km Nearest 3 schools**	
	(1)	(2)	(3)	(4)	(5)	(6)
Top Univ. Entrance Rate (t)	2.254***	1.964***	1.176***	1.324***	1.095***	1.207***
	(0.346)	(0.361)	(0.227)	(0.238)	(0.244)	(0.21)
$(t-1)$	3.787***	2.74***	2.421***	1.653***	2.282***	1.926***
	(0.325)	(0.28)	(0.202)	(0.229)	(0.209)	(0.19)
$(t-2)$	2.800***	1.588***	1.975***	1.287***	1.954***	1.325***
	(0.283)	(0.321)	(0.182)	(0.213)	(0.190)	(0.247)
$(t-3)$	2.481***	1.793***	2.092***	1.731***	1.950***	1.06***
	(0.255)	(0.298)	(0.179)	(0.201)	(0.180)	(0.222)
$(t-4)$	1.884***	1.313***	0.965***	1.47***	0.704***	0.799***
	(0.254)	(0.307)	(0.162)	(0.178)	(0.164)	(0.155)
$(t-5)$	1.693***	0.871***	0.844***	1.077***	0.677**	0.9***
	(0.325)	(0.231)	(0.216)	(0.228)	(0.226)	(0.175)
$(t-6)$	0.689***	0.831**	0.643***	0.791***	0.544***	0.552**
	(0.202)	(0.388)	(0.132)	(0.16)	(0.137)	(0.24)
$(t-7)$	1.244***	0.409	0.702***	0.726***	0.618***	0.575**
	(0.237)	(0.359)	(0.153)	(0.152)	(0.153)	(0.266)
$(t-8)$	0.452*	0.736***	0.220	0.056	0.108	0.454**
	(0.219)	(0.273)	(0.152)	(0.178)	(0.149)	(0.188)
$(t-9)$	0.656***	0.847***	0.656***	0.754***	0.452***	0.584***
	(0.182)	(0.226)	(0.122)	(0.143)	(0.121)	(0.153)
Sum($t \sim t-7$)	**16.831*** **	**11.509*** **	**10.818*** **	**10.058*** **	**9.822*** **	**8.345*** **
	(1.615)	(1.725)	(1.018)	(1.045)	(1.075)	(1.166)
Sum($t \sim t-9$)	**17.939*** **	**13.091*** **	**11.694*** **	**10.868*** **	**10.383*** **	**9.383*** **
	(1.935)	(1.955)	(1.144)	(1.183)	(1.196)	(1.255)
W/ All lags		**Yes**		**Yes**		**Yes**
Obs.	72,246	72,246	72,246	72,246	72,246	72,246
Adj R^2	0.991	0.993	0.991	0.993	0.991	0.993

Standard errors clustered at IDs are in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: All coefficients are reported as β*100. All models include fixed effects (11 school zone by monthly level) and the other variables listed in Table 1.1 and A1 in the model.

Table A5 Reputation Effect (House Fixed-Effect)

– 25 District –

Dependent Variable = Log(Price)						
Variable	**Distance-Weight Radius : 1km**		**Nearest 3 Schools**		**Radius : 1km Nearest 3 schools**	
	(1)	(2)	(3)	(4)	(5)	(6)
Top Univ. Entrance Rate (t)	1.618***	1.315***	1.149***	1.112***	1.124***	0.727***
	(0.270)	(0.290)	(0.217)	(0.258)	(0.231)	(0.191)
$(t-1)$	2.598***	2.322***	1.716***	1.121***	1.544***	1.117***
	(0.312)	(0.392)	(0.196)	(0.236)	(0.199)	(0.231)
$(t-2)$	2.142***	1.640***	1.224***	1.023***	1.168***	1.104***
	(0.335)	(0.336)	(0.178)	(0.245)	(0.178)	(0.258)
$(t-3)$	1.303***	1.093**	0.875***	0.638**	0.830***	0.853***
	(0.285)	(0.402)	(0.183)	(0.202)	(0.186)	(0.227)
$(t-4)$	1.313***	1.186**	0.557***	0.668**	0.544***	0.801**
	(0.259)	(0.381)	(0.141)	(0.207)	(0.147)	(0.264)
$(t-5)$	1.109***	1.078**	0.571***	0.546**	0.527**	0.689*
	(0.332)	(0.377)	(0.162)	(0.177)	(0.164)	(0.268)
$(t-6)$	0.663**	0.941*	0.385	0.529*	0.384	0.350
	(0.223)	(0.425)	(0.217)	(0.237)	(0.225)	(0.286)
$(t-7)$	0.668**	0.797	0.424**	0.495	0.380*	0.452
	(0.235)	(0.429)	(0.160)	(0.274)	(0.161)	(0.291)
$(t-8)$	0.636***	0.0130	0.524***	-0.531*	0.0728	0.139
	(0.181)	(0.326)	(0.119)	(0.208)	(0.151)	(0.224)
$(t-9)$	0.348	0.105	0.0602	0.191	0.376**	0.230
	(0.223)	(0.256)	(0.155)	(0.159)	(0.120)	(0.176)
Sum($t \sim t-7$)	**11.413*** **	**10.373*** **	**6.903*** **	**6.133*** **	**6.501*** **	**6.093*** **
	(1.577)	(2.087)	(0.992)	(1.214)	(1.022)	(1.332)
Sum($t \sim t-9$)	**12.397*** **	**10.491*** **	**7.487*** **	**5.793*** **	**6.950*** **	**6.462*** **
	(1.749)	(2.367)	(1.110)	(1.367)	(1.131)	(1.505)
W/ All lags		**Yes**		**Yes**		**Yes**
Obs.	72,246	72,246	72,246	72,246	72,246	72,246
Adj R^2	0.208	0.264	0.197	0.241	0.207	0.262

Standard errors clustered at IDs are in parentheses

*** p < 0.01, ** p < 0.05, * p < 0.1

Note: All coefficients are reported as β*100. All models include fixed effects (25-district by monthly level) and the other variables listed in Table 1.1 and A1 in the model.

Table A6 Reputation Effect (Time-varying)

– 25 District –

Dependent Variable = Log(Price)						
Variable	**Distance-Weight Radius : 1km**		**Nearest 3 Schools**		**Radius : 1km Nearest 3 schools**	
	(1)	(2)	(3)	(4)	(5)	(6)
Top Univ. Entrance Rate (t)	2.262***	1.306***	1.203***	1.184***	1.406***	0.850***
	(0.362)	(0.371)	(0.234)	(0.235)	(0.189)	(0.254)
($t-1$)	2.953***	1.829***	1.972***	1.349***	1.765***	1.302***
	(0.338)	(0.313)	(0.212)	(0.242)	(0.219)	(0.203)
($t-2$)	1.869***	1.736***	1.627***	1.156***	1.168***	0.820***
	(0.286)	(0.301)	(0.187)	(0.213)	(0.254)	(0.181)
($t-3$)	1.758***	1.098***	1.272***	0.986***	1.154***	0.721***
	(0.299)	(0.398)	(0.193)	(0.188)	(0.200)	(0.273)
($t-4$)	1.653***	1.021***	0.691***	0.671***	0.557***	0.648***
	(0.274)	(0.337)	(0.141)	(0.235)	(0.149)	(0.164)
($t-5$)	1.279***	0.676***	0.735***	0.698***	0.602***	0.669***
	(0.335)	(0.244)	(0.165)	(0.222)	(0.170)	(0.224)
($t-6$)	0.713**	0.461	0.599**	0.607***	0.428	0.593**
	(0.229)	(0.378)	(0.217)	(0.165)	(0.227)	(0.236)
($t-7$)	0.536*	0.544	0.624***	0.606***	0.355*	0.410*
	(0.256)	(0.332)	(0.169)	(0.160)	(0.172)	(0.249)
($t-8$)	-0.192	0.163	-0.0367	-0.446**	-0.253	0.102
	(0.241)	(0.298)	(0.161)	(0.191)	(0.160)	(0.200)
($t-9$)	0.509**	0.660***	0.591***	0.454***	0.360**	0.485***
	(0.186)	(0.237)	(0.123)	(0.15)	(0.123)	(0.159)
Sum($t \sim t-7$)	**13.023*** **	**8.671*** **	**8.724*** **	**7.256*** **	**7.435*** **	**6.013*** **
	(1.729)	(1.825)	(1.064)	(1.090)	(1.125)	(1.166)
Sum($t \sim t-9$)	**13.340*** **	**9.494*** **	**9.278*** **	**9.278*** **	**7.542*** **	**6.6*** **
	(1.935)	(2.089)	(1.196)	(1.244)	(1.252)	(1.315)
W/ All lags		**Yes**		**Yes**		**Yes**
Obs.	72,246	72,246	72,246	72,246	72,246	72,246
Adj R^2	0.992	0.993	0.991	0.993	0.992	0.994

Standard errors clustered at IDs are in parentheses

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: All coefficients are reported as β*100. All models include fixed effects (25-district by monthly level) and the other variables listed in Table 1.1 and A1 in the model.

Table A7 WTP for Educational Environments
– 25 District –

	Dependent Variable = log(price)					
	House Fixed-Effect Model			Time Varying Unobservable		
		WTP ($)			WTP ($)	
VARIABLE	**Coefficient**	**(1 unit)**	**(1 SD)**	**Coefficient**	**(1 unit)**	**(1 SD)**
Top Univ Entrance Rate	9.392***	3,024	2,155	8.425***	**2,699**	**1,923**
N-SAT Score	0.044	14	131	0.076***	23	226
Top High School Entrance Rate	-0.116	- 36	- 32	-0.085	- 26	- 23
Private High_school Entrance Rate	0.096	29	140	0.108***	33	157
Neighborhood % BA or Above	0.003	1	10	0.308*	95	**1,091**
Private Academy	0.146*	45	967	0.121**	37	**808**
Student per Class	-0.529***	- 162	- 375	-0.584***	- 179	- 414
Student per Teacher	-0.774***	- 237	- 317	-1.274***	- 389	- 520
% Regual Teacher	-0.041	- 13	- 50	0.018	6	22
Distance to Secondary School				-3.80**	**- 1,190**	- 523
Distance Elementary School				-9.59**	**- 2,925**	- 463
PM10 Air Pollution	-0.729***	- 223	- 255	-0.882***	- 270	- 309

All coefficients are report as β*100.
Willingness to pay is annualized at rate of 7% for mean house price of $438,656
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A8 Housing Price Gap Decomposition
– 25 District –

Variable	Mean value of Variable: Gangnam	Mean value of Variable: The other	Coefficient	($) Price Gap Decomposition	% of Difference(Mean)
Top Univ Entrance Rate	2.02	0.60	8.425***	**68,028**	**19.75%**
N-SAT Score	109.45	100.07	0.076***	4,041	1.17%
Top High School Entrance Rate	2.71	2.09	-0.085	- 303	-0.09%
Private High_school Entrance Rate	12.83	8.53	0.108***	2,626	0.76%
Neighborhood % BA or Above	70.01	43.44	0.308*	**46,470**	**13.49%**
Private Academy	50.61	32.74	0.121**	**12,340**	**3.58%**
Student per Class	33.76	31.40	-0.584***	**- 7,817**	**-2.27%**
Student per Teacher	15.85	14.47	-1.274***	**- 10,028**	**-2.91%**
% Full-time Teacher	84.44	86.36	0.018	- 200	-0.06%
Distance to Secondary School	1.01	1.13	-3.80**	2,427	0.70%
Distance Elementary School	0.33	0.32	-9.59**	- 799	-0.23%
Mean Housing Price (900 ft^2)	$ 757,076	$ 412,610	**SUM**	**116,785**	**33.90%**

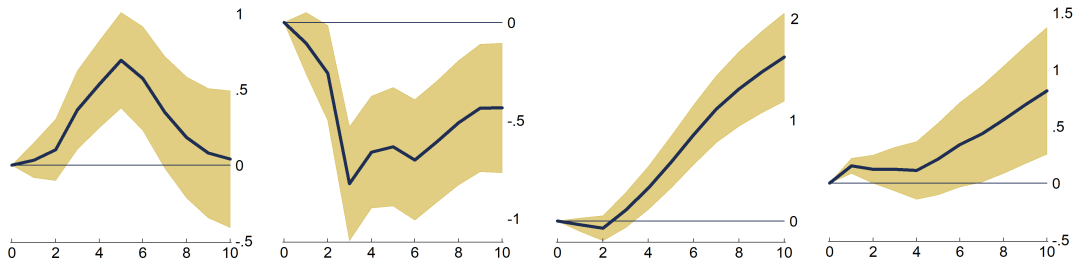
All coefficients are report as β*100.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A9 WTP for the School Zone Adjustment without Weight

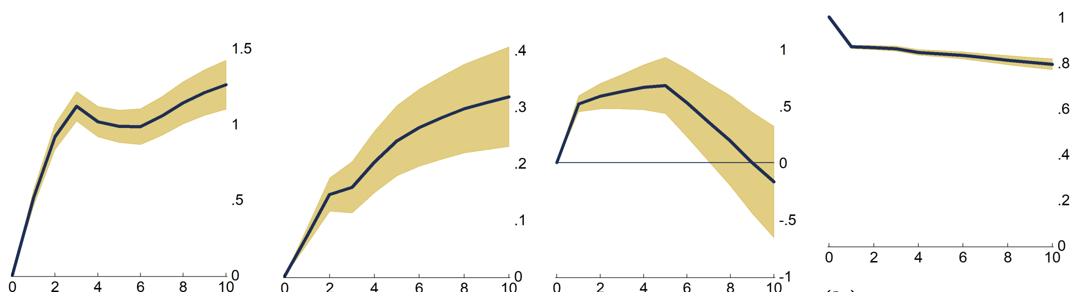
Area	Time	WTP %		WTP($) / Year		Obs
		Median	Mean	Median	Mean	
All	Short-term	0.32	0.32	83	65	34,818
	Long-term	2.87	- 0.05	548	- 660	
Lower Score	Short-term	0.48	0.69	119	187	22,315
	Long-term	3.43	3.27	775	807	
Higher Score	Short-term	- 0.27	- 0.33	- 132	- 152	12,503
	Long-term	- 5.85	- 5.98	- 2,846	- 3,278	

Table A10 The Changes in Experimental Area without Weight

Area	Time	Housing Price ($1,000)			Top Univ. Entrance (%)			BA or Above (%)			Obs
		Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	
	Initial	427.0	510.1	280.2	0.61	1.18	0.95	52.94	56.71	12.95	
All	Short-term	429.0	511.0	278.9	0.74	1.20	0.88				34,818
	Long-term	434.2	500.6	255.5	0.87	1.18	0.66	53.66	56.26	10.09	
	Initial	345.2	365.4	135.7	0.45	0.54	0.31	49.43	48.38	6.18	
Lower Score	Short-term	347.8	368.1	137.1	0.51	0.63	0.37				22,315
	Long-term	356.4	376.9	138.7	0.71	0.79	0.26	50.87	49.81	5.07	
	Initial	736.8	768.3	285.8	2.40	2.34	0.51	72.35	71.59	7.34	
Higher Score	Short-term	734.3	766.1	285.5	2.29	2.22	0.55				71,046
	Long-term	690.6	721.4	267.1	1.73	1.87	0.59	68.31	67.77	5.55	

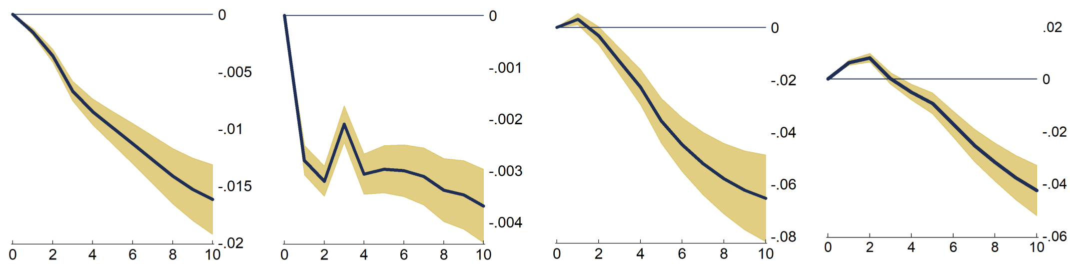


(a) **Student/Class** (b) **Student/Teacher** (c) **Bank** (d) **Culture**

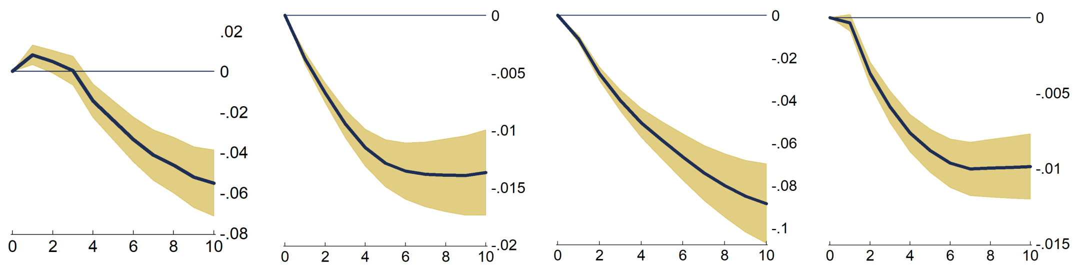


(e) **Sports** (f) **Restaurant** (g) **Population (%)** (h) **Housing Price (%)**

Figure A3 (Cumulative) IRF
– Response = Housing Price (%) –



(a) BA (%p) (b) Top Univ (%p) (c) Private Academy (d) Population (%)



(e) Employment (%) (f) Bank (g) Culture (h) Library

Figure A4 (Cumulative) IRF
– Impulse = Housing Price (1%) –

A.3 Structural Panel VAR

$$\Phi Y_{it} = \sum_{p=1}^{T} \Gamma_p Y_{it-p} + \Theta X_{it} + D_{jt} + \xi_i + \varepsilon_{it} \tag{6}$$

This paper also adopts a structural panel VAR model, which is presented in equation (6) to consider the simultaneous relationships among variables. Unlike the panel VAR model (equation (5)) in Section 1.5, the model includes the parameter matrix Φ defining the contemporaneous relations among dependent variables in the system. Since the structural model requires restrictions for the parameter specification, to apply restrictions such as variable ordering, the model reduces the number of endogenous variables and includes the key variables of this research. Specifically, the structural panel VAR model includes the neighborhood composition (percent of BA or above degree holders in the 40∼49 age group), top university entrance rate, log of housing price, population growth, and log of employment (full- and part-time employment) as endogenous variables and controls for all the other variables as exogenous variables.[60]

Note that in the structural model that has simultaneous relationships among endogenous variables, the variance-covariance matrix of the shocks is not likely to be diagonal. In other words, a shock is correlated with shocks of other variables and they are not independent of each other. Thereby, we need to make them orthogonal to isolate shocks uncorrelated with other shocks. In this process, this paper uses the following restrictions. The variable ordering this paper assumes is: top university entrance rate - neighborhood composition - population growth - employment - housing price. That is, we assume that the variable on the left side contemporaneously affects variables on its right side, however, the opposite is not true. The increase in the top university entrance rate of a school can attract more educated people to the neighborhood around the school at the same time. Given that, however, it is rare for households to move during the academic year when their children take the university entrance exam, this paper assumes the neighborhood composition cannot contemporaneously affect school performance. Also, based on the impulse-response analysis in Section 1.5, the housing

[60] That is, the number of students per class and per teacher, the number of private academies, libraries, supermarkets, culture related stores, shopping stores, clinics, restaurants, sport facilities, banks, and grocery stores around a house. Also, the model controls distances from a house to the nearest subway station, department store, supermarket, elementary school, middle school, and high school as exogenous variation.

price is treated as the most endogenous variable among the five variables. For the estimation, the first differenced variables are used and IV-GMM is adopted as in Section 1.5.

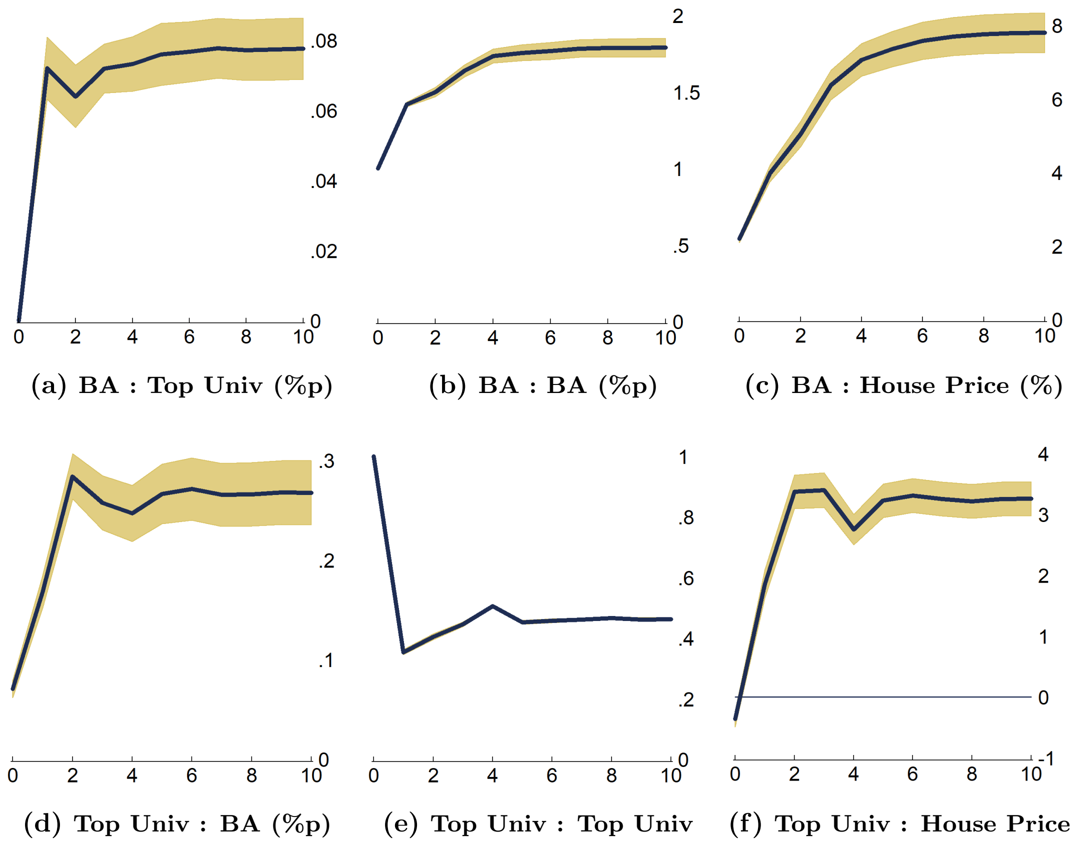
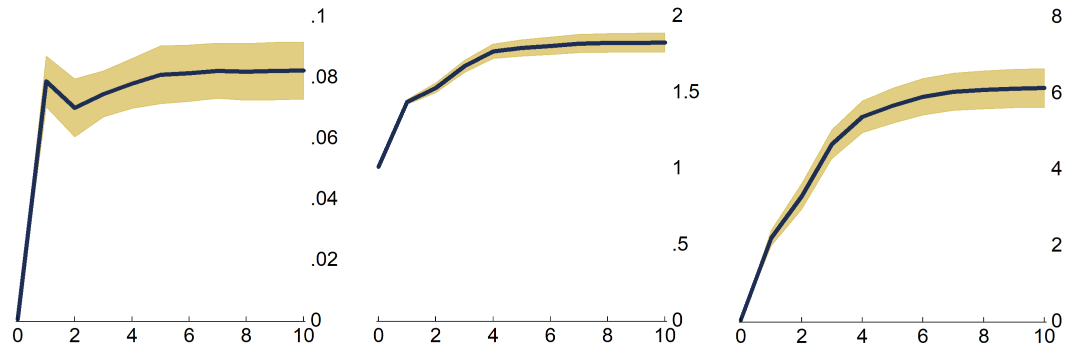


Figure A5 Structural Panel VAR, (Cumulative) IRF

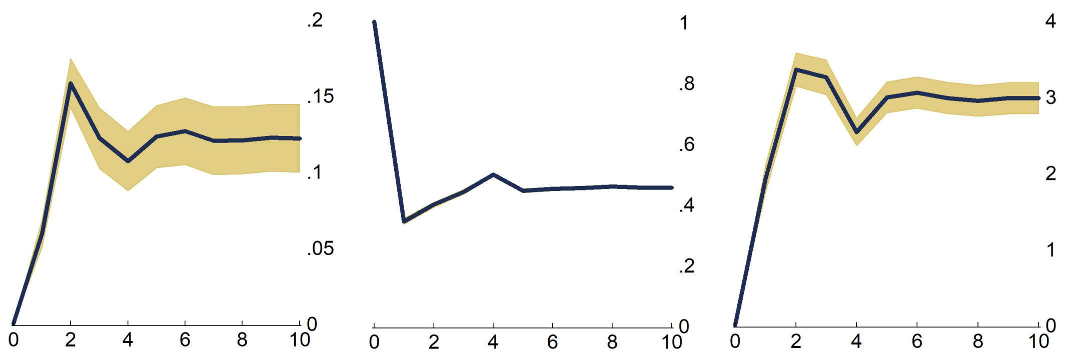
Figure A5 shows the results from the structural panel VAR adopting Cholesky decomposition with the ordering, and the impulse response functions are generated from the orthogonalized shocks. According to the results, compared to the results in Section 1.5, the duration and the size of the effect from a shock becomes shorter and smaller compared to the results presented in Figure 1.5. The differences between these two results can be caused by two sources. First, the diminished number of endogenous variables in the system VAR could simplify and eliminate relationships among variables. Second, considering the contemporaneous relation among endogenous variables, which is represented by the parameter matrix Φ, can lead to the differences.

To figure out the factor driving the differences, this paper also estimates the equation (5) while using the same variables as with the structural panel VAR model.

That is, for the estimation, five variables – in the structural panel VAR model – are used as endogenous variables, and all the other variables are controlled as exogenous variables. However, the model does not consider the contemporaneous relations among endogenous variables. Thus, the model is a non-structural panel VAR model which uses five endogenous variables.



(a) BA : Top Univ (%p) **(b)** BA : BA (%p) **(c)** BA : House Price (%)



(d) Top Univ : BA (%p) **(e)** Top Univ : Top Univ (%p) **(f)** Top Univ : House Price (%)

Figure A6 Panel VAR, (Cumulative) IRF
– (Impulse) : (Response) –

Figure A6 shows the results of the model. Note that the estimated impulse-response functions are quite similar to the results in Figure A5. This result implies that the differences – the shortened durations and the smaller size of effect from a shock – are mainly caused by the diminished number of endogenous variables that eliminate the relationships among amenities and attributes of a house.

Chapter 2. Impact of Resale Price Maintenance (RPM) Regulation on the Book Market of Korea

2.1 Introduction

Resale Price Maintenance (RPM) regulations in book markets are now in force in many countries in Asia and Europe, while it was abolished a long time ago in the United States and the United Kingdom.[61] Although RPM is not allowed generally, many Asian and European countries made an exception for their book markets, considering a book to be unique goods, which is different from other goods in the market. They allow RPM in a book market under the policy purpose of protecting a variety of cultures and knowledge. The proponents of RPM regulation for books argue a proper price should be guaranteed for a book – especially if the book is not popular – to survive in the market, and it is justified to protect the value of cultural variety and knowledge.

Even though "A Book is Different" in terms of its cultural and knowledge perspective, "A Book is the Same" in that it cannot be an exception of the law of demand and supply in the market. First of all, a decent margin of a book by fixing the retail price does not guarantee a higher profit for the author and publishing company. Because the higher price will reduce the demand for the book, fixed retail price can help the publishing company (the price setter under the RPM regulation) only if they can set the optimal price by accurately forecasting the market demand when they publish a book. However, it is generally known that forecasting the demand for a new book is hard. Unexpected new publications of similar books and social trends make the demand for a book highly volatile.

Second, a higher retail price also does not mean a higher wholesale price for the publishing company and higher royalty for the author of a book. If some retailers have overwhelming bargaining power over the publishing companies, resale price maintenance regulation can be used as a method to help increase the profit of market-dominant retail firms at the cost of diminishing consumer surplus, profit of publishing

[61] Fifteen OECD countries, ten of which EU members, have regulation for fixing the price of books. In the Unites States, the resale price maintenance was the per se illegal under 1 of the Sherman Act, 15 U.S.C. 1. However, a decade ago, the U.S. Supreme Court overturned a nearly century-old precedent when it issued its *Leegin* opinion. After that, the resale price maintenance is considered to be restrictedly allowed if there is a reasonable reason. In the United Kingdom, the RPM for a book took effect in 1900 and abolished in 1997 by the Restrictive Practices Court

companies and authors. De los Santos and Wildenbeest (2017) analyze the fixing e-book price case of the United States. In 2010, five of the six largest publishing companies concurrently adopted the agency model where publishers set retail prices. They empirically analyze the effect of vertical price restraints on the price of e-books. They show that e-book prices were 18 percent higher at Amazon and 8 percent higher at Barnes & Noble when publishing companies adopt the agency model which controls the retail price. However, it is important to note that the higher retail price during the agency period did not enhance the profit margin of publishing companies. On the contrary, publishers' profit margins from e-books decrease under the agency period mainly due to higher commission fee for the retailers.

In most cases, online companies sell books at lower prices than local offline stores because they have more efficient cost structures compared to the small offline bookstores. Online firms operate Internet servers and large-scale distribution warehouses instead of offline stores, and therefore decrease the average cost by the economy of scale. On the other hand, offline stores should pay rent for the store and a substantial cost for displaying the books and operating the store. Also, note that major online firms can buy books from publishing companies at lower prices than offline book-stores because big online firms have higher bargaining power based on their dominant market shares.[62] According to a recent survey (The Survey on the Book Publishing Industry) which is investigated by the Statistics Korea, online companies – of which sales are over 10 billion (₩) – buy a book from publishing companies on average at 63.6% of the list price. However, the small offline stores (sales below one hundred million (₩)) buy books from publishing companies on average at 70.1% of list prices. In Korea, There were 3,429 offline book-stores in 2005; however, the number decreased to 2,116 (– 38.3%) in 2015 while top online firms have continuously expanded their market shares.[63] The number of small offline book stores was expected to decrease more given the price gaps between online and offline stores based on the cost efficiency and the irreversible trend of e-commerce.

In these circumstances, on the purpose of protecting cultural variety and helping small publishing companies and offline local bookstores, the National Assembly of

[62] For example, the Kyobo Book Centre (No.1 book company in Korea) alone accounts for 25% of the total book market share in terms of sales.

[63] The top 7 online firm's market share was over 50% in 2012 and have been continuously increasing.

Korea legislated the new RPM regulation (reinforcing the price regulation on book markets), which took effect in November of 2014, despite oppositions from consumers. Table 2.1 shows what was changed in the RPM regulation from November 2014.

Table 2.1 Amendment of RPM Regulation

	RPM regulation (Before 11.2014)	**New RPM regulation (After 11.2014)**
Regulation Coverage	Newly-published book[64] (published within the last 18 months)	All books
Discount	Up to 19% from list price • Direct discount up to 10% • Indirect discount up to 9% (coupon, points, other benefits)	Up to 15% from list price • Direct discount up to 10% • Indirect discount up to 5% (coupon, points, other benefits)
Changing List Prices		Publishing companies can change the list price only after 18 months pass from the new-release

Before the New RPM Regulation

Before the new regulation was in effect, in most cases, a local offline store had sold its book at its list price while an online firm had usually sold it at a 10% discounted price (maximum discount rate by the regulation) for newly-published books. However, the price gap between online and offline stores was more prominent for books published more than 18 months ago. No price regulation for the non-newly-published books[65] and price competition among online firms make a prices fall to its marginal cost. Moreover, online users can easily compare the online firm's price because search engines such as NAVER.com and DAUM.net offer the price comparison among all online book-stores[66] and it makes search-cost almost zero. If users enter the title of a book, they show all prices of each online stores for comparison. Because of the above reasons, prices of non-newly-published books were usually much lower in online stores than local offline stores. That is, the price gap was just 10% for a newly-published book; however, in case of a non-newly-published book which is published more than 18 months ago, the

[64] According to the regulation, a book is categorized as a newly-published book if the book was published in recent 18 months. However, if a book is published more than 18 months ago, it had not been under the RPM regulation until 11. 2014

[65] In this paper, a non-newly-published book means a book which is published more than 18months ago.

[66] In Korea, NAVER and DAUM have more market share than Google. Each portion was 80% (NAVER), 13% (DAUM) and 5% (Google) in 2015.

price gap between online and offline store was larger than 10% until the new RPM regulation took effect.

After the New-RPM Regulation

The most critical change in the reinforced new RPM regulation is that the books which were published more than 18 months ago are newly included in the new regulation coverages. Many countries are enforcing resale price maintenance regulation on books. However, no country has price regulation for books published a long time ago[67] because price regulation for all kinds of books can cause some problems.

First of all, regulating the price of all books can make serious inefficiency in the inventory disposal process. It was possible under past RPM regulation to lower the price of unpopular books if a book store wanted to sell them. However, this is not the case under the new regulation system. Now, the bookstore should wait until the publishing companies change the list price of a book. A worse problem is that to change the list price, publishing companies should recall all books from the bookstores and re-distributes again after changing the list prices. Before the new RPM was in effect, the inventory disposal was processed efficiently through online book stores. The cost for changing the price (menu cost) of a book at online stores is almost zero because changing the price on an Internet server is not costly and this process can be done systematically by a program accoroding to the situation of inventory stock. However, under the new RPM system, the market no more can use the efficient inventory disposal process of online firms.

That is, the new RPM regulation restricting prices of all books increase inventory disposal cost. Because of the increased burden of inventory controls, it is also possible that bookstores become reluctant to buy and retain unpopular books from publishing companies. Then, it is against the purpose of the new policy to support the survival of unpopular books. Moreover, if the new regulation increases menu cost, inventory management cost and logistical cost, retailers (online firms and small offline stores) and publishing companies should share and divide the extra burdens, and it is likely that they divide the extra burden depending on the bargaining power between a publishing

[67] Even in the countries enforcing more restrict resale price maintenance on book market such as France and Germany, the RPM regulation is applied for a newly-published books for a limited time (two years in France and 18 months in Germany)

company and an online store or an offline store. Thus, one might expect that publishing companies and small offline stores might be not in a favorable position in this process when considering the dominant market share of large online firms.

The proponents of new RPM regulation anticipate that some portions of online firm's customers move to local offline stores because of a diminished price gap between online and offline stores and it would help the business of small offline local stores. However, converted demands from online stores to offline stores could be smaller than expected for some reasons.

First, even though the regulation restricts the indirect discount up to 5% of a list price, online firms can legally avoid the law by finding the blind spot of the regulation. According to the report of the Publication Industry Promotion Agency of Korea (KPIPA, 2016), most online firms are using various alliance credit card discounts which are not under the price restriction. For example, the YES24 adopts more than 30 alliance credit cards, and if customers use one of these credit cards, they can get more discounts over the regulation limit. Moreover, some indirect discounts are hard to be measured accurately; thus it is difficult for the authorities to regulate it. Also, lots of other benefits such as gifts and coupons for other goods and services are not under the price restriction.[68] Because of these problems, according to a recent publication industry survey administered by KPIPA, more than 60% offline stores responded that there should be more restrictions about indirect discounts of online stores including alliance credit card discounts.

Second, most online firms are also operating offline stores in addition to their online stores. They are usually large-scale stores and also located in a good place such as the central business district of a city. Thus, it is possible that a large portion of converting demands from online to offline can be absorbed by the large offline stores operated by online firms. After the new regulation, the profit of online firms significantly increases[69] and they have expanded their business aggressively opening many large-scale offline stores based on their increased profits. Kyobo Book Centre (No.1

[68] According to a recent news report (Large Online Book Stores Out of RPM Regulation, 4.14.2017, Seoul Economy, http://www.sedaily.com/NewsView/1OEN7SNHXB/GK01) a person obtained various coupons of which total value reach 50% of a book price after he bought a book. However, surprisingly, all gifts and discount coupons he obtained on the day did not violate the indirect discount regulations of the new RPM regulation.

[69] The empirical analysis about this point is in Section 2.3.

book store) has opened 17 more offline stores (among their 39 offline stores) after the new RPM regulation took effect.[70] Aggressive offline entry of large online companies could make the converted demand from online to offline small local stores more difficult.

The bottom line is that various indirect discounts out of law's restriction, large-scale offline stores operated by online firms of which numbers are sharply increasing recently make it harder to achieve the policy purpose of new RPM regulation. However, paradoxically, all the above things are caused by the new RPM regulation itself. Note that online companies having grip on the book market can avoid price regulations by using various methods explained above. Due to the increased price margin of a book, large online firms' profit ratio and the amount of profits are enhanced significantly after the new regulation took effect as I show the empirical evidences in the following Section 2.3. As a result, the market share of top 7 online firms even increase rapidly after the new regulation took effect even though the new regulation intended the opposite results.

At the same time, the amount of sales of publishing companies and small offline stores have decreased significantly after the new regulation. In addition, after the new regulation, the number of newly-published book which can be used as the index for cultural variety is decreased by 7% annually after de-trending. Even though the RPM regulation took effect under the purpose to preserve cultural variety and knowledge by supporting small publishing companies and book stores to survive in the market, many empirical evidence imply the opposite results.

Also, the new RPM regulation could affect the average price of books, transaction quantity, and consequently the consumer surplus. Cho (2015) shows the new regulation increases the price of best-seller books by 16.2% and steady-seller books by more than 40%. The study also estimates that the new regulation results in the decrease of book transactions by 12.1%. In terms of welfare change, he estimates the consumer surplus decreases more than 13 billion (₩). Among them, 9 billion (₩) is transferred to online firms and offline stores lose 3.5 billion (₩) owing to the new regulation.

This paper proceeds as follows: Section 2.2 describes and analyzes the data illustrating the changes in the whole book market situation before and after the new

[70] The Aladin has opened 17 out of 37, the Yongpung has opened 9 out of total 30, and the Bandi & Lunis have opened 4 out of total 14 offline stores during the same short period. The YES24 (the second largest online book store) opened their first offline store last April. 2016 and have opened four more offline stores until now.

RPM regulation. More specifically, Section 2.3 investigates the effect of new regulation on each firms based on firm-level dataset and various surveys. Section 2.4 concludes and summarize the empirical analysis of this paper.

2.2 Impact of RPM Regulation on the Book Market

2.2.1 Market Size and Online Share

Table 2.2 The Book Market of Korea

Year	(1) Market Size 1		(2) Expediture / Household		(3) Expenditure / Person		(4) Market Size 2		(5) On-line Market Size		(6) Online Share (5)/(1)	(7) Online Share (5)/(4)	(8)=(1)-(5) Offline Market Size	
2016	26,288	**-4%**	12,066	**-8%**	51,299	**-4%**			13,406	**16%**	**51%**		12,882	**-19%**
2015	27,331	**-7%**	13,108	**-10%**	53,575	**-8%**	25,385	**-14%**	11,512	**-10%**	**42%**	**45%**	15,819	**-5%**
2014	29,471	0%	14,614	-4%	58,074	-1%	29,438	16%	12,804	7%	43%	43%	16,667	-5%
2013	29,465	0%	15,147	-2%	58,428	-1%	25,397	5%	11,961	-6%	41%	47%	17,503	4%
2012	29,498	-8%	15,502	-8%	58,761	-8%	24,133	-15%	12,728	0%	43%	53%	16,770	-13%
2011	31,963	-5%	16,937	-6%	64,006	-5%	28,504	5%	12,743	9%	40%	45%	19,220	-12%
2010	33,530	3%	17,939	3%	67,664	2%	27,258	0%	11,691	14%	35%	43%	21,839	-2%
2009	32,600	-6%	17,484	-6%	66,116	-7%	27,244	6%	10,298	18%	32%	38%	22,302	-14%
2008	34,738	4%	18,586	5%	70,815	3%	25,810	-18%	8,752	18%	25%	34%	25,986	0%
2007	33,377	-2%	17,710	-5%	68,559	-2%	31,461	22%	7,442	19%	22%	24%	25,935	-6%
2006	33,959		18,607		70,107		25,810		6,277		18%	24%	27,682	

Table 2.2 shows the estimated size of the book market and online firms' market share by year. The amount of market size in column (4) is estimated and published by the Korean Publishers Association (KPA). According to KPA, they use the following formula to calculate the total market size: the average price of books $\times$ the number of calculation for a year $\times$ 2, by every 12 categories and the summation of the total amounts becomes the estimated market size. Thus, it is multiplication of Table B2 and Table B3.[71] On the other hand, I estimate the total market size presented in the column (1) by using the two statistics published by Statistics Korea. It publishes the average expenditure on books by the household size and also publishes the number of household by the number of household member. Thus, we can estimate the total amount of expenditure on books by using the two statistics. The column (2) shows the average amount of expenditure on books per household for a year, which is also published by Statistics Korea. The column (3) comes from the calculation that the estimated market size in the column (1) is divided by the total population of Korea.

[71] However, the calculation results are different from column (4), because the final amount is decided after an expert meeting about the total market size based on the calculation. The ground for the multiplication of 2 in the KPA's formula is based on the rough estimation that the total amount of transaction for the books published more than one year ago is similar with that for a newly-published book during the year.

Thus, it means the average expenditure for books per person. The column (5) shows the total amount of transactions for books at online stores which comes from the Online Shopping Survey by Statistics Korea. The column (6) and (7) show the market share of online stores by using the ratio (5)/(1) and (5)/(4).

According to Table 2.2, the total market size, the average expenditure on book per household and per person are all decreased after the new RPM regulation. Although all of them were on the decreasing trends, the new regulation could not change the trend and the decreasing trend even speeds up. The online sales decreased just after the new regulation in 2015, however, it rapidly rebounded in 2016. As a result, the market share of online exceeded 50% for the first time while it stayed around 40% before the new regulation. On the other hand, the amount of offline sales continued to decrease. It even decreases by 19% in 2016. Based on the above statistics, it is hard to say some demands are converted to offline stores as the new RPM regulation intended. Even after the new regulation took effect, both increasing trend of online and decreasing trend of offline have accelerated. Figure 2.2 shows the trends of total book market size, the number of online sales and the online market share.

Table 2.3 The 7 Biggest Online Firms

Year	(1) Top 7 Firms Sales[72]		(2) Top 7 Market Share Maket 1	(3) Market 2	(4) Top 7 Firms Profit[73]		(5) All the Other Firms Sales	
2016	16,979	**8%**	**65%**		4,826	**12%**	9,310	**-19%**
2015	15,791	**-1%**	**58%**	**62%**	4,302	**16%**	11,541	**-15%**
2014	15,954	0%	54%	54%	3,694	5%	13,517	0%
2013	15,934	1%	54%	63%	3,520	1%	13,531	-2%
2012	15,728	0%	53%	65%	3,492	4%	13,770	-15%
2011	15,685	2%	49%	55%	3,362	0%	16,278	-10%
2010	15,401	8%	46%	57%	3,377	0%	18,129	-1%
2009	14,326		44%	53%	3,387		18,274	

Table 2.3 shows the top 7 online firms' amount of sales, profit (= sales - cost) and their market share by year. In 2015, the sales of top 7 firms decrease a little (-1%), however, the amount of profit increase by 16% owing to the increased price margin.[74]

[72] The top 7 (in terms of sales) firms in the book market is 1) KYOBO 2) YES24 3) ALADIN 4) INTERPARK 5) YOUNGPUNG 6) BANDI & LUNIS 7) LIBRO.

[73] This amount of profit is sum of 6 firms except INTERPARK. Because INTERPARK is operating 4 business parts and they have only published the sales of each business, I could not figure out the cost of INTERPARK book business part.

[74] Table 2.11 shows the enhanced profit ratio of top 7 online firms. Before the new regulation, there was no price regulation for the books which were published more than 18 months ago. Price competition among

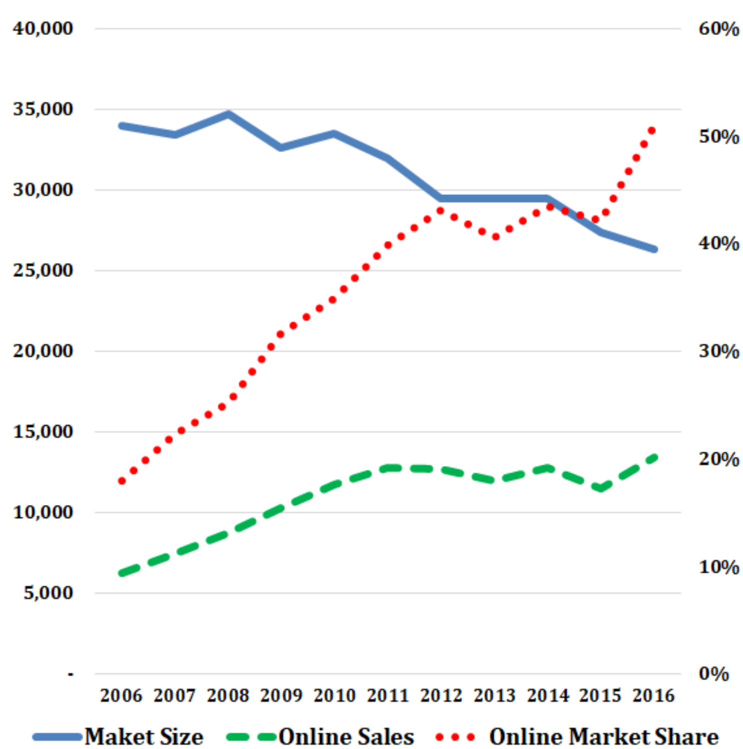


Figure 2.1 Market Size and On-line Share

In 2016, their amount of sales rebounded and increased by 8% while the amount of profit increase 12%. If we check the top 7 firms' market share and all the other firms' sales, the Top 7 online firms' market share rapidly increases after 2015 and it was over 60% in 2016. On the contrary, all the other firms' sales decrease by 15% and 19% after 2015. In summary, the degree of market concentration is aggravated after 2015.

2.2.2 Variety of New Book Publications

In addition to supporting the business of small publishing companies and offline book stores, protecting the cultural variety and the value of knowledge was another important aim of the new RPM regulation. The variety of newly-published books can be an index to evaluate the effect of the reinforced RPM policy on a cultural variety. Table 2.4 shows the number of newly-published books by 11 categories by year. This paper uses the book registration dataset from the National Library of Korea (the copyright library of Korea)[75] in which all new publications are registered.

Table 2.5 shows various regression results by using the data of Table 2.4. For the analysis, the equation (7) is used as the baseline model to estimate the effect of new RPM regulation on the number of newly-published books.

$$log(New\ Publication)_{it} = \alpha_t + \beta_{1t}\,RPM15_t + \beta_{2t}\,RPM16_t + \delta_t\,log(Price_{it}) + \xi_{it} \quad (7)$$

The dependent variable is the log of the number of newly-published books of book category i at time (year) t. RPM15 and RPM16 represent time dummy variable for year 2015 and 2016 respectively. The model also uses the average price of books of the category i at time (year) t as a control variable, because the average price of books could

online book stores and zero search cost drove the price to its marginal cost. However, resale price maintenance regulation restricting price competitions among them enhanced their price margin significantly.

[75] According to the library law, all newly-published books should be registered on the National Library of Korea within 30 days from the new publication. There is another dataset about statistics of the number of new books published by the Korean Publishers Association (KPA). However, the statistics from KPA only included the books which were registered on the National Library of Korea through KPA's system. Because of direct registration to the National Library of Korea, KPA dataset does not include the all-new publication book. Table B1 in the appendix shows the data from KPA. Publication Industry Promotion Agency of Korea (KPIPA) also compiles and publishes the statistics about newly-published books. However, it extended its data source in 2015 and 2016. In 2015, books recorded in Yongpung and Bandi & Lunis are included as the data source, and they again added Aladin in 2016 to make the new publication statistics and did not retroactively compile the previous data with the same standard. Thus, we cannot compare the statistics comes from the broadened data source with the previous statistics. KPA also broadened their data source to make new publication statistics from 2016 and also did not retroactively compile the previous statistics based on the same data source.

Table 2.4 The Number of Newly-Published Books by Categories

Category	2009	2010	2011	2012	2013	2014	**2015**	**2016**
General	1,992	1,829	1,870	1,610	1,628	1,994	1,946	2,545
Philosophy	1,262	1,390	1,395	1,665	1,892	2,074	2,084	1,505
Religion	3,906	3,595	3,788	3,639	3,503	3,459	3,169	3,132
Social Science	13,248	13,158	13,542	12,717	13,307	14,559	13,612	13,141
Science	964	981	968	855	960	1,064	1,117	900
Engineering	6,966	7,043	7,666	7,441	8,068	8,149	9,242	8,871
Art	2,471	2,521	2,556	2,638	2,669	3,095	3,200	2,898
Language	2,122	2,420	2,536	2,422	2,582	2,685	2,216	2,114
Literature	10,042	10,662	10,501	10,738	11,240	13,502	12,591	12,162
History	2,085	2,202	2,145	2,128	2,221	2,459	2,364	2,208
Children	12,417	12,154	12,722	12,062	10,133	9,032	7,872	7,715
Total	57,475	57,955	59,689	57,915	58,203	62,072	**59,413**	**57,191**

Data : National Library of Korea

affect the number of new publication. Lastly, ξ_{it} denotes the omitted or unobservable attributes affecting the number of newly-published books of category i at time t. By using two consecutive years of data and additionally assuming that the unobservable attributes of a house and the implicit prices for time-varying variables do not change over two years ($\beta_{t'} = \beta_t = \beta, \;\; \delta_{t'} = \delta_t = \delta, \;\; \xi_{it'} = \xi_{it}, \, (t' > t)$), the equation (7) can be re-written as the following model by first differencing:

First-differencing Model

$$\Delta log(New\ Publication)_i = \Delta\alpha \, + \beta_1 \, \Delta RPM15 + \beta_2 \, \Delta RPM16 + \delta \, \Delta log(Price)_i \quad (8)$$

This study also flexibly allows the unobservables to vary within the two consecutive years given the assumption that it follows the first-order Markov process:

$$\xi_{it'} = \gamma \, \xi_{it} + \eta_{it'} \qquad (t' > t) \qquad (9)$$

here, $\gamma\xi_{it}$ is the expected value of the unobservable attributes at time t', and $\eta_{it'}$ is the stochastic innovation in the unobservable attributes. Note that this study assumes $E[\eta_{it'}|I_t] = 0$, where I_t denotes the available information at time t. Then, equation (7) can be re-written as follows by using the Markov process assumption.

Time-varying Unobservable Model

$$
\begin{aligned}
log(New\,Books)_{it'} &= \alpha_{t'} + \beta_{1t'}\,RPM15_{t'} + \beta_{2t'}\,RPM16_{t'} + \delta_{t'}\,log(Price_{it'}) + \xi_{it'} \\
&= \alpha_{t'} + \beta_{1t'}\,RPM15_{t'} + \beta_{2t'}\,RPM16_{t'} + \delta_{t'}\,log(Price_{it'}) + \gamma\xi_{it} + \eta_{it'} \\
&= \alpha_{t'} + \beta_{1t'}\,RPM15_{t'} + \beta_{2t'}\,RPM16_{t'} + \delta_{t'}\,log(Price_{it'}) \\
&\quad + \gamma[log(New\,Books)_{it} - \alpha_t - \beta_{1t}\,RPM15_t - \beta_{2t}\,RPM16_t \\
&\quad - \delta_t\,log(Price_t)] + \eta_{it'} \\
&= (\alpha_{t'} - \gamma\alpha_t) + \gamma log(New\,Books)_{it} + \beta_1[RPM15_{t'} - \gamma RPM15_t] \\
&+ \beta_2[RPM16_{t'} - \gamma RPM16_t] + \delta[log(Price_{it'}) - \gamma log(Price_{it})] + \eta_{it'}
\end{aligned}
\tag{10}
$$

Table 2.5 Effect of RPM on the Number of New Publications

	Dependent = Log(Number of Newly-Published Books)					
VARIABLES	(1) FD	(2) FD	(3) FD-IV	(4) TV	(5) TV	(6) TV-IV
Constant	.0241*	.0307**	.0277	.1144	.1180	.1047
	(.0125)	(.0131)	(.0183)	(.1013)	(.1354)	(.1113)
RPM15	**-.0610****	**-.0691****	**-.0649***	**-.0599***	**-.0675****	**-.0600***
	(.0305)	(.0308)	(.0353)	(.0312)	(.0309)	(.0336)
RPM16	**-.1425*****	**-.1574*****	**-.1501*****	**-.1401*****	**-.1541*****	**-.1406*****
	(.0466)	(.0474)	(.0564)	(.0466)	(.0474)	(.0521)
Log(Price)		-.2636	-.0767		-.2425	.0514
		(.1787)	(.4404)		(.1792)	(.2004)
γ				.9889***	.9916***	.9896***
				(.0122)	(.0136)	(.0139)
Observation	77	77	66	77	77	66
Adj. R^2	.1149	.1407	.1281	.9889	.9892	.9879

Standard errors clustered at IDs are parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2.5 shows the various estimation results by using above models. Column (1), (2) and (3) show estimation results from the first differencing model (equation (8)), and column (4), (5) and (6) come from the time-varying unobservable model (equation (10)). The column (3) and (6) adopt two-stage least square (2SLS) method to control a simultaneous relation between price and new publication of books. I use the average price of books in the previous year and all other predetermined variables as instruments for the average price of books. It is reasonable that higher average price leads to more publication. In column (6), the coefficient changes to positive but is not significant.

According to various panel regression results, the number of newly-published books decreases by 6∼7% in 2015 after de-trending. In 2016, the accumulated amount of decrease reaches 14∼15% implying the amount of decrease in 2016 is about 8∼9%. Based on the above results, it is hard to insist that the new RPM regulation help to enhance the number of new publication as an index for a variety of knowledge and culture. According to the result of the column (6), the time-varying unobservable model using instrument variables, the number of newly-published books decreased by 6% in 2015 and 8% more in 2016.

2.2.3 Price of Books

Another controversial issue of new RPM regulation is its effect on the price of books. The proponents of new RPM regulation insist that the new RPM regulation help to stabilize the rapid increase in the average price of a book. The ground for this argument is that because of high discount rate of online firms (the result from price competition among online firms and zero search cost) especially for a book out of RPM regulation (a book which is published more than 18 months ago), the publishing company could increase the list price to offset the discount rate of online firms. If the publishing companies sell their product, for example, at 50% of the list price, they can earn more by increasing the list price of a book.

However, there are some problems with this argument. First, the actual selling price of a book is tied to its list price for 18 months after the initial publication. If a publishing company increase its list price, it could increase its earnings per book; however, the selling quantity can be decreased by the increased list price and the RPM regulation applied to a newly-published book. Accordingly, increasing the list price does not guarantee a higher profit for publishing companies. Second, the important thing for a customer is not a list price but the actual selling price of a book. Even though the new regulation can help to stabilize the list price, it does not mean the stabilization of the actual selling price.

In the short-run, we can guess the new RPM regulation may increase the actual price of a book at online stores especially for a book which is published more than 18 months ago because the book is newly included in the new RPM regulation coverage.[76]

[76] According to Table 2.1, there was no price regulation for a book published more than 18 months ago

According to Cho (2015) that analyzes the change of actual price at online stores, the introduction of the new RPM regulation increases the price of best-selling books by 16.2% and steady-selling books by more than 40% on average.[77]

However, it is not simple to show the effect of new regulation on the actual price of books in the long-run. First, the new RPM regulation can slow down the increasing rate of list price of books, thereby the actual price fixed with the list price can be lower in the long-run compared to the counter-factual situation that the new regulation was not introduced and a book is sold at, for example, more than 20% discounted price from the list price. Second, according to the new regulation, only the publishing companies can change the list price after 18 months from the initial publication.[78] Thus, we need to compare the newly-set-price by a publishing company with the counter-factual-price under no price regulation. In other words, the new RPM regulation can affect both the list price of a new book and the newly-set-price of a book published more than 18 months ago by publishing companies, accordingly makes it hard to evaluate the effect of the new regulation on the overall price level of books. Moreover, I could not obtain information about the actual selling price of books – which were published more than 18 months ago – at online stores before the new regulation took effect, which is essential to analyze the effect of new regulation on the price of non-newly-published books.

Thus, in this paper, I focus on the list price of newly-published books. Because this kind of books are already under the price regulation, the amount of change in list price mean the amount of actual price change. I use the book registration dataset published by the National Library of Korea. This dataset include the information of

before the new RPM regulation took effect. However, the maximum discount rate of those books becomes 15% under the new RPM regulation.

[77] The difference in the increasing rate between best-sellers and steady-sellers comes from the difference in their compositions. Best-sellers are mostly composed of newly-published books while non-newly-published books – which are published more than 18 months ago – consist of steady-sellers. Because there was no price regulation for a non-newly-published book before the new regulation took effect, those books had room for a further price increase.

[78] It is clear that this system is not efficient in that it will increase menu cost and logistical cost. Before the new regulation took effect, an online store can change the price of a book flexibly according to the market situation and their stocks. Because they can quickly change the price of a book by changing the price on the screen, the menu cost was almost zero. However, after the new regulation, the publishing company should print again the cover showing the new price and collect all books in the market and re-distribute the book with a new price. Moreover, online stores could frequently change the actual price to cope with a volatile market situation (e.g., the new publication of similar books and their amount of stocks. However, the publishing company usually have less information about the market, and because the newly-set list price should be fixed for 18 months again, they cannot deal with the unexpected situations by flexibly changing the price.

each book's title, category, author, publishing company, edition, publication year, page, size, the physical characteristics of a book such as (include color picture, partial color and so on), the list price, ISBN, the number of books and supplements information (CD-ROM, DVD, MP3, map, picture and so on) from 2009 to 2017.

Table 2.6 The Average List Price of Books by Category

Category	2009	2010	2011	2012	2013	2014	**2015**	**2016**	**2017**
General	17,514	19,142	18,597	16,833	18,676	19,586	19,435	18,656	20,041
Philosophy	17,320	16,324	17,607	18,277	17,310	17,059	16,715	17,518	17,049
Religion	12,019	12,566	12,840	13,149	13,840	14,289	14,367	14,056	14,632
Social Science	18,750	19,095	20,276	20,300	20,550	21,818	20,878	20,865	20,886
Natural Science	18,523	19,699	20,494	21,796	20,548	21,342	21,345	20,773	20,268
Engineering	23,346	24,716	28,013	26,649	28,032	27,002	26,309	26,517	23,560
Art	19,698	19,108	19,941	19,342	26,655	23,610	24,941	26,395	26,725
Language	14,519	14,146	14,597	15,236	14,588	16,021	16,756	15,445	15,790
Literature	10,142	10,074	10,594	10,765	10,794	11,537	10,847	11,028	10,065
History	17,861	17,982	18,627	18,436	18,499	18,888	18,667	19,761	17,300
Children	6,546	7,389	6,995	7,451	7,579	8,003	8,042	7,846	8,274
Total	13,525	13,662	14,165	14,461	15,072	15,873	15,599	15,766	15,741

Data : National Library of Korea

Table 2.6 shows the average list price of newly-published books which use the book registration data published by the National Library of Korea. Based on the data represented in Table 2.6, this paper analyzes the effect of new RPM regulation on the list price of newly-published books. Table 2.7 shows the various regression result according to category classification of the National Library of Korea. The full table including all the other explanatory variables and fixed effects is represented in Table B4 in the appendix. The column (1) presents the regression result for all books. After the new regulation took effect, the list price decrease by 4% (=11.83%/3) annually after eliminating the effect of time trend. The effects of new RPM regulation are somewhat different depending on the book category, however, in general, the new regulation decreased the list prices.

Table 2.8 shows the results when we consider the case that there was a discontinuity or a jump in the list price when the new regulation took effect and the time trend of the list price is changed after the new regulation. "Time Trend" variable shows the time trend of the list price before the new RPM regulation by each book category. The coefficient of "RPM15" represents how much the list price jumped in 2015 when the new regulation took effect. The coefficient of "RPM Trend" means the amount

Table 2.7 Effect of RPM on List Price

	Dependent Variable = Log(Price per Book)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	ALL	Non-Children	Children	General	Religion	Philosophy	Literature
Time Trend	2.968***	2.575***	3.086***	4.077***	2.951***	1.656***	2.795***
	(0.0546)	(0.0376)	(0.131)	(0.207)	(0.151)	(0.169)	(0.0708)
RPM15	-4.366***	-2.232***	-2.941***	-0.885	-0.925	-1.439	-2.877***
	(0.280)	(0.213)	(0.777)	(1.072)	(0.893)	(0.878)	(0.383)
RPM16	-6.391***	-3.191***	-8.585***	-6.611***	-4.709***	1.628	-2.709***
	(0.382)	(0.245)	(0.948)	(1.147)	(1.075)	(1.092)	(0.453)
RPM17	-11.83***	-8.318***	-5.023***	-8.976***	-3.605***	-1.139	-8.686***
	(0.463)	(0.317)	(1.293)	(1.682)	(1.298)	(1.215)	(0.576)
Obs.	401,266	327,034	74,232	13,382	26,109	12,993	65,382
Adj. R^2	0.619	0.696	0.287	0.670	0.739	0.662	0.709
	(8)	(9)	(10)	(11)	(12)	(13)	
VARIABLES	Social	Natural	Engineering	Language	History	Art	
Time Trend	2.832***	1.228***	1.852***	2.712***	1.878***	1.713***	
	(0.0596)	(0.225)	(0.0990)	(0.148)	(0.158)	(0.206)	
RPM15	-2.755***	2.041	-1.753***	-3.982***	-0.895	-0.603	
	(0.354)	(1.272)	(0.529)	(0.914)	(0.890)	(1.095)	
RPM16	-4.254***	2.173	-3.636***	-8.741***	1.309	4.085***	
	(0.383)	(1.529)	(0.597)	(0.982)	(1.023)	(1.427)	
RPM17	-8.593***	-2.280	-7.002***	-7.865***	-6.772***	-1.367	
	(0.487)	(1.788)	(0.773)	(1.261)	(1.223)	(1.751)	
Obs.	95,740	6,901	53,317	16,895	15,126	18,916	
Adj. R^2	0.703	0.721	0.589	0.696	0.662	0.495	

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

NOTE: All coefficient are reported as β*100. Fixed effects of each book category and various characteristics of a book are included in the all above model. Full table including all the other explanatory variables is presented in **B4** in the appendix.

Table 2.8 Effect of RPM on List Prices (RPM Trend)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	ALL	Non-Children	Children	General	Religion	Philosophy	Literature
(A) Time Trend	2.969***	2.575***	3.086***	4.076***	2.951***	1.656***	2.795***
	(0.0546)	(0.0376)	(0.131)	(0.207)	(0.151)	(0.169)	(0.0708)
(B) RPM15	-3.913***	-1.673***	-4.082***	-1.401	-1.562*	-0.778	-2.029***
	(0.280)	(0.205)	(0.762)	(1.032)	(0.884)	(0.852)	(0.370)
(C) RPM Trend	-3.405***	-2.663***	-2.188***	-4.445***	-1.894***	0.588	-2.376***
	(0.176)	(0.138)	(0.569)	(0.722)	(0.548)	(0.535)	(0.247)
(A)+(C)	-.4368***	-.0877	.8972	-.3690	1.057**	2.244***	.4187*
	(.159)	(.132)	(.551)	(.692)	(.520)	(.507)	(.235)
Obs.	401,266	327,034	74,232	13,382	26,109	12,993	65,382
Adj. R^2	0.619	0.696	0.287	0.670	0.739	0.662	0.709
	(8)	(9)	(10)	(11)	(12)	(13)	
VARIABLES	Social	Natural	Engineering	Language	History	Art	
(A) Time Trend	2.832***	1.229***	1.852***	2.711***	1.877***	1.711***	
	(0.0596)	(0.225)	(0.0990)	(0.148)	(0.158)	(0.206)	
(B) RPM15	-2.370***	2.678**	-1.551***	-4.715***	0.426	0.631	
	(0.337)	(1.229)	(0.509)	(0.871)	(0.855)	(1.072)	
(C) RPM Trend	-2.661***	-1.809**	-2.506***	-2.512***	-1.869***	0.772	
	(0.225)	(0.808)	(0.344)	(0.604)	(0.572)	(0.792)	
(A)+(C)	.1714	-.5796	-.6545**	.1993	.0077	2.483***	
	(.217)	(.776)	(.330)	(.585)	(.549)	(.763)	
Obs.	95,740	6,901	53,317	16,895	15,126	18,916	
Adj. R^2	0.703	0.720	0.589	0.696	0.661	0.494	

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

NOTE: All coefficient are reported as β*100. Fixed effects of each book category and various characteristics of a book are included in the all above model. Full table including all the other explanatory variables is presented in **B5** in the appendix.

of change in time trend after the new RPM regulation, thus the sum of (A) and (C) becomes the new time trend and (C) becomes the difference between new and past time-trend. Overall, the new regulation immediately decreases the list price by 3.9% and diminishes the time trend by 3.4%, thereby it changes the time trend from 2.9% to -0.43%. In most cases, except books included in natural science, history, and art, new regulation decreases the list price immediately and also slows down the increasing time trend except philosophy and art related books. However, the cross-sectional analysis has problems in that there are lots of unobservable omitted variables which were not included in explanatory variables affecting the list price of a book such as an author's reputation, the popularity of the book's topic and advertising cost expenditure of publishing companies and book stores. Thus, the omitted variables and unobservable attributes of a book can cause bias problems in regression estimation.

To overcome this problem, I construct panel dataset by using the same books with different editions from the National Library of Korea database. The different editions of the same books could have different characteristics such as the change of pages, different supplement and so on. To control the possible difference within the different editions, I additionally control the various attributes of a book in panel regression analysis. The full table including all control variables is Table B6 and Table B7 in appendix. According to results in Table 2.9, the main results are similar to cross-sectional results in Table 2.7, however, the amount of decrease in list price is diminished in panel regression results. The estimated time trend is 2.9% in cross section while it is estimated as 2.4% in panel regression. In panel regression, after the new regulation took effect, the list price decrease by 1.36% (=4.14%/3) annually while it is 4% (=11.83%/3) in cross-section analysis. The list price of books included in general, natural science and language category shows more rapid decrease whereas that of literature books increases significantly in 2017.

If we consider the case that there was a discontinuity in list prices of books when the new regulation took effect and it followed different time trend after that, the regression results about the case are presented in Table 2.10. The results are similar to the cross-sectional analysis in Table 2.8, however, the effects of new regulation on list prices are estimated as smaller than the result of the cross-sectional analysis. The introduction of the new RPM regulation immediately decreased the list price by 1.6%

Table 2.9 Effect of RPM on List Prices (Panel)

VARIABLES	(1) ALL	(2) Non-Children	(3) Children	(4) General	(5) Religion	(6) Philosophy	(7) Literature
Time Trend	2.396***	2.393***	2.913***	4.773***	2.030	4.106***	-0.408
	(0.116)	(0.117)	(1.059)	(1.183)	(2.659)	(1.446)	(4.293)
RPM15	-1.740***	-1.813***	4.797	-3.283	-1.921	-8.348	2.91
	(0.373)	(0.374)	(4.055)	(2.745)	(16.02)	(5.608)	(2.55)
RPM16	-2.636***	-2.760***	2.223	-9.437*	8.410	-5.589	2.32
	(0.427)	(0.430)	(3.796)	(5.154)	(20.34)	(7.483)	(2.84)
RPM17	-4.144***	-4.198***	-6.261	-13.20**	-9.138		4.82***
	(0.588)	(0.591)	(5.002)	(5.454)	(13.08)		(1.22)
Obs.	10,855	10,764	91	118	104	91	43
Adj. R^2	0.588	0.586	0.845	0.801	0.712	0.890	0.722
# of IDs	4,334	4,296	39	63	54	54	23
VARIABLES	(8) Social	(9) Natural	(10) Engineering	(11) Language	(12) History	(13) Art	
Time Trend	2.265***	3.614***	2.006***	3.832**	3.345***	1.355	
	(0.137)	(0.577)	(0.176)	(1.569)	(1.064)	(1.311)	
RPM15	-1.569***	-3.502*	-1.662***	-7.109***	1.086	-1.690	
	(0.481)	(2.029)	(0.457)	(2.580)	(3.840)	(3.306)	
RPM16	-2.161***	-5.170*	-2.585***	-11.64**	5.058	-4.772	
	(0.492)	(2.872)	(0.610)	(4.595)	(4.321)	(4.109)	
RPM17	-3.318***	-9.659***	-3.963***	-15.34**	-3.176	-3.773	
	(0.726)	(2.973)	(0.819)	(6.966)	(4.102)	(6.369)	
Obs.	6,173	144	3,512	232	232	70	
Adj. R^2	0.568	0.779	0.683	0.751	0.576	0.632	
# of IDs	2,456	57	1,411	114	109	31	

Standard errors clustered at IDs are parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

NOTE: All coefficient are reported as β*100. Fixed effects of each book category and various characteristics of a book are included in the all above model. Full table including all the other explanatory variables is presented in **B6** in the appendix.

Table 2.10 Effect of RPM on List Prices (Panel, RPM Trend)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	ALL	Non-Children	Children	General	Religion	Philosophy	Literature
(A) Time Trend	2.395***	2.392***	2.866***	4.759***	2.229	4.106***	-0.960
	(0.116)	(0.117)	(1.045)	(1.182)	(2.597)	(1.446)	(3.568)
(B) RPM	-1.666***	-1.755***	5.235	-3.426	-3.617	-8.348	21.12
	(0.353)	(0.354)	(4.029)	(2.700)	(15.39)	(5.608)	(24.60)
(C) RPM Trend	-1.141***	-1.144***	-4.268*	-5.261**	4.944	2.759	7.227
	(0.229)	(0.230)	(2.476)	(2.441)	(6.197)	(4.188)	(6.256)
(A)+(C)	1.254***	1.248***	-1.402	-.5018	7.1728	6.864*	6.267
	(.201)	(.202)	(2.076)	(1.954)	(5.796)	(4.143)	(8.619)
Obs.	10,855	10,764	91	118	104	91	43
Adj. R^2	0.588	0.586	0.843	0.800	0.707	0.890	0.719
# of IDs	4,334	4,296	39	63	54	54	23

	(8)	(9)	(10)	(11)	(12)	(13)
VARIABLES	Social	Natural	Engineering	Language	History	Art
(A) Time Trend	2.264***	3.640***	2.006***	3.830**	3.409***	1.349
	(0.137)	(0.579)	(0.176)	(1.564)	(1.041)	(1.309)
(B) RPM	-1.501***	-3.177	-1.603***	-7.175***	2.049	-2.361
	(0.453)	(2.005)	(0.442)	(2.575)	(3.650)	(3.410)
(C) RPM Trend	-0.822***	-3.078***	-1.105***	-4.265	-0.728	-1.803
	(0.298)	(1.109)	(0.325)	(2.746)	(2.801)	(2.085)
(A)+(C)	1.442***	.5619	.9003***	-.4347	2.681	-.4541
	(.276)	(.914)	(.278)	(1.733)	(2.414)	(2.596)
Obs.	6,173	144	3,512	232	232	70
Adj. R^2	0.568	0.777	0.683	0.751	0.566	0.630
# of IDs	2,456	57	1,411	114	109	31

Standard errors clustered at IDs are parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

NOTE: All coefficient are reported as β*100. Fixed effects of each book category and various characteristics of a book are included in the all above model. Full table including all the other explanatory variables is presented in **B7** in the appendix.

(in cross section, – 3.9%) and decrease the time trend by 1.14% (in cross section, -3.4%). Thus, the time trend after the new regulation is changed to 1.25% from 2.4% (in cross section, from 2.9% to -0.44%). If we consider the change of regulation for a newly-published book (the maximum discount limit is changed to 15% from 19%, thus 4% increase in actual price at the maximum discount), in short run, it is hard to say that the new regulation decreases the actual price of a new book. According to the panel regression results, it is estimated that the new regulation can offset the 4% decreased discount limit after 3 years, that is from 2018.[79]

2.3 Effect of RPM on Firms

2.3.1 Online Firm

One of the main aims of reinforced RPM regulation is to support small-scale offline bookstores and publishing companies in the context of expanding market dominance of large online firms. In order to assess the effectiveness of the reinforced RPM regulation, this section analyzes the impact of the RPM on the profit ratio and sales of top 6 online firms by using firm-level data. I analyze the profit and loss account of the 6 largest online book store companies from 2009 to 2016.[80] Figure 2.2 shows the total amount of sales, costs, and profit of the top 6 online firms. According to the graph, the amount of profit had stayed almost the same from 2009 to 2014, however, it increased in both 2015 and 2016 in the result of enhanced price margin, that is the increased gap between sales and costs. Table 2.3 shows the top 6 firms' profit rapidly increases by 16% and 12% in 2015 and 2016 compared to the previous year and their market share reaches 65% in 2016 while it was 54% in 2014.[81]

Profit Ratio

First, to show the change of price margin of top 6 online firms, I use the profit ratio (=

[79] 1.666%(immediate jump) + 2*1.141%(time trend) = 3.948% compared to the counter-factual situation that there is no change in RPM regulation and also assume that there is no change in time trend.

[80] Because the new RPM regulation took effect from Nov. 2014, I consider the year 2015 and 2016 as the period under the new regulation.

[81] The top 7 (in terms of sales) firms in the book market is 1) KYOBO, 2) YES24, 3) ALADIN, 4) INTERPARK, 5) YOUNGPUNG, 6) BANDI & LUNIS, and 7) LIBRO. However, sales, cost, and profit in the figure is sum of six firms except INTERPARK. Because INTERPARK is operating 4 business parts and they have only published the sales of each business. Thus, I could not figure out the cost and the profit of INTERPARK book business part.

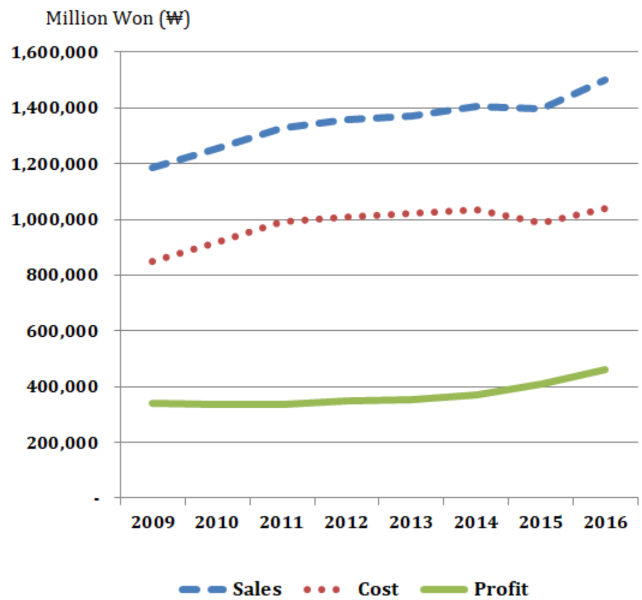


Figure 2.2 Top 6 Online Firm's Sales, Cost and Profit

1 – the rate of cost to selling price) variable[82] which can capture the pure effect of new RPM regulation on firm's profit structure. we can think of using operating surplus as a proxy value for profit condition of a firm under the new regulation. However, it is affected by lots of other factors such as the change of wage or cost for other input for operating the firm (e.g.: rent, investment, and other management costs.).[83] Figure 2.3 shows the changes in profit ratio by each firm.[84] From Figure 2.3 we can see that the profit ratios of all firms were enhanced after the new regulation took effect.

This paper also analyzes the effect of new RPM regulation on top 6 online firms' profit ratio by using regressions and Table 2.11 shows the results. The column (1) and (2) are the results from pooled OLS, and the columns from (3) to (6) are results from panel regression models. In column (3) and (4), I use the fixed-effect model, and time-varying unobservable model is adopted in column (5) and (6). In all three models (column (1), (3) and (5)), the profit ratio is significantly increased by 5.2∼5.5%.

[82] For example, if an online firm sells a book at 100 and buy the book from a publishing company at 70, then the profit ratio is 0.3.

[83] Actually, top 6 online firms have been expanding their business aggressively after 2015 based on the increased earnings. Refer to the footnote 70.

[84] INTERPARK, the 4th biggest online book store have many business areas besides the online book store. The total amount of sales is published by each business sector, however, the cost of sales is not published by each sector. Thus, INTERPARK is not included in the analysis of the profit ratio.

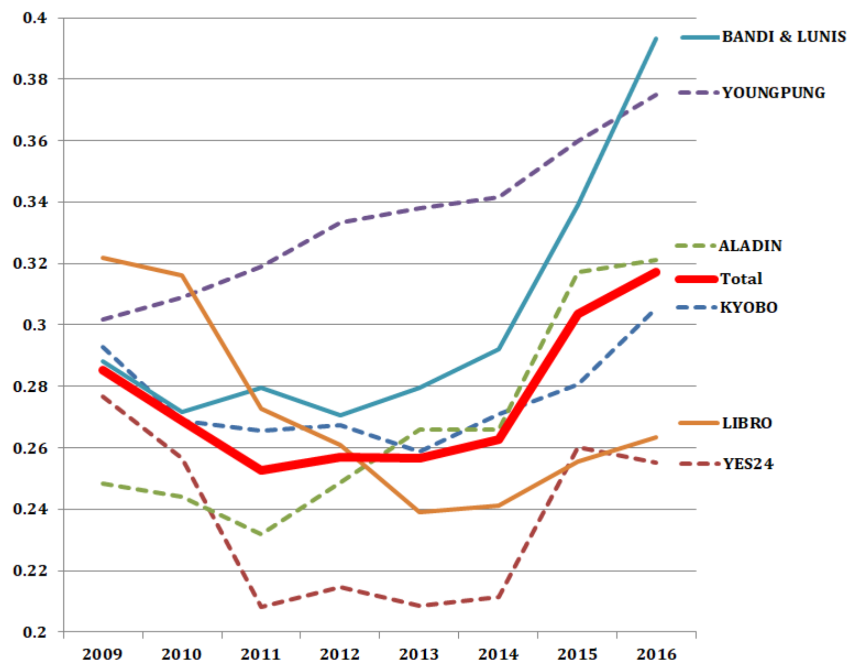


Figure 2.3 The Profit Ratio of online Firms

According to model (2), (4) and (6) showing the effect separately by each year in 2015 and 2016, the profit ratio is increased more in 2016 compared to 2015. Time trend shows negative values in all models; however, it is not significant. All models also control the sales amount to consider a plausible size effect of profit ratio; however, it is not significant in all models. Another interesting thing is the coefficient of "Only Online" which is -4.4% (OLS) and -5.16% (Time-varying). It means that if a firm only operates the online store without an offline store, its profit ratio is lower roughly by 5% than a company operating both online and offline stores. This results can be caused by the fact that the price margin of the offline store is larger than the online store. Another possible reason for this is that top online firms have been opening new offline stores aggressively after the new RPM regulation took effect.[85] Thus, it is highly possible that the big offline stores operated by top online firms absorb a large portion of converting demand from online to offline. Consequently, online firms simultaneously operating an offline store could have obtained two gains: enhanced price margin at the online store and more customer at their offline stores.

[85] Refer to the footnote 70.

Table 2.11 Effect of RPM on Online Firm's Profit Ratio

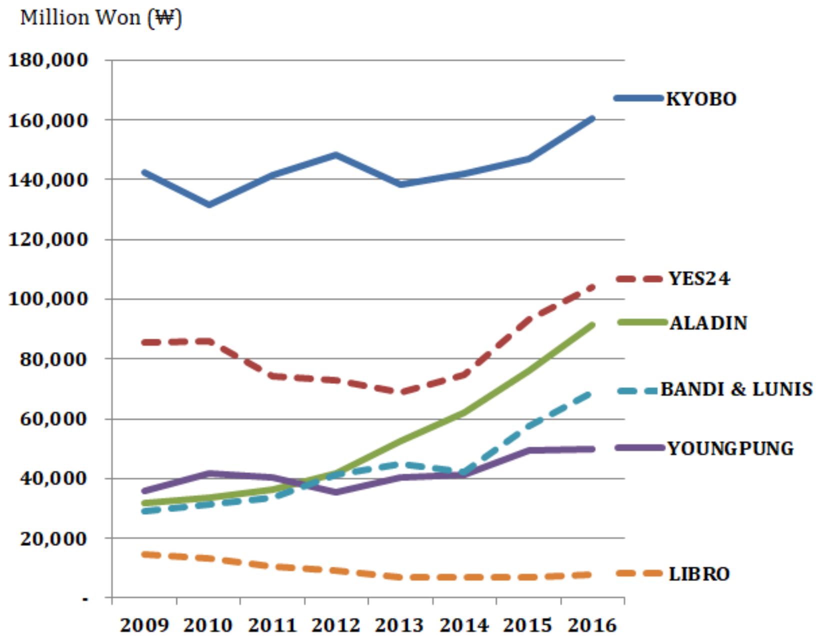
Dependent, Profit Ratio = (Sales - Cost) / Sales						
VARIABLES	OLS		Fixed Effect		Time-varying	
	(1)	(2)	(3)	(4)	(5)	(6)
RPM	**5.243****		**5.553****		**5.240*****	
	(2.615)		(2.616)		(1.931)	
RPM15		**4.526**		**4.976***		**4.348***
		(2.886)		(2.905)		(2.384)
RPM16		**6.370****		**6.431****		**6.206****
		(3.219)		(3.217)		(2.468)
Time Trend	-.1028	-.1537	-.425	-.4538		
	(.495)	(.506)	(.548)	(.557)		
Sales	-.000	-.000	.000	.000	.000	-.000
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Only Online	-4.461**	-4.455**			-5.169*	-5.166*
	(2.112)	(2.127)			(2.703)	(2.763)
γ					.1381	.1502
					(.166)	(.168)
Constant	27.481***	27.667***	18.554***	18.997***	22.872***	22.539***
	(2.113)	(2.150)	(6.436)	(6.566)	(4.681)	(4.746)
Obs.	48	48	48	48	42	42
Adj. R^2	0.2353	0.2420	0.1999	0.2047	0.2846	0.2928
*** p < 0.01, ** p < 0.05, * p < 0.1						

Note: Robust standard errors (OLS) and standard errors clustered at IDs (FE and Time-varying) are parentheses. "RPM" is dummy variable which is equal to one if the year of data is 2015 or 2016. "RPM15" and "RPM16" are also dummy variables representing 2015 and 2016 respectively. "Only Online" is dummy variable which is one if the firm only operates at online without offline stores. γ is the coefficient of lagged profit ratio in the Time-varying Unobservable model.

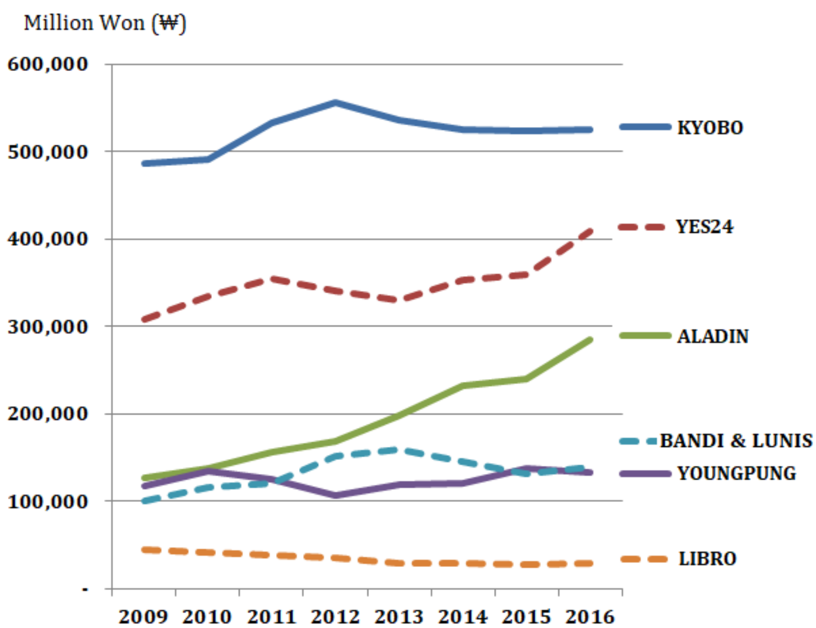
Profits and Sales of the Top 6 Online Firms

The Figure 2.4a shows the amount of profit by each online firm. In general, the amount of profit rapidly increase after the new regulation took effect in 2015 and 2016. Because of plausible zero-sum competition among online firms, the picture becomes more clear if we see the Figure 2.2. Until 2014, the total amount of profit had stayed at almost same level, however, it started increase after the new regulation took effect. The upper panel of Table 2.12 shows the effect of new regulation on the top 6 online firms' profit by using various regression models. If we focus on the result of the Time-varying Unobservable model in the column (3), the new regulation increases the online firm's profit by 24% after controlling the firms' sales. However, the effect is not significant in OLS and fixed effect model in the column (1) and (2).

Figure 2.4b shows the changes in sales by each online firm. After the new



(a) Profit



(b) Sales

Figure 2.4 Profits and Sales of Top 6 Online Firm

Table 2.12 Effect of the New RPM on Online Firms

Dependent = Log(Profit)			
VARIABLES	(1) OLS	(2) Fixed Effect	(3) Time-varying
RPM	**.206**	**.238**	**.241***
	(.210)	(.203)	(.145)
Time Trend	0.001	-0.023	
	(.039)	(.039)	
Log(Sales)	1.042***	1.955***	1.048***
	(.073)	(.380)	(.077)
Only Online	-.168		-.186
	(.174)		(.186)
γ			-.0548
			(.163)
Constant	-1.864**	-12.74***	-2.053*
	(.871)	(4.517)	(1.005)
Obs.	48	48	42
Adj. R^2	0.8399	0.4498	0.8329

Dependent = Log(Sales)			
VARIABLES	(4) OLS	(5) Fixed Effect	(6) Time-varying
RPM	**-.037**	**-.037**	**-.059**
	(.434)	(.085)	(.052)
Time Trend	.0260	.026	
	(.082)	(.015)	
Only Online	.928***		.087
	(.330)		(2.23)
γ			.999***
			(.023)
Constant	11.72***	11.87***	-0.252
	(.325)	(.062)	(.271)
Obs.	48	48	42
Adj. R^2	0.1548	0.0933	0.9842

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: Robust standard errors (OLS) and standard errors clustered at IDs (Fixed effect and Time-varying) are parentheses. "RPM" is dummy variable which is equal to one if the year of data is 2015 or 2016. "Only Online" is dummy variable which is one if the firm only operates at online without offline stores. γ is the coefficient of lagged profit ratio in the Time-varying Unobservable model.

regulation took effect, the sales of KYOBO, YOUNGPUNG, and LIBRO have stayed almost the same level while those of YES24 and ALADIN increase and BANDI & LUNIS' sales have decreased. The growth of YES24 and ALADIN is remarkable in that both companies are leading online stores in used book market.[86] The lower panel of Table 2.12 shows the effect of new regulation on online firm's sales. The effect of new regulation on the firm's sales shows negative values (3.7% in the column (4) and (5), and 5.9% in the Time-varying Unobservable model in the column (6)); however, it is not significant.

If we sum up the above results, we can conclude that online firms' sales are not significantly affected by the RPM regulation. However, both top 6 online firms' profits and profit ratios increase significantly by the enhanced price margin. One might expect that the prices of books are increased by the new RPM regulation. However, the cost of buying books for online firms are not increased relatively. As a result, ironically, the market share of top online firms even increase rapidly after the new regulation took effect even though the new regulation intended the opposite results.

2.3.2 Publishing Company

In this section, this paper analyzes the effect of reinforced RPM regulation on business circumstances of publishing companies. Figure 2.5 shows the Publishing Industry Production Index published by Statistics Korea in the Survey of Service Industry. This statistics is compiled by using publishing companies' sales data. This paper also uses firm-level data including more than 70 firms from KIS database to show the effect of new RPM regulation on the publishing industry in the following regression analysis. However, the firms in the KIS database are included in the biggest firms among all publishing companies. A company is required to be externally audited only if the firm's sale is above a certain level. Because the KIS database is generated from companies which take external audit, the dataset does not include smaller publishing companies. On the other hand, the Publishing Industry Production Index uses data of all publishing companies including various size of companies which is not opened to public. Thus,

[86] ALADIN has opened 17 more used book offline stores (total number of used book stores becomes 37 in 2016) after the new RPM regulation. YES24 also opened the first offline used book store last April. 2016 and have opened four more offline used book stores. According to a news report, YES24's revenue from used book part has been growing on average 30% for each month. http://news.heraldcorp.com/view.php?ud=20160425000116

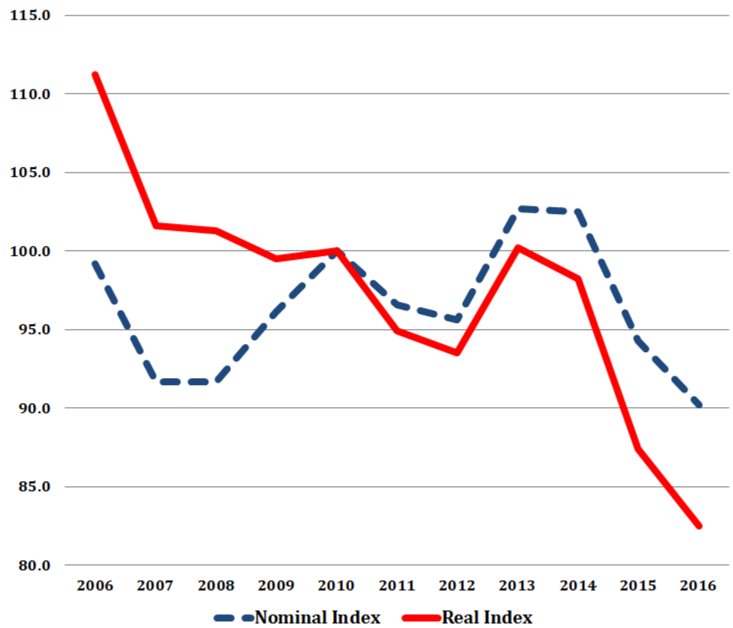


Figure 2.5 The Publishing Industry Production Index

the index shows the overall situation of sales in the publishing industry. According to the nominal index, it tends to increase until 2014. However, it decreases rapidly in 2015 and 2016. The nominal index changed from 100 to 102.5 during 2010∼2014; however, it decreases 8% in 2015 and again 4.3% in 2016. If we check the real index eliminating the effect of an inflation, it shows a downward trend but rebounded in 2013 and have been decreasing more rapid speed in 2015 and 2016. The real index decreases 10% in 2015 and again 5.6% in 2016.

That is, after the new regulation took effect, sales of publishing companies did not increase. Rather, the Publishing Industry Production Index shows the opposite result. Now, this paper investigates each firm-level data from KIS database including more than 70 publishing companies.[87]

Sales, Profit Ratio and Profit of Publishing Companies

Table 2.13 shows the panel regression results. I use time-varying unobservable models for various categories.[88] According to the result in the column (1) including all

[87] The number of companies required to get outside audit differs each year. Some companies disappeared from the sample, and some companies newly appear on the list. Thus, the dataset is the unbalanced panel.

[88] "Study" category includes companies that mainly publish textbooks and study related books such

Table 2.13 Effect of RPM on Publishing Firm's Sales

	Dependent = Log(Sales)				
VARIABLE	(1)	(2)	(3)	(4)	(5)
	Time-varying Unobservable Model				
RPM15	**-.086***	**-.0343**	**-.1712***	**-.1076**	**.0235**
	(.024)	(.043)	(.051)	(.044)	(.040)
RPM16	**-.122***	**-.0934***	**-.2268***	**-.1098***	**.0426**
	(.031)	(.053)	(.068)	(.056)	(.067)
Log(Mean Sales)	1.033***	1.024***	1.056***	.9951***	1.246**
	(.016)	(.025)	(.037)	(.023)	(.494)
γ	.549***	.5379***	.6342***	.4356***	.952***
	(.033)	(.046)	(.104)	(.082)	(.085)
Constant	-.167**	-.119	-.229*	.002***	-.185
	(.076)	(.124)	(.135)	(.128)	(.119)
Obs.	471	198	78	142	61
Category	**ALL**	**Study**	**Children**	**General**	**English**
Adj. R^2	0.9768	0.9764	0.9891	0.9769	0.9937

Standard errors clustered at IDs are parentheses
*** p < 0.01, ** p < 0.05, * p < 0.1

Note: "RPM15" and "RPM16" are dummy variables which are equal to one if the year of data is 2015 and 2016 respectively. γ is the coefficient of lagged profit ratio in the Time-varying Unobservable model.

companies, publishing companies' sales decrease by 8.6% and 12.2% in 2015 and 2016 respectively. To control a plausible size effect on dependent variable,[89] I include the log of the mean sales of a firm. The companies publishing books for children show the largest decrease in their sales among all categories. In 2015, their sales decreased by 17% and more than 22% in 2016. On the other hand, the sales of companies publishing English study books did not decrease after the RPM regulation was strengthened. The results show that, after the new regulation took effect, the top 70 firms' sales are decreased significantly. However, the amount of decrease is somewhat different depending on the category of companies' books. The companies related to publishing study related material ("Study" and "English") comparatively retained their sales. One could expect that the different price elasticity of demand for a book results in different outcomes. In general, a book related to study and exam preparation could have lower price elasticity compared to other books such as literature work; thereby its sale could decrease less or even increase when the price of book increase.

Table 2.14 shows the effect of new regulation on publishing firms' profit ratio.

as books for exam preparation. "Children" includes companies publishing books for children. Companies publishing a single book (such as literature, philosophy, history and so on) are included in the "General" category and "English" includes the companies publishing English study materials.

[89] It is possible that the effect of the new RPM regulation can be different according to a firm's size.

Table 2.14 Effect of RPM on Publishing Firm's Profit Ratio

Dependent = Profit Ratio = (Sales - Cost) / Sales					
VARIABLE	(1)	(2)	(3)	(4)	(5)
	Time-varying Unobservable Model				
RPM15	**1.203**	**1.544**	**1.240**	**1.949**	**1.930**
	(.838)	(1.262)	(1.674)	(1.707)	(1.836)
RPM16	**2.939****	**4.712****	**-2.014**	**3.579**	**5.699****
	(1.262)	(1.897)	(2.479)	(2.517)	(2.699)
Log(Sale)	.5490	-.3474	2.568	1.296	-.4448
	(1.263)	(1.677)	(3.044)	(1.949)	(.4.660)
γ	.9266***	.9273***	.9371***	.7726***	.9186***
	(.015)	(.020)	(.041)	(.053)	(.052)
Constant	2.885**	3.260**	1.311	6.759	5.1036
	(1.171)	(1.834)	(1.973)	(1.949)	(4.808)
Obs.	471	198	78	142	61
Category	**ALL**	**Study**	**Children**	**General**	**English**
Adj. R^2	0.8902	0.9174	0.9073	0.6586	0.8529
Standard errors clustered at IDs are parentheses					
*** p < 0.01, ** p < 0.05, * p < 0.1					

Note: "RPM15" and "RPM16" are dummy variables which are equal to one if the year of data is 2015 and 2016 respectively. γ is the coefficient of lagged profit ratio in the Time-varying Unobservable model.

Overall, after the new regulation was enacted, the profit ratio of publishing companies increase by 1.2% in 2015 (not significant at 10% significance level) and 2.9% (its p-value is 0.02) in 2016. On the other hand, the enhanced profit ratio of "Study" and "English" category is remarkable. There were 4.7% and 5.7% improvement in the profit ratio for each category. Similar to the effect of new RPM regulation on firms' sales, the low price elasticity of these books can be the reason for the enhanced price margin.

Table 2.15 shows the panel regression result about the effects of new regulation on firms' amount of profit. In all columns, the coefficients of (RPM15) and (RPM16) shows positive sign except the children's book publishing companies in 2016. However, the only the coefficient of (RPM15) in column (2) – which shows the effect of new regulation on profits of companies which are included in (Study) category in 2015 – is significant under 10% significance level. Also, the coefficient of (RPM16) in the column (5) implies the significant effect of the RPM regulation on the profits of firms which publish English study books.

To sum up, according to the above results analyzing the effect of new regulation on publishing companies' sales, profit ratio, and the amount of profit, the reinforced RPM regulation significant decreased firms' sales, increased in profit ratio and made no

Table 2.15 Effect of RPM on Publishing Firm's Profit

	Dependent = Log(Profit = Sale - Cost)				
VARIABLE	(1)	(2)	(3)	(4)	(5)
	Time-varying Unobservable Model				
RPM15	**.0255**	**.0614***	**.0474**	**.0356**	**.0332**
	(.0313)	(.0367)	(.0476)	(.0515)	(.0322)
RPM16	**.0321**	**.0581**	**-.0600**	**.0478**	**.0945****
	(.0443)	(.0554)	(.0703)	(.0624)	(.0468)
Log(Sale)	.9630***	.9133***	1.033***	1.094***	.9805***
	(.0346)	(.0514)	(.0847)	(.0212)	(.0721)
γ	.7994***	.9383***	.9317***	.2701***	.8960***
	(.0211)	(.0240)	(.0452)	(.0430)	(.0592)
Constant	-.0921**	.0031	-.0791	-1.291***	-.0363
	(.0737)	(.0373)	(.0943)	(.1765)	(.0843)
Obs.	471	198	78	142	61
Category	**ALL**	**Study**	**Children**	**General**	**Eng**
Adj. R2	0.9681	0.9834	0.9930	0.9745	0.9950

Standard errors clustered at IDs are parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note: "RPM15" and "RPM16" are dummy variables which are equal to one if the year of data is 2015 and 2016 respectively. γ is the coefficient of lagged profit ratio in the Time-varying Unobservable model.

significant changes in the profit of publishing companies. That is, given the decreasing sales, publishing companies kept their profits by enhancing the profit ratio. Enhanced profit ratio could be attained by some changes in firms' profit structure and strategy. First, if publishing companies sell their books to a retail book store at a higher price than before, it could increase the profit ratio of publishing companies. Increased retail price owing to the reinforced RPM regulation can raise the wholesale price of a book, and this is the effect that the government intended from the introduction of the new regulation. However, second, it is also possible that publishing companies decrease their costs against the decrease in sales by reducing the variety of publication and concentrating on publishing some popular books. The diminished variety of new publication as we can see in (Table 2.4) implies that the enhanced profit ratio could be the result of the second reason.

2.3.3 Offline Book Stores

As I discuss in the introduction part, one of the important purposes of reinforced RPM regulation is to support small offline bookstores. In this section, this paper introduces some survey results to investigate the effect of the new RPM regulation on small offline bookstores, because there is no available firm-level data in case of the small-scale offline bookstores.

Table 2.16 Survey Result of Offline Stores

	Year	Response	Increase	Almost Same	Decrease
Total					(%)
	2014	839	4.1	37.1	58.8
	2015	813	8.7	42.3	49
	2016	774	5.0	28.8	66.2
Less Than 2					
	2014	604	3.0	34.4	62.6
	2015	552	8.2	36.6	55.3
	2016	665	5.0	28.4	66.6
3∼4 Employees					
	2014	95	4.2	48.4	47.4
	2015	152	9.2	44.7	46.1
	2016	113	2.7	33.6	63.7
5∼9 Employees					
	2014	59	10.2	44.1	45.8
	2015	74	9.5	68.9	21.6
	2016	46	6.5	28.3	65.2
More Than 10					
	2014	16	25.0	43.8	31.3
	2015	35	14.3	65.7	20.0
	2016	15	20.0	13.3	66.7

DATA: KPIPA (2017)

Table 2.16 shows the results of the Publishing Industry Survey by Publication Industry Promotion Agency of Korea (KPIPA). The survey asked offline bookstores whether sales increased, decreased, or remained unchanged compared to the previous year. The table presents the number of respondents and the portion of each response into three categories by the number of employees. According to the results, 66.2% of offline book store responded that sales decreased compared to the previous year in 2016. The proportion of respondents who reported a decrease in sales overall declined temporarily in 2015; however, increased again in 2016. The results by the number of employees show a similar pattern; however, when the number of respondents is small, the volatility of the result becomes large.

It is necessary to note that analyses based on the survey data are highly limited because the survey is not panel data and also there is no quantitive information such as the number of changes in sales of offline stores. However, according to the results of the survey, it is hard to insist that the introduction of new RPM regulation helps to support the business of small offline book stores.

2.4 Conclusion

The RPM regulation on Korean book market was enacted and reinforced to protect a variety of culture and knowledge by helping unpopular books to survive in the market and also to support small publishing companies and offline book stores under the situation that some large online bookstores have been dominating the book market. Based on the purposes of the RPM regulation, this paper investigates and evaluates whether the regulation achieves its goal.

Various empirical analyses imply that the RPM regulation causes various adverse effects opposite to the original intention of the regulation. First of all, the market share of the seven online firms, which stay around at 54%, increased to 65% and the total sales of all the remaining bookstores decreased by 34 and consequently, the degree of market concentration became worse. Second, this paper evaluates the effect of RPM regulation on the number of new book publications as a proxy for the variety of culture and knowledge by using the data of the National Library of Korea. According to various panel regression results, the number of newly-published books decreased by 6% (in 2015) and 8% more (in 2016) after the regulation. Third, this paper also analyzes the effects of RPM regulation on the management conditions of large online firms, publishing companies, and small offline bookstores, respectively. Both profit ratios and profits of large online firms significantly enhanced after the regulation was reinforced. On the other hand, the Publishing Industry Production Index based on the sales of publishing companies decreased 16% after the new regulation. Also, regression results using the data of the top 70 publishing companies from KIS database show similar results. Overall, sales of the publishing companies decrease 8% in 2015 and 12% in 2016. As there is no data available for small offline bookstores, this paper refers to the KPIPA's survey results to examine the impact of the new regulation on offline bookstores. From the survey results, we might estimate that the introduction of the new RPM regulation has not helped to improve the management condition of small bookstores.

Also, the reinforced RPM regulation decreased the list price of the newly-published book by 4.1% for three years based on the panel regression results in Table

2.9. Nonetheless, we could not insist that the new RPM regulation decreases the retail price of a book because the regulation diminishes the discount rate. When we consider the significant discount rate of online book stores for a non-newly-published book which has no price regulation before the new RPM regulation took effect, 4.1% (from the panel regression in Table 2.9) or 11.8% (from the cross-sectional result in Table 2.7) decrease for three years might be insufficient to offset the diminished discount rate by the new RPM regulations.

In conclusion, the results of various empirical studies show that new RPM regulation is not accomplishing the purpose of its introduction. I suggest that the government of Korea should review other methods such as the direct subsidies for small publishing companies and offline book stores instead of price regulation such as the resale price maintenance.

APPENDIX

APPENDIX B

Table B1 The Number of Newly-Published Books

Category	2008	2009	2010	2011	2012	2013	2014	**2015**
General	822	805	703	715	613	822	1,078	930
Philosophy	946	915	1,055	1,152	1,237	1,335	1,367	1,303
Religion	2,009	2,177	1,899	1,925	1,889	1,899	2,078	1,668
Social Science	6,335	6,483	6,017	5,919	6,089	7,097	8,015	7,561
Science	593	542	541	647	521	645	718	665
Engineering	2,754	3,054	3,206	3,628	3,552	3,880	4,417	4,508
Art	1,451	1,407	1,382	1,354	1,329	1,402	1,605	1,729
Language	1,823	1,660	1,625	1,385	1,192	1,399	1,408	1,392
Literature	8,482	8,718	8,192	8,184	7,963	9,296	10,671	10,899
History	1,139	1,008	1,031	989	1,083	1,283	1,291	1,403
Study	1,787	1,803	2,512	2,159	1,379	1,356	1,462	1,393
Children	8,417	7,884	7,352	9,546	7,495	7,424	7,269	5,572
Sub-Total	36,558	36,456	35,515	37,603	34,342	37,838	41,379	39,023
Comic	6,541	5,735	4,776	6,433	5,425	5,308	6,210	6,190
Total	43,099	42,191	40,291	44,036	39,767	43,146	47,589	45,213

Data : Korean Publisheres Association

Table B2 The Average Price of Newly-Published Books

Category	2008	2009	2010	2011	2012	2013	2014	**2015**	**2016**
General	18,782	19,927	18,456	18,932	19,999	20,365	24,050	20,046	21,170
Philosophy	16,835	18,231	16,061	16,883	17,012	17,405	17,796	17,628	18,581
Religion	13,292	12,677	13,497	13,778	14,158	13,952	14,762	14,929	15,486
Social Science	17,587	18,795	19,108	19,855	19,821	21,955	22,347	20,574	22,453
Science	22,960	19,162	19,963	20,916	21,569	21,679	19,771	21,625	22,879
Engineering	21,142	21,776	21,459	21,647	22,585	24,579	24,400	23,169	25,586
Art	18,465	23,731	18,600	19,928	19,438	21,262	28,590	27,332	29,684
Language	15,668	15,319	14,263	16,710	18,704	16,636	17,554	18,081	16,644
Literature	9,845	10,227	10,352	10,887	11,297	11,485	13,229	11,852	11,805
History	19,963	19,088	18,733	19,666	19,760	20,398	19,713	22,043	20,704
Study	10,373	10,328	9,365	10,314	11,113	11,875	11,845	10,889	11,495
Children	8,536	8,992	9,427	9,813	10,617	9,932	10,527	9,955	10,545
Sub-Total	13,494	14,148	13,965	14,459	15,333	16,055	17,232	16,505	18,607
Comic	4,413	4,441	4,309	4,541	4,720	4,865	4,959	4,998	5,186
Total	12,116	12,829	12,820	13,010	13,885	14,678	15,631	14,929	17,356

Data : Korean Publisheres Association

Table B3 The Number of Circulation

(Unit = 1000)

Category	2008	2009	2010	2011	2012	2013	2014	**2015**
General	1,616	1,514	1,405	1,336	1,190	1,448	1,574	1,675
Philosophy	1,729	1,488	1,980	2,153	2,162	2,240	1,892	2,036
Religion	4,393	4,368	3,790	3,997	3,328	3,384	3,496	2,986
Social Science	10,853	10,937	10,765	9,364	9,774	9,618	9,651	9,490
Science	761	907	727	1,113	675	783	796	798
Engineering	3,641	3,902	4,397	4,977	4,634	4,872	4,986	4,798
Art	2,265	2,202	2,116	2,151	2,007	1,936	1,941	2,415
Language	4,048	3,591	4,338	2,720	1,871	2,185	2,000	2,062
Literature	17,641	18,644	17,280	15,837	14,796	15,945	15,176	15,610
History	2,151	1,826	1,829	1,816	1,866	2,064	2,029	2,197
Study	13,621	14,297	22,007	17,217	10,547	10,630	16,713	16,524
Children	26,885	29,275	26,200	37,705	26,537	24,863	26,167	16,837
Sub-Total	89,605	92,951	96,833	100,387	79,388	79,968	86,420	77,429
Comic	16,911	13,263	9,477	9,163	7,518	6,546	7,746	7,589
Total	106,516	106,215	106,310	109,550	86,907	86,513	94,166	85,018

Data : Korean Publisheres Association

Table B4 Effect of RPM on List Price

VARIABLE	(1) ALL	(2) Non-Child	(3) Child	(4) General	(5) Religion	(6) Philosophy	(7) Literature
Time Trend	2.968***	2.575***	3.086***	4.077***	2.951***	1.656***	2.795***
	(0.0546)	(0.0376)	(0.131)	(0.207)	(0.151)	(0.169)	(0.0708)
RPM15	-4.366***	-2.232***	-2.941***	-0.885	-0.925	-1.439	-2.877***
	(0.280)	(0.213)	(0.777)	(1.072)	(0.893)	(0.878)	(0.383)
RPM16	-6.391***	-3.191***	-8.585***	-6.611***	-4.709***	1.628	-2.709***
	(0.382)	(0.245)	(0.948)	(1.147)	(1.075)	(1.092)	(0.453)
RPM17	-11.83***	-8.318***	-5.023***	-8.976***	-3.605***	-1.139	-8.686***
	(0.463)	(0.317)	(1.293)	(1.682)	(1.298)	(1.215)	(0.576)
log(page)	42.14***	53.15***	22.26***	57.14***	72.95***	56.55***	39.37***
	(0.211)	(0.180)	(0.363)	(0.701)	(0.836)	(0.905)	(0.337)
size	1.346*	6.145***	0.242	4.716***	6.519***	6.260***	9.097***
	(0.700)	(0.0353)	(0.158)	(0.145)	(0.188)	(0.147)	(0.151)
color picture	14.48***	14.88***	-5.250***	0.897	22.13***	10.82***	16.17***
	(0.251)	(0.214)	(0.820)	(0.950)	(1.330)	(0.902)	(0.330)
partial color	9.297***	8.162***	9.185***	-7.935***	13.35***	9.718***	-3.086***
	(0.489)	(0.348)	(1.831)	(1.414)	(1.762)	(1.840)	(0.660)
picture	2.328***	1.018***	-6.487***	3.772***	-0.322	1.334**	3.805***
	(0.174)	(0.140)	(0.871)	(0.792)	(0.580)	(0.540)	(0.302)
graph	1.726***	-2.829***	11.49***	-0.273	-0.0563	-5.237***	2.291
	(0.510)	(0.153)	(1.026)	(0.658)	(1.394)	(0.736)	(1.410)
etc	-2.611***	-0.0144	-4.239***	-3.542***	-1.640***	-1.561**	2.906***
	(0.356)	(0.165)	(0.467)	(0.819)	(0.617)	(0.696)	(0.390)
number of books	-28.39***	-67.26***	-15.75***	-65.18***	-57.85***	-61.62***	-74.63***
	(2.461)	(1.439)	(1.765)	(0.872)	(9.942)	(1.799)	(0.551)
DVD, CD, MP3	12.12***	1.267***	-2.933	0.764	7.200**	14.75***	10.53***
	(0.696)	(0.428)	(2.005)	(0.971)	(3.118)	(3.449)	(2.884)
# of Appendix			23.29***				
			(1.170)				
Constant	-5,246***	-4,590***	-5,405***	-7,609***	-5,495***	-2,770***	-5,026***
	(105.2)	(75.87)	(265.6)	(416.1)	(309.8)	(341.9)	(141.9)
Obs.	401,266	327,034	74,232	13,382	26,109	12,993	65,382
Adj. R^2	0.619	0.696	0.287	0.670	0.739	0.662	0.709

VARIABLE	(8) Social	(9) Natural	(10) Tech	(11) Language	(12) History	(13) Art
Time Trend	2.832***	1.228***	1.852***	2.712***	1.878***	1.713***
	(0.0596)	(0.225)	(0.0990)	(0.148)	(0.158)	(0.206)
RPM15	-2.755***	2.041	-1.753***	-3.982***	-0.895	-0.603
	(0.354)	(1.272)	(0.529)	(0.914)	(0.890)	(1.095)
RPM16	-4.254***	2.173	-3.636***	-8.741***	1.309	4.085***
	(0.383)	(1.529)	(0.597)	(0.982)	(1.023)	(1.427)
RPM17	-8.593***	-2.280	-7.002***	-7.865***	-6.772***	-1.367
	(0.487)	(1.788)	(0.773)	(1.261)	(1.223)	(1.751)
log(page)	60.74***	60.59***	50.23***	55.32***	55.75***	35.46***
	(0.353)	(0.942)	(0.431)	(0.623)	(1.251)	(0.689)
size	5.039***	4.635***	5.250***	4.558***	5.790***	5.008***
	(0.0613)	(0.162)	(0.0777)	(0.0872)	(0.134)	(0.102)
color picture	9.457***	21.91***	21.95***	6.236***	-1.201	25.29***
	(0.518)	(1.138)	(0.488)	(0.703)	(0.789)	(0.905)
partial color	5.681***	14.96***	21.25***	4.513**	3.498***	20.39***
	(0.740)	(1.755)	(0.776)	(1.944)	(1.150)	(1.092)
picture	0.870***	0.996	3.076***	1.161**	-4.493***	2.192**
	(0.201)	(0.945)	(0.425)	(0.519)	(0.723)	(0.894)
graph	-2.993***	-4.830***	1.005***	3.807***	-2.129***	-2.067**
	(0.211)	(0.812)	(0.334)	(0.921)	(0.803)	(0.943)
etc	0.388	0.843	1.147*	4.901***	-1.652***	-4.348***
	(0.262)	(0.791)	(0.618)	(0.903)	(0.380)	(0.620)
# of Books	-61.80***	-66.17***	-68.36***	-66.66***	-63.32***	-74.06***
	(2.258)	(1.326)	(0.856)	(0.496)	(0.774)	(0.981)
DVD, CD, MP3	3.297***	20.49***	5.142***	3.937***	15.41***	1.770
	(1.096)	(5.702)	(0.884)	(0.581)	(3.939)	(1.489)
Constant	-5,134***	-1,890***	-3,100***	-4,852***	-3,192***	-2,719***
	(119.7)	(452.4)	(199.0)	(297.8)	(317.7)	(413.7)
Obs.	95,740	6,901	53,317	16,895	15,126	18,916
Adj. R^2	0.703	0.721	0.589	0.696	0.662	0.495

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

Table B5 Effect of RPM on List Price (RPM Trend)

VARIABLES	(1) ALL	(2) Non-Child	(3) Child	(4) General	(5) Religion	(6) Philosophy	(7) Literature
Time_Trend	2.969***	2.575***	3.086***	4.076***	2.951***	1.656***	2.795***
	(0.0546)	(0.0376)	(0.131)	(0.207)	(0.151)	(0.169)	(0.0708)
RPM	-3.913***	-1.673***	-4.082***	-1.401	-1.562*	-0.778	-2.029***
	(0.280)	(0.205)	(0.762)	(1.032)	(0.884)	(0.852)	(0.370)
RPM_Trend	-3.405***	-2.663***	-2.188***	-4.445***	-1.894***	0.588	-2.376***
	(0.176)	(0.138)	(0.569)	(0.722)	(0.548)	(0.535)	(0.247)
log(page)	42.13***	53.15***	22.28***	57.12***	72.96***	56.56***	39.34***
	(0.211)	(0.180)	(0.363)	(0.700)	(0.837)	(0.905)	(0.338)
size	1.345*	6.145***	0.242	4.726***	6.520***	6.252***	9.100***
	(0.700)	(0.0353)	(0.158)	(0.145)	(0.188)	(0.148)	(0.151)
color picture	14.48***	14.87***	-5.286***	0.971	22.13***	10.85***	16.17***
	(0.251)	(0.214)	(0.820)	(0.948)	(1.330)	(0.902)	(0.330)
partial color	9.308***	8.177***	9.232***	-7.905***	13.34***	9.734***	-3.093***
	(0.489)	(0.348)	(1.834)	(1.415)	(1.762)	(1.844)	(0.661)
picture	2.327***	1.016***	-6.504***	3.837***	-0.337	1.342**	3.818***
	(0.174)	(0.140)	(0.871)	(0.791)	(0.579)	(0.540)	(0.302)
graph	1.727***	-2.829***	11.37***	-0.260	-0.0336	-5.234***	2.324*
	(0.510)	(0.153)	(1.023)	(0.658)	(1.393)	(0.737)	(1.408)
etc	-2.604***	-0.00103	-4.228***	-3.574***	-1.641***	-1.555**	2.935***
	(0.356)	(0.165)	(0.468)	(0.821)	(0.617)	(0.698)	(0.391)
# of Books	-28.39***	-67.28***	-15.77***	-65.15***	-57.87***	-61.58***	-74.69***
	(2.461)	(1.439)	(1.766)	(0.871)	(9.944)	(1.795)	(0.551)
DVD, CD, MP3	12.12***	1.263***	-2.989	0.749	7.186**	14.75***	10.47***
	(0.696)	(0.428)	(2.000)	(0.970)	(3.119)	(3.441)	(2.882)
# of Appendix			23.32***				
			(1.166)				
Constant	714.5***	580.2***	792.1***	577.8***	431.4***	556.0***	587.3***
	(16.95)	(2.207)	(5.559)	(4.808)	(12.64)	(6.063)	(3.600)
Obs.	401,266	327,034	74,232	13,382	26,109	12,993	65,382
Adj. R^2	0.619	0.696	0.287	0.670	0.739	0.662	0.709

VARIABLES	(8) Social	(9) Natural	(10) Tech	(11) Language	(12) History	(13) Art
Time_Trend	2.832***	1.229***	1.852***	2.711***	1.877***	1.711***
	(0.0596)	(0.225)	(0.0990)	(0.148)	(0.158)	(0.206)
RPM	-2.370***	2.678**	-1.551***	-4.715***	0.426	0.631
	(0.337)	(1.229)	(0.509)	(0.871)	(0.855)	(1.072)
RPM_Trend	-2.661***	-1.809**	-2.506***	-2.512***	-1.869***	0.772
	(0.225)	(0.808)	(0.344)	(0.604)	(0.572)	(0.792)
log(page)	60.74***	60.58***	50.23***	55.32***	55.75***	35.41***
	(0.353)	(0.941)	(0.431)	(0.623)	(1.251)	(0.693)
size	5.039***	4.634***	5.249***	4.562***	5.793***	5.003***
	(0.0613)	(0.162)	(0.0777)	(0.0872)	(0.134)	(0.102)
color picture	9.456***	21.90***	21.96***	6.233***	-1.198	25.15***
	(0.518)	(1.139)	(0.488)	(0.703)	(0.790)	(0.909)
partial color	5.700***	15.08***	21.27***	4.577**	3.558***	20.32***
	(0.741)	(1.754)	(0.776)	(1.949)	(1.149)	(1.095)
picture	0.869***	0.997	3.083***	1.153**	-4.529***	2.070**
	(0.201)	(0.945)	(0.425)	(0.519)	(0.724)	(0.898)
graph	-2.993***	-4.836***	1.004***	3.790***	-2.116***	-2.116**
	(0.211)	(0.812)	(0.334)	(0.923)	(0.803)	(0.942)
etc	0.387	0.846	1.145*	4.903***	-1.607***	-4.264***
	(0.262)	(0.792)	(0.618)	(0.902)	(0.380)	(0.622)
# of Books	-61.80***	-66.16***	-68.36***	-66.66***	-63.36***	-74.21***
	(2.258)	(1.329)	(0.856)	(0.495)	(0.775)	(0.983)
DVD, CD, MP3	3.282***	20.40***	5.143***	3.920***	15.55***	1.764
	(1.096)	(5.705)	(0.884)	(0.581)	(3.967)	(1.488)
Constant	553.5***	577.2***	618.7***	592.7***	579.1***	721.2***
	(3.229)	(6.557)	(3.079)	(4.327)	(7.248)	(5.620)
Obs.	95,740	6,901	53,317	16,895	15,126	18,916
Adj. R^2	0.703	0.720	0.589	0.696	0.661	0.494

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table B6 Effect of RPM on List Price (Panel)

VARIABLES	(1) ALL	(2) Non-Child	(3) Child	(4) General	(5) Religion	(6) Philosophy	(7) Literature
Time Trend	2.396***	2.393***	2.913***	4.773***	2.030	4.106***	-0.408
	(0.116)	(0.117)	(1.059)	(1.183)	(2.659)	(1.446)	(4.293)
RPM15	-1.740***	-1.813***	4.797	-3.283	-1.921	-8.348	20.91
	(0.373)	(0.374)	(4.055)	(2.745)	(16.02)	(5.608)	(24.55)
RPM16	-2.636***	-2.760***	2.223	-9.437*	8.410	-5.589	21.32
	(0.427)	(0.430)	(3.796)	(5.154)	(20.34)	(7.483)	(28.84)
RPM17	-4.144***	-4.198***	-6.261	-13.20**	-9.138		40.82***
	(0.588)	(0.591)	(5.002)	(5.454)	(13.08)		(12.12)
log(page)	7.008***	7.026***	12.19	17.14	61.31***	45.34***	27.38
	(1.561)	(1.567)	(11.82)	(15.66)	(13.95)	(4.371)	(23.02)
size	3.146***	3.133***	16.16*	0.0672	6.318***	6.308***	12.91***
	(0.736)	(0.736)	(8.006)	(1.684)	(1.444)	(1.074)	(3.555)
color picture	1.730	1.674	-0.685	-8.736	2.487	6.337	20.20
	(1.893)	(1.933)	(6.448)	(9.582)	(6.807)	(6.251)	(12.97)
graph	-0.163	-0.162	2.431**	-4.416	-45.42***	4.913	
	(0.408)	(0.408)	(1.170)	(3.932)	(16.42)	(4.174)	
etc	1.186	1.166	-0.396	14.19*	1.855	0.237	
	(0.997)	(1.023)	(3.619)	(7.590)	(3.534)	(3.973)	
# of Books	-65.03***	-64.81***	-98.12***	-60.93***	-72.14***		
	(1.406)	(1.408)	(13.47)	(15.29)	(10.58)		
# of Appendix			-2.456*				
			(1.240)				
partial color	0.325	0.313		-32.69	95.07***	4.469	-52.71**
	(1.460)	(1.464)		(20.36)	(18.04)	(6.915)	(19.98)
picture	1.263**	1.267**		-3.890	-12.91	-1.200	-18.85
	(0.510)	(0.509)		(4.371)	(14.41)	(2.996)	(18.40)
DVD, CD, MP3	5.920**	5.897**		5.410			
	(2.431)	(2.431)		(9.015)			
Constant	-3,866***	-3,860***	-5,286**	-8,652***	-3,575	-7,712**	1,331
	(232.2)	(233.2)	(2,101)	(2,340)	(5,315)	(2,899)	(8,669)
Obs.	10,855	10,764	91	118	104	91	43
Adj. R^2	0.588	0.586	0.845	0.801	0.712	0.890	0.722
# of IDs	4,334	4,296	39	63	54	54	23

VARIABLES	(1) Social	(2) Natural	(3) Tech	(4) Language	(5) History	(6) Art
Time Trend	2.265***	3.614***	2.006***	3.832**	3.345***	1.355
	(0.137)	(0.577)	(0.176)	(1.569)	(1.064)	(1.311)
RPM15	-1.569***	-3.502*	-1.662***	-7.109***	1.086	-1.690
	(0.481)	(2.029)	(0.457)	(2.580)	(3.840)	(3.306)
RPM16	-2.161***	-5.170*	-2.585***	-11.64**	5.058	-4.772
	(0.492)	(2.872)	(0.610)	(4.595)	(4.321)	(4.109)
RPM17	-3.318***	-9.659***	-3.963***	-15.34**	-3.176	-3.773
	(0.726)	(2.973)	(0.819)	(6.966)	(4.102)	(6.369)
log(page)	8.720***	40.69***	3.427***	54.23***	7.923	96.25*
	(3.223)	(7.962)	(1.220)	(14.03)	(5.458)	(50.20)
size	2.063**	-3.912	1.559**	7.133**	2.660	4.498***
	(0.948)	(2.411)	(0.650)	(3.353)	(1.989)	(0.687)
color picture	-3.018	1.503	3.781*		-8.622	10.08
	(3.233)	(3.958)	(2.165)		(20.61)	(14.33)
partial color	-0.189	-7.672**	2.203	4.118		-6.510
	(1.964)	(3.244)	(1.771)	(3.103)		(9.525)
picture	0.0698	-3.415	2.115***	7.309**	10.81**	-9.275
	(0.550)	(2.919)	(0.746)	(2.950)	(4.917)	(9.595)
graph	-0.204	-0.608	-0.166	3.108	-0.133	-0.304
	(0.625)	(4.516)	(0.488)	(2.786)	(3.730)	(2.210)
etc	-0.110	0.0710	-0.181	-14.47	1.867	0.679
	(1.235)	(3.417)	(1.474)	(10.67)	(3.397)	(5.408)
# of Books	-65.26***		-67.86***	-68.87***	-47.15***	
	(2.267)		(1.526)	(3.488)	(4.974)	
DVD, CD, MP3	4.732*		9.269*	11.70**		
	(2.485)		(5.627)	(5.003)		
Constant	-3,578***	-6,403***	-3,019***	-7,168**	-5,822***	-2,419
	(267.4)	(1,125)	(353.1)	(3,154)	(2,162)	(2,393)
Obs.	6,173	144	3,512	232	232	70
Adj. R^2	0.568	0.779	0.683	0.751	0.576	0.632
# of IDs	2,456	57	1,411	114	109	31

Standard errors clustered at IDs are parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table B7 Effect of RPM on List Price (Panel, RPM Trend)

VARIABLES	(1) ALL	(2) Non-Child	(3) Child	(4) General	(5) Religion	(6) Philosophy	(7) Literature
Time Trend	2.395***	2.392***	2.866***	4.759***	2.229	4.106***	-0.960
	(0.116)	(0.117)	(1.045)	(1.182)	(2.597)	(1.446)	(3.568)
RPM	-1.666***	-1.755***	5.235	-3.426	-3.617	-8.348	21.12
	(0.353)	(0.354)	(4.029)	(2.700)	(15.39)	(5.608)	(24.60)
RPM Trend	-1.141***	-1.144***	-4.268*	-5.261**	4.944	2.759	7.227
	(0.229)	(0.230)	(2.476)	(2.441)	(6.197)	(4.188)	(6.256)
log(page)	7.016***	7.032***	12.58	16.49	60.75***	45.34***	28.13
	(1.562)	(1.567)	(11.44)	(15.36)	(13.88)	(4.371)	(23.61)
size	3.150***	3.136***	16.27**	-0.146	6.345***	6.308***	13.46***
	(0.736)	(0.736)	(7.852)	(1.338)	(1.561)	(1.074)	(2.900)
color picture	1.717	1.661	0.0603	-8.174	4.496	6.337	21.85*
	(1.893)	(1.933)	(6.767)	(9.504)	(7.141)	(6.251)	(10.79)
graph	-0.156	-0.156	2.369**	-4.597	-47.56***	4.913	
	(0.408)	(0.408)	(1.149)	(3.993)	(15.69)	(4.174)	
etc	1.185	1.165	-0.335	14.11*	0.475	0.237	
	(0.998)	(1.024)	(3.528)	(7.497)	(3.354)	(3.973)	
# of Books	-65.04***	-64.81***	-99.77***	-60.94***	-72.31***		
	(1.406)	(1.408)	(12.97)	(15.37)	(10.36)		
# of Appendix			-3.158**				
			(1.226)				
partial color	0.329	0.315		-33.19	95.97***	4.469	-46.87***
	(1.459)	(1.463)		(20.66)	(18.47)	(6.915)	(14.47)
picture	1.259**	1.265**		-3.574	-11.02	-1.200	-19.41
	(0.510)	(0.510)		(4.350)	(13.56)	(2.996)	(18.07)
DVD, CD, MP3	5.907**	5.886**		5.404			
	(2.430)	(2.430)		(9.024)			
Constant	937.1***	937.7***	551.7**	927.6***	497.2***	519.9***	499.3***
	(22.16)	(22.18)	(230.3)	(101.0)	(72.18)	(32.32)	(117.6)
Obs.	10,855	10,764	91	118	104	91	43
Adj. R^2	0.588	0.586	0.843	0.800	0.707	0.890	0.719
# of IDs	4,334	4,296	39	63	54	54	23

VARIABLES	(8) Social	(9) Natural	(10) Tech	(11) Language	(12) History	(13) Art
Time Trend	2.264***	3.640***	2.006***	3.830**	3.409***	1.349
	(0.137)	(0.579)	(0.176)	(1.564)	(1.041)	(1.309)
RPM	-1.501***	-3.177	-1.603***	-7.175***	2.049	-2.361
	(0.453)	(2.005)	(0.442)	(2.575)	(3.650)	(3.410)
RPM Trend	-0.822***	-3.078***	-1.105***	-4.265	-0.728	-1.803
	(0.298)	(1.109)	(0.325)	(2.746)	(2.801)	(2.085)
log(page)	8.723***	40.29***	3.437***	54.26***	7.657	95.95*
	(3.224)	(7.980)	(1.220)	(14.04)	(5.580)	(49.98)
size	2.065**	-3.795	1.562**	7.128**	2.746	4.477***
	(0.948)	(2.501)	(0.653)	(3.353)	(1.943)	(0.701)
color picture	-3.062	0.921	3.772*		-7.701	10.97
	(3.228)	(3.945)	(2.164)		(20.61)	(13.83)
partial color	-0.234	-7.647**	2.206	4.300		-6.357
	(1.958)	(3.413)	(1.769)	(3.310)		(9.372)
picture	0.0696	-3.692	2.111***	7.292**	10.29**	-8.523
	(0.551)	(2.989)	(0.747)	(2.932)	(5.010)	(9.085)
graph	-0.198	-0.402	-0.157	3.009	-1.866	-0.431
	(0.625)	(4.530)	(0.489)	(2.654)	(3.620)	(2.143)
etc	-0.120	0.121	-0.178	-14.17	2.353	0.700
	(1.235)	(3.430)	(1.475)	(10.45)	(3.540)	(5.130)
# of Books	-65.26***		-67.86***	-68.71***	-47.56***	
	(2.267)		(1.526)	(3.412)	(4.950)	
DVD, CD, MP3	4.712*		9.266*	11.71**		
	(2.484)		(5.627)	(4.970)		
Constant	963.4***	842.8***	1,003***	513.9***	884.6***	299.5
	(33.94)	(80.36)	(19.07)	(63.21)	(55.45)	(300.9)
Obs.	6,173	144	3,512	232	232	70
Adj. R^2	0.568	0.777	0.683	0.751	0.566	0.630
# of IDs	2,456	57	1,411	114	109	31

Standard errors clustered at IDs are parentheses
*** p<0.01, ** p<0.05, * p<0.1

Chapter 3. Impact of Resale Price Maintenance (RPM) Regulation on the Used Book Market

3.1 Introduction

The most significant difference from the existing resale price maintenance (RPM) regulation, implemented in November 2014, is a book which is published more than 18 months ago is newly included in the new regulation coverage. Before the new RPM regulation, price competition among large online bookstores and low search cost (various search engines instantly compare prices of a book at all online bookstores if we enter the title of the book) result in price cuts for books that are not subject to regulation. Discount rates for those books differed depending on the sales and popularity of a book and inventory management of each online bookstore. However, in many cases, a book published more than 18 months ago was sold at online bookstores at sometimes more than 40% discounted price. In this situation, newly-established price regulation for the non-newly published book,[90] directly led to massive price increases and have unintentionally led to increasing demand for used books that are substitution goods for new books.

As a result of the increased price of books, the number of potential buyers in the used book market platform increases, and it causes the matching cost (or searching cost, or transaction cost) to decrease in the used book market platform. This kind of network effect is the property of a two-sided market. Subsequently, decreased searching and matching cost in the used book market caused by more buyers on the platform also could raise an incentive of used-book sellers to attend the platform even though the price of a used book has not changed after the new RPM regulation. More sellers in the platform decrease the matching or transaction cost of used book buyers again, and it consecutively attracts customers of a new book and converts the demand to an online used book market platform. Accordingly, the RPM regulation which significantly increases the price of a non-newly published book could trigger the positive network effect in the used book market platform and sequentially and repeatedly increase the number of players in the platform and decrease the demand in the new book market.

The used book market has rapidly grown since the new RPM regulation was

[90] In this study, a non-newly-published book refers to a book that is published more than 18 months ago.

enacted. For example, 523,401 kinds of used books were available at Aladin (No.1 online book company in the used book market (www.aladin.co.kr)) immediately before the new RPM regulation was in effect (at Oct.25.2014). However, this number increased to 814,306 on Apr. 24. 2016.[91] The Aladin also opened three more used book stores interworking with their online platform in 2017. The YES24 (The second biggest online book company (www.yes24.co.kr)) newly started their business for dealing used books after Nov. 2014 and established a strategic alliance with the Yongpung (The fifth biggest firm operating both online and offline stores) to buy used books for sales. According to a recent news report, their revenue from the used book part has been growing on average 30% for each month.[92] As a result of the remarkable growth, the YES24 also opened offline used bookstores in Seoul.

In addition to aggressive entry into the used book market, large online firms devised a new sales strategy to cope with the new RPM regulations. They call it the "Buy-Back" service, which is not under the RPM restriction. By using this service, customers can buy a new book at a largely discounted price (40∼50%) under the contract that they will re-sell their books to online stores after a particular predetermined time. The bottom line is that the rapid growth of the used book market – which is also dominated by online firms – makes it harder to achieve the policy purpose of new RPM regulation. However, paradoxically, all the above changes are triggered by the new RPM regulation itself. Also, as we can see from the case of the "Buy-Back" service, large online firms are capable of avoiding regulations by looking at the blind spot of regulation.

Next, this paper examines the characteristics of the used book market and investigates how the new regulations led to the rapid growth of the used book market. When trading a used book online, buyers want to check the quality of the used book, unlike a new book. Thus, asymmetric information problems between the seller and the buyer occur in the process. The asymmetric information can be an obstacle in the online used book market to grow. However, an online platform can solve the asymmetric information problem by using various methods. First, sellers in the platform can upload the picture of their used books in addition to the specific descriptions to

[91] This number was over one million in 2019, and some news reported the rapid growth of the used book market. (http://news.heraldcorp.com/view.php?ud=20160425000116)

[92] The growth rate becomes 2300% annually.

inform the condition of a book. Second, an online platform has system evaluating and managing the sellers' previous transactions. After a transaction, buyers can evaluate their purchasing experience and check about the quality of the used book (e.g., same as expected, above expected, and below expected). Based on the buyers' evaluation, other buyers can observe the seller's reputation from previous transactions when they buy a used book from a seller. That is, an online platform can manage seller reputation to overcome asymmetric information exiting on online market for used goods. Third, online platforms have a payment system suitable for dealing in second-hand goods. Under the payment system, sellers can obtain the money from the platform firm after a buyer receives the product and decides to buy it.

Furthermore, online firms operating used book market platforms have critical advantages in used book tradings compared to small offline bookstores. In the case of a new book, an offline store has almost all books for sale. However, it is not the case with a used book. An offline bookstore only has a little portion of used books among all publications. Consequently, it is highly likely that there is no stock for a book that a buyer wants to purchase in the offline store. Also, one more condition is needed to achieve a transaction for a used book. Even though an offline store holds a used book a buyer wants to buy, the quality of the used book should meet with the quality that the buyer wants. Thus, the possibility that there exists a used book that a consumer wants with the desired condition is quite low (double coincidence). On the other hand, online used book stores have a lot more books with diverse quality than an offline bookstore.

Moreover, there is also a transaction cost to visit the local store to check the condition of a used book. We should visit the store to check the quality of a used book, and if it does not meet the buyer's preference, the transaction will not occur after wasting the buyer's visiting cost. To summarize, much more books with a variety of quality, various systems and methods to overcome asymmetric information, and no physical visiting cost become reasons to make people prefer online bookstores to local offline stores when customers buy a used book.

According to the Publishing Industry Survey by Publication Industry Promotion Agency of Korea (KPIPA) which asks offline bookstore owners about the effect of the rapid growth of used book market on their business (Table 3.1), the proportion of respondents who answered as positive, neutral and negative was 4.3%, 30%, and 60%, respectively. The negative answer was higher for smaller bookstores based on sales and the number of employees. The percentage of respondents who answered positively was 17% in the case of bookstores with 10 or more employees, which is much higher than those of small bookstores.

Table 3.1 Effect of Used Book Market Growth on Offline Stores

(Unit: %)

	Response	Positive	Neutral	Negative
Total	814	4.2	35	60.8
By Sales (₩, Won)				
Less Than 10 Million	48	6.3	25	68.8
10 ~ 50 Million	525	3.6	33.7	62.7
50 ~ 100 Million	143	2.1	41.3	56.6
Greater than 100 Million	90	10	35.6	54.4
By Employees				
Less Than 2	552	4.2	33.9	62
3~4 Employees	153	1.3	38.6	60.1
5~9 Employees	74	4.1	39.2	56.8
More Than 10	35	17.1	28.6	54.3

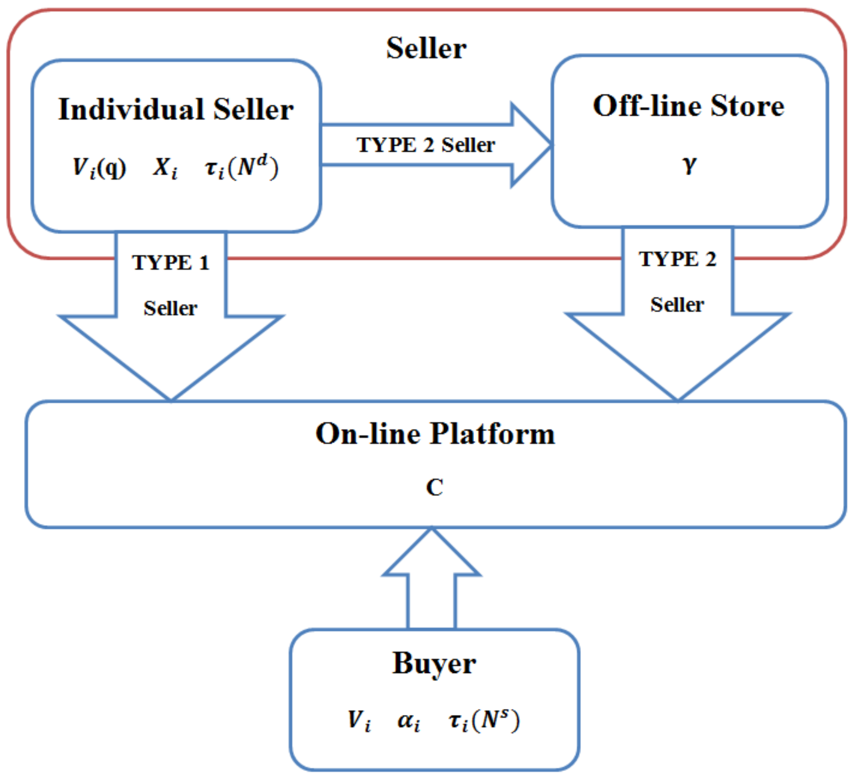
DATA: KPIPA (2017)

In the following Section 3.2, I construct the comprehensive theoretical model including the used book market, which has properties of a two-sided market and is linked with the new book market. From the model, this paper investigates the effects of new RPM regulations on the used book market in Section 3.3. Section 3.4 summarizes the theoretical analysis and suggest another method instead of price regulation to achieve the purpose of resale price maintenance regulation on the book market of Korea.

3.2 Model

3.2.1 Structure of Used Book Market

In the case of a new book, if the books have the same ISBN, they are the exact same products. However, even though used books have the same ISBN, the quality of used books could be different. Thus, we need to introduce the quality variable (q) of a used book. As I discuss in the introduction, online used book market platforms have some critical advantages compared to offline stores. Based on the advantages – much more variety of used books with various quality, and lower search costs compared to offline stores – this paper assumes that all used books are traded on an online platform. Figure 3.1 shows the structure of the used book market that this paper assumes and analyzes.



Note: $V_i(q)$: Seller's valuation for a used book with quality q, X_i: Seller's location on a linear city. Sellers are uniformly located on the linear city. Thus, $X_i \sim$ U[0,1]. $\tau_i(N^d)$: Transaction cost of the Type 1 seller, N^d represent the number of buyers on the platform. γ: commission fee imposed by an offline book store. C: commission fee imposed by an online platform to Type 1 and Type 2 sellers. V_i is a buyer's valuation for a new book. α_i: an individual parameter representing the preference for a used book. $\tau_i(N^s)$: Buyer's transaction cost where N^s is the number of sellers on the platform.

Figure 3.1 Used Book Market

Online Firm

Online firms operate platforms on which sellers and buyers transact each other. The platform has properties of a two-sided market; that is, there exists a network effect. If the number of seller increases, buyer's transaction cost or matching cost decrease and sellers' waiting cost or a matching cost also decreases as the number of buyers increases. This paper assumes that online platform firms charge a commission fee (C) only to the sellers when a transaction is completed, and it becomes their revenue from operating a platform.

Large online firms can quickly expand their business to the used book market by adding a menu or a category for used books on their web pages. When they expand their business to used book markets, their market share in the new book market would significantly affect the market share of the used book market. If a consumer usually buys a new book in "A" online store, the person is highly likely to buy a used book on "A" online platform because one-click converts the web page from a new book market to used book platform. If a firm wants to enter the used book market, it should pay a substantial initial investment cost, because the used book market is already an oligopoly market by some large online firms. In Korea, only four major online firms explain more than 90% of the total revenue of the online book market. Thus, this paper additionally supposes that there is no entrant to the used book market platform.

Offline Store

Based on the above the assumption that all used books are traded on an online platform, offline stores buy used books at the local market from individual sellers and sell those used books at an online platform. Thus, this paper supposes that an offline store buys a used book from an individual living in the local at a price $P_q - \gamma$ and sell it at a price P_q at an online platform. P_q is the used book price of a certain quality (q), and γ is commission fee charged by an offline store for selling a used book instead of an individual seller. Thus, offline stores become agents of the Type 2 Seller in Figure 3.1.

Buyer

Buyers maximize their utility by comparing utilities from buying a used book of a certain quality or a new book. This paper assumes that buyer's valuation (V_i) for a new book is uniformly distributed from 0 to 1. Thus, $V_i \sim U[0, 1]$. Also, this paper introduces the individual parameter α_i representing an individual's preference for a used book and shows the degree of diminishing marginal utility from used book quality. People have different preference over used-book quality and α_i captures individual heterogeneity over the quality of a used book. When a person buys a used book at an online platform, there is a transaction cost or searching cost ($\tau_i(N^s)$). It decreases as the number of sellers (N^s) increase because of a network effect.

Seller

Sellers decide whether they sell their used books directly at an online platform or sell it to an offline store located at the center of the linear city. This paper assumes that sellers are uniformly distributed over the linear city. X_i shows the location of seller i and $X_i \sim U[0,1]$. Each seller also has the valuation for an own used book with a certain quality (q). Even though the quality of a used book is the same, the individual value for the used book can be different. Thus, this paper introduces another individual parameter $V_i(q)$ representing the individual $i's$ heterogeneous valuation for a used book of quality q. When they sell their used book directly via an online platform (Type 1 Seller), they should pay the commission fee (C) and a transaction cost ($\tau_i(N^d)$, N^d means the number of buyers). If they choose to sell it to an offline store which is located at the center of the linear city (Type 2 Seller), they pay the commission fee (γ) imposed by the local offline store and also a transaction cost depending on their location (X_i).

3.2.2 Demand of Used Book Market

Discrete Used Book Quality Choice Model

The model assumes a used book has one of the two qualities value (q_L (low-quality) or q_H (high-quality)) and they are given satisfying the condition that $0 \ll q_L \ll q_H \ll 1$ which mean the values are apart from each other enough to have interior solution.[93] If the value of quality is equal to 1, it is the same with a new book. A buyer $i's$ utility function for buying a used book of quality $j \in [L, H]$ is the following:

$$U_{ij}^{used} = V_i \cdot q_j - \frac{1}{2}\alpha_i \cdot q_j^2 - \tau_i(N^s) - P(q_j) \tag{11}$$

where $V_i \sim U[0,1]$ is person $i's$ value for a new book. α_i is the degree of diminishing marginal utility for used book quality which captures individual heterogeneity over the

[93] To buy a used book, there is a fixed transaction cost to buy it. Thus if the quality is too low, there is no market for a low-quality used book.

quality of a used book. $\tau_i(N^s)$ is transaction cost or searching cost that buyers need to pay when they use the online platform, and it shows the network effect of a two-sided market; that is the online platform.[94] $P(q_j)$ is price of used book with quality $j \in [L, H]$. The price of a used book is endogenously decided by supply and demand for the used book.

Conditions and Properties of Utility Function

This paper imposes some conditions that the above utility function should satisfy.

(Condition 1) Utility from getting a used book without any cost cannot be greater than the utility from the new book. From this condition, the lower boundary of α_i is derived. That is, $\alpha_i > 0$.

$$V_i \cdot q_j - \frac{1}{2}\alpha_i \cdot q_j^2 < V_i$$

(Condition 2) Marginal utility of increasing used book quality should be always greater than 0 for all possible quality levels. The upper boundary of α_i is derived from this condition as $\alpha_i < V_i$

$$V_i - \alpha_i \cdot q_j > 0$$

From above utility function and parameter conditions, we can derive the following properties.

(Property 1) Buyer $i's$ utility from buying a used book is the increasing function of V_i

(Property 2) Buyer i's marginal utility from increasing used book quality is also the increasing function of V_i.

$$\frac{d(U_{ij}^{used})}{d(quality)} = V_i - \alpha_i q_j$$

[94] I assume that $\frac{d(\tau_i(N^s))}{d(N^s)} < 0$ and $\frac{d^2(\tau_i(N^s))}{d(N^s)d(N^s)} > 0$ Thus, the changes in the network effect become smaller as the number of seller increase.

(Property 3) Marginal utility is decreased as the quality increase (by the law of diminishing marginal utility). However, this effect is different depending on α_i. If $\alpha_A > \alpha_B$, then person A's speed of decreasing marginal utility from increasing used book quality is faster than person B's. This individual characteristic makes a person choose the different quality of used book even though they have the same value for the new book (e.g., $V_A = V_B$). Even though a group of people has the same value for a new book, they could have different preference over used-books according to the quality of the used book. For instance, some people care much about the quality of a used book while other people do not. α_i captures this heterogenous property of people. The speed of decreasing marginal utility can be independent of V_i. However, α_i is not independent of V_i. It has a different boundary according to V_i.

(Property 4) α_i has different upper and lower boundary according to V_i. First, to satisfy (Condition 1), α_i should be greater than 0. That is the low boundary of α_i. Second, to satisfy the (Condition 2), α_i should be less than V_i. An intuitive explanation for different α boundary according to V_i is the following: If V_A is low, the marginal utility of increasing used book quality at q (close to 0) should also be lower than that of V_B ($> V_A$). However, even though the value of marginal utility is low, it should stay greater than 0 for all quality levels. Thus, the marginal utility of increasing used book quality should decrease with relatively slow speed for the person whose V_i is low. Thus, low V_i restricts the maximum speed of decreasing marginal utility. On the other hand, if V_i has high value, α_i could have a wide range of value for V_i.

Based on the above utility function and individual parameters (V_i, α_i), It is decided whether to buy a new book, to buy a high-quality used book or a low-quality used book.

Conditions for Buying High-Quality Used Book

(1) $U_{iH}^{used} > U_{iL}^{used}$

Utility from buying high-quality used book should be greater than that of buying low-quality used book.

$$
\begin{aligned}
& U_{iH}^{used} - U_{iL}^{used} > 0 \\
\Rightarrow\ & V_i \cdot (q_H - q_L) - (1/2) \cdot \alpha_i \cdot (q_H^2 - q_L^2) - [P(q_H) - P(q_L)] > 0 \\
\Rightarrow\ & \frac{2[V_i \cdot (q_H - q_L) - (P(q_H) - P(q_L))}{q_H^2 - q_L^2} > \alpha_i
\end{aligned}
\tag{12}
$$

This paper writes the above equation (12) for the condition as $f_{HL}(V_i) > \alpha_i$. In Figure 3.2, it is area below the red line shows the combination of (V_i, α_i) where the utility of buying high-quality used book is the same with that of buying low-quality used book.
(2) $U_{iH}^{used} > Max[U_i^{new}, 0] = Max[V_i - C_{new}, 0]$
The second condition means that the utility of buying high-quality used book should be greater than that of buying a new book. C_{new} is cost of buying new book including transaction or searching cost. Utility from buying high-quality used book should greater than both buying a new-book and zero (buy nothing).

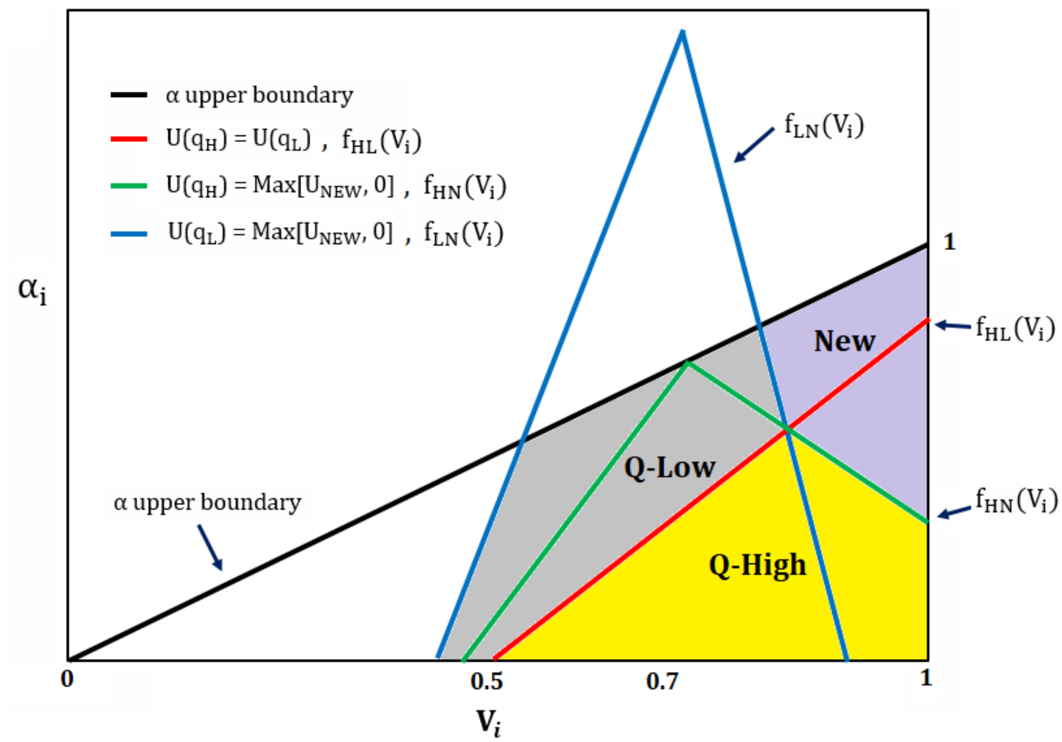
$$
\begin{aligned}
& U_{iH}^{used} > Max[U_i^{new}, 0] \\
\Rightarrow\ & V_i \cdot q_H - (1/2)\alpha_i \cdot q_H^2 - \tau_i(N^s) - P(q_H) > Max[V_i - C_{new}, 0] \\
\Rightarrow\ & \frac{2[V_i \cdot q_H - \tau_i(N^s) - P(q_H) - Max[V_i - C_{new}, 0]]}{q_H^2} > \alpha_i
\end{aligned}
\tag{13}
$$

This paper writes the above equation (13) for the condition as $f_{HN}(V_i) > \alpha_i$. In Figure 3.2, it is area below the green line. Thus, the area satisfying both $f_{HL}(V_i) > \alpha_i$ and $f_{HN}(V_i) > \alpha_i$ buy high-quality used book. It is overlapped area where below the green line and below the red line, that is yellow shaded area in Figure 3.2.

$$
\begin{aligned}
& Area(Buy\ High\ Quality\ Used\ Book) = (V_i, \alpha_i) \\
& s.t \qquad min[V_i,\ f_{HL}(V_i),\ f_{HN}(V_i)] > \alpha_i > 0
\end{aligned}
$$

Conditions for Buying Low Quality Used Book

(1) $U_{iH}^{used} < U_{iL}^{used}$
This condition is the same with the equation (12) with the reverse inequality sign. Thus, the condition can be written $f_{HL}(V_i) < \alpha_i$. In Figure 3.2, it is area above the



This graph is drawn by using the parameter values as following: $C_{new} = 0.7$, $q_H = 0.7$, $q_L = 0.4$, $P(q_H) = 0.3$, $P(q_L) = 0.15$ and $\tau(N^s) = 0.02$.

Figure 3.2 Discrete Used Book Quality Model

red line.

(2) $U_{iL}^{used} > Max[U_i^{new}, 0] = Max[V_i - C_{new}, 0]$

$$\Rightarrow V_i q_L - (1/2)\alpha_i q_L^2 - \tau_i(N^s) - P(q_L) > Max[V_i - C_{new}, 0]$$
$$\Rightarrow \frac{2[V_i q_L - \tau_i(N^s) - P(q_L) - Max[V_i - C_{new}, 0]]}{q_L^2} > \alpha_i \quad (14)$$

The second condition means that the utility of buying low-quality used book should be greater than that of buying a new book. This paper also writes the equation (14) as $f_{LN}(V_i) > \alpha_i$. In Figure 3.2, it is area below the blue line. Thus, the area satisfying both conditions – $f_{HL}(V_i) < \alpha_i$ and $f_{LN}(V_i) > \alpha_i$ – buy low-quality used book (gray shaded area).

$$Area(Buy\ Low\ Quality\ Used\ Book) = (V_i, \alpha_i)$$
$$s.t \qquad min[V_i,\ f_{LN}(V_i)] > \alpha_i > f_{HL}(V_i)$$

Conditions for Buying a New Book

(1) $U_{iH}^{used} < Max[U_i^{new}, 0] = Max[V_i - C_{new}, 0]$

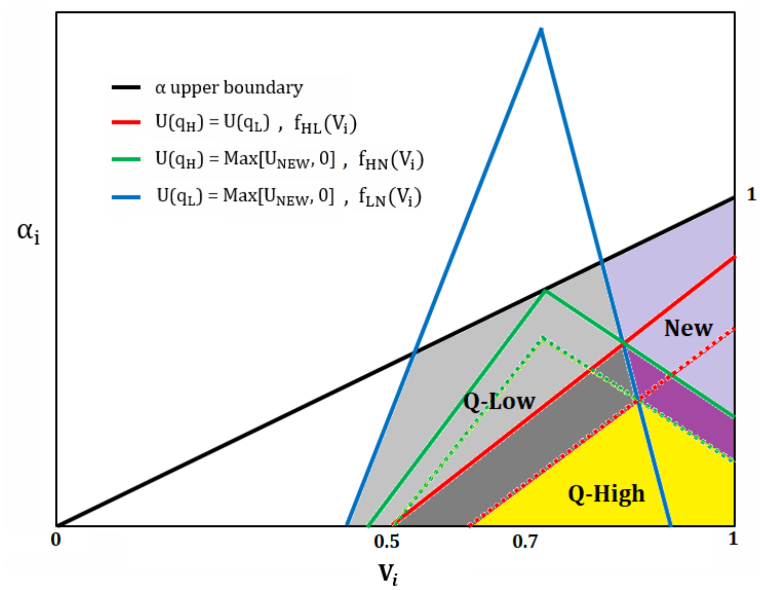
This condition can be written as $f_{HN}(V_i) < \alpha_i$ from the above equation (13) which represents the area above the green line.

(2) $U_{iL}^{used} < Max[U_i^{new}, 0] = Max[V_i - C_{new}, 0]$ This condition is the same with the above equation (14) with reverse inequality, thus it can be written as $f_{LN}(V_i) < \alpha_i$ which represents the area above the blue line. The area satisfying both conditions – $f_{LN}(V_i) < \alpha_i$ and $f_{HN}(V_i) < \alpha_i$ – buy a new book (purple shaded area).

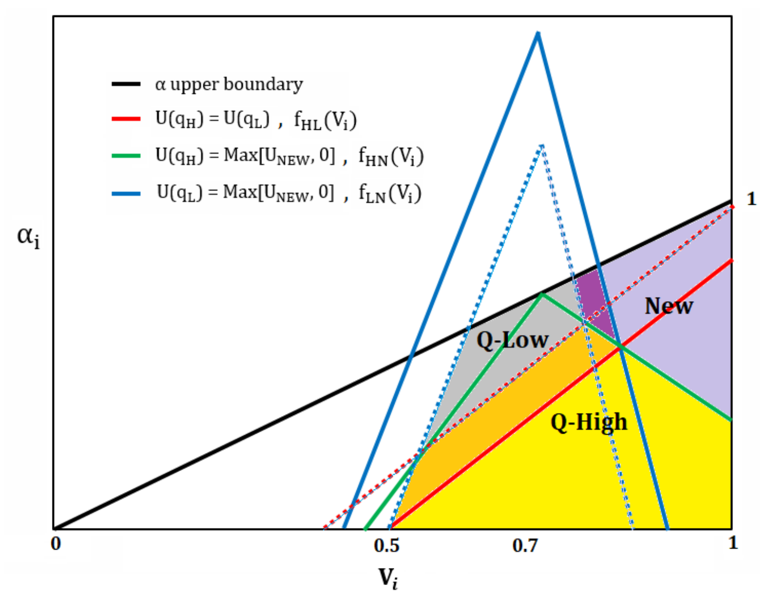
Effects of Changing Prices and Parameters

Figure 3.3 and 3.4 show the variations in Figure 3.2 according to changes in prices and transaction cost. Figure 3.3a show the situation when the price of high-quality used book ($P(q_H)$) increases from 0.3 to 0.33. Dashed lines show the effect of increasing $P(q_H)$ for each line. Both $f_{HL}(V_i)$ and $f_{HN}(V_i)$ shift downward, thereby the area for the high-quality used book (yellow shaded area) shrink. The dark purple area is transferred to the new book area, and the dark gray area is converted to the low-quality used book area from the high-quality used book area. Figure 3.3b shows the situation when the price of low-quality used book ($P(q_L)$) increases from 0.15 to 0.18. Also, dashed lines show the effect of increasing $P(q_L)$ for each line – $f_{HL}(V_i)$ shifts upward and $f_{HN}(V_i)$ shifts downward. As a result, the area for low-quality used book (gray shaded area) shrink. The dark yellow area is changed to high-quality used book area, and the dark purple area is changed to the new book area from the demand for a low-quality used book.

Figure 3.4a shows the changes in functions when buyers' transaction cost at online platform ($\tau_i(N^s)$) increases from 0.02 to 0.05. Dashed lines show the effect of the increased transaction cost for each function. Both $f_{LN}(V_i)$ and $f_{HN}(V_i)$ shift downward. As a result, both the low and high-quality used book areas shrink and the area for buying new book expands. The dark purple area is changed to the new book area from demands for the high and low-quality used book. On the other hand, Figure 3.4b shows the situation when the cost of buying a new book (C_{new}) increases from 0.7 to 0.75. Again, dashed lines show the effect of increasing C_{new} for each function



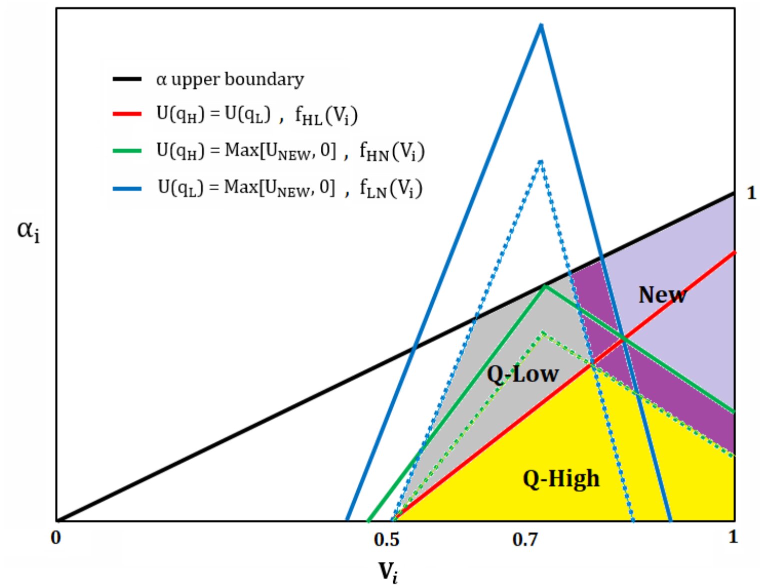
(a) Effect of Increasing $P(q_H)$



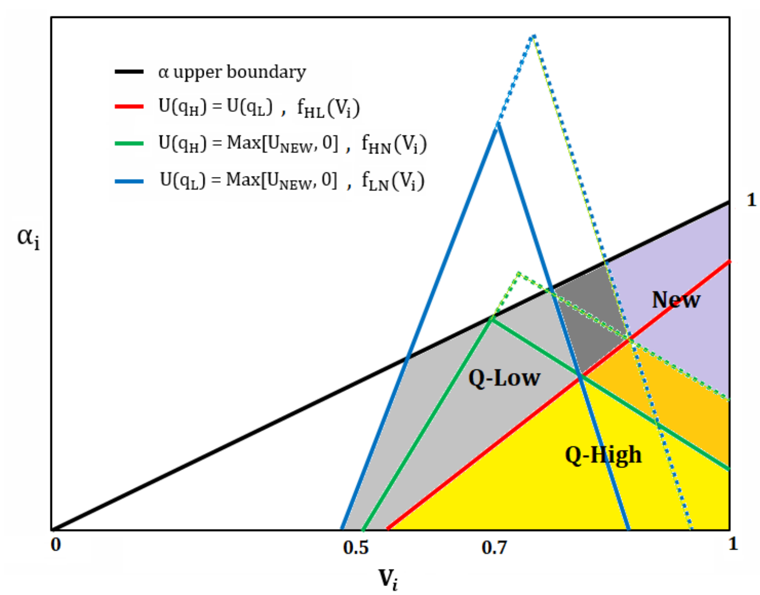
(b) Effect of Increasing $P(q_L)$

Figure 3.3 The Effect of Increasing Used Book Price

NOTE: Graphs are drawn by using the parameter values as following: $C_{new} = 0.7$, $q_H = 0.7$, $q_L = 0.4$, $P(q_H) = 0.3$, $P(q_L) = 0.15$ and $\tau(N^s) = 0.02$. (a) shows the effect of increased $P(q_H)$ from 0.3 to 0.33. (b) illustrates the effect of increased $P(q_L)$ from 0.15 to 0.18.



(a) **Effect of Increasing** τ



(b) **Effect of Increasing** C_{new}

Figure 3.4 The Effect of Changing Parameters

NOTE: Graphs are drawn by using the parameter values as following: $C_{new} = 0.7$, $q_H = 0.7$, $q_L = 0.4$, $P(q_H) = 0.3$, $P(q_L) = 0.15$ and $\tau(N^s) = 0.02$. (a) shows the effect of increased τ from 0.02 to 0.05. (b) illustrates the effect of increased cost for buying a new book from 0.7 to 0.75

– both $f_{HN}(V_i)$ and $f_{LN}(V_i)$ shift upward. Consequently, dark yellow area converts to high-quality used book demand, and the dark gray area turns to low-quality used book area from the new book demand area.

Probability and Demand

Next, we can calculate demand for each product (high-quality used book, low-quality used book, new book) from the above (V_i, α_i) plane graph (Figure 3.2, 3.3 and 3.4). The probability density function (pdf) of (V_i, α_i) plane is $\frac{1}{V_i}$ given that $\alpha_i \sim U[0, V_i]$ and $V_i \sim U[0, 1]$.

Probability of Buying a High-Quality Used Book

The probability of buying a high quality used book for V_i is calculated by integrating the area of high quality used book in Figure 3.2. This paper writes the probability function as $f_{pr}^{q_H}(V_i)$. The blue line in Figure 3.5 shows the probability of buying high quality used book according to V_i.

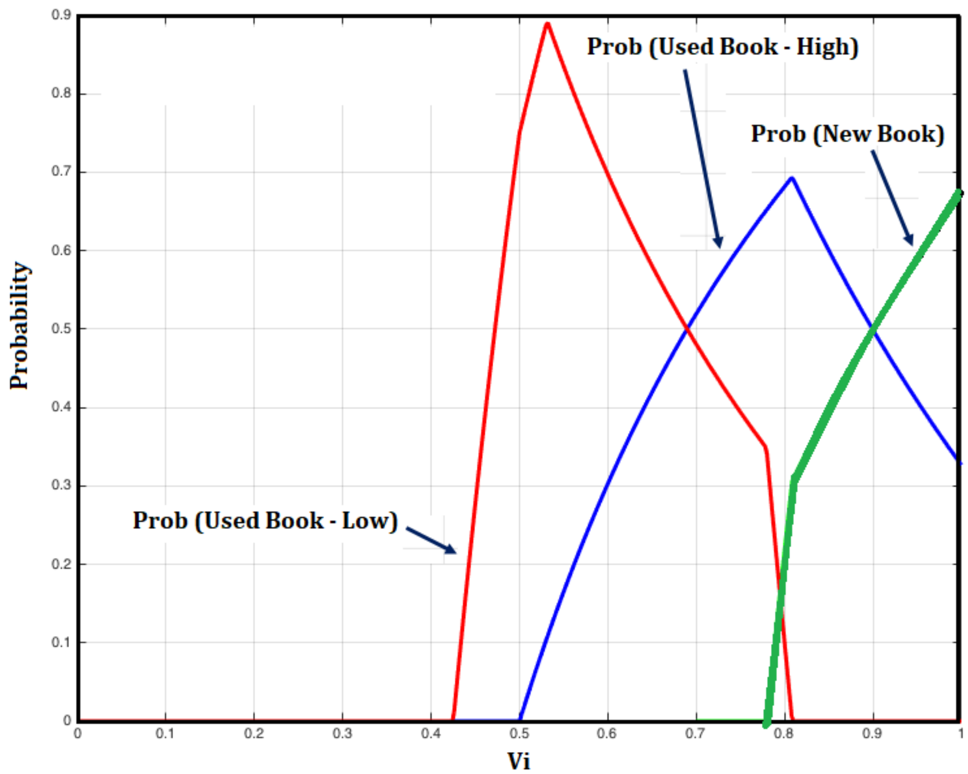
$$
\begin{aligned}
Pr(min[V_i, f_{HL}(V_i), f_{HN}(V_i)] > \alpha_i > 0) &= f_{pr}^{q_H}(V_i) \\
&= min[V_i, f_{HL}(V_i), f_{HN}(V_i)] \cdot p.d.f \\
&= min[V_i, f_{HL}(V_i), f_{HN}(V_i)] \cdot 1/V_i
\end{aligned}
$$

Probability of Buying a Low-Quality Used Book

At the same way, the probability of buying a low-quality used book is obtained by integration of low-quality area which is as follows:

$$
\begin{aligned}
Pr(min[V_i, f_{LN}(V_i)] > \alpha_i > f_{HL}(V_i)) &= f_{pr}^{q_L}(V_i) \\
&= (min[V_i, f_{LN}(V_i)] - f_{HL}(V_i)) \cdot p.d.f \\
&= (min[V_i, f_{LN}(V_i)] - f_{HL}(V_i)) \cdot 1/V_i
\end{aligned}
$$

The red line in Figure 3.5 shows the probability of buying low-quality used book according to V_i.



NOTE: Graphs are drawn by using the parameter values as following: $C_{new} = 0.7$, $q_H = 0.7$, $q_L = 0.4$, $P(q_H) = 0.3$, $P(q_L) = 0.15$ and $\tau(N^s) = 0.02$.

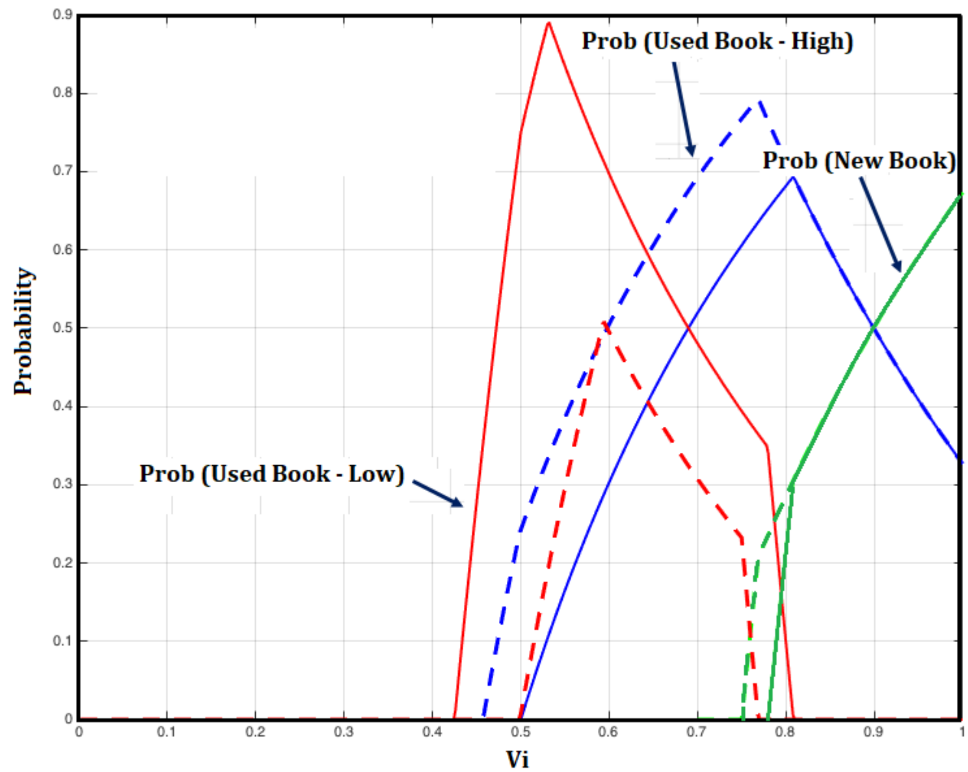
Figure 3.5 Probability of Buying Each Book for V_i

Probability of Buying a New Book

The integration of the new book area which means the probability of buying a new book in Figure 3.2 is calculated as the following:

$$\begin{aligned} Pr(V_i > \alpha_i > Max[f_{HN}(V_i), f_{LN}(V_i) \Big| V_i > C_{new}]) &= f_{pr}^{new}(V_i) \\ &= (V_i - Max[f_{HN}(V_i), f_{LN}(V_i) \Big| V_i > C_{new}] \cdot 1/V_i \end{aligned}$$

The green line in Figure 3.5 represents the probability of buying a new book according to V_i. According to Figure 3.5, if a group of people's value for a new book is 0.7, then 50% of people in the group buy a low-quality used book, and the rest 50% of people buy a high quality used book. In the case of people whose value for a new book is 0.9, 50% of people buy a high-quality used book and the rest 50% of people buy a new book. Figure 3.6 shows the situation when the price of low-quality used book increases from 0.15 to 0.18. The dashed line shows the effect of increasing the low-quality book price. As a result, the probability of buying a low-quality used book decrease and probabilities for both a high quality used book and a new book increase for each V_i.



NOTE: Graphs are drawn by using the parameter values as following: $C_{new} = 0.7$, $q_H = 0.7$, $q_L = 0.4$, $P(q_H) = 0.3$, and $\tau(N^s) = 0.02$. $P(q_L)$ is increased from 0.15 to 0.18.

Figure 3.6 Changes in Probability Functions by Increased $P(q_L)$

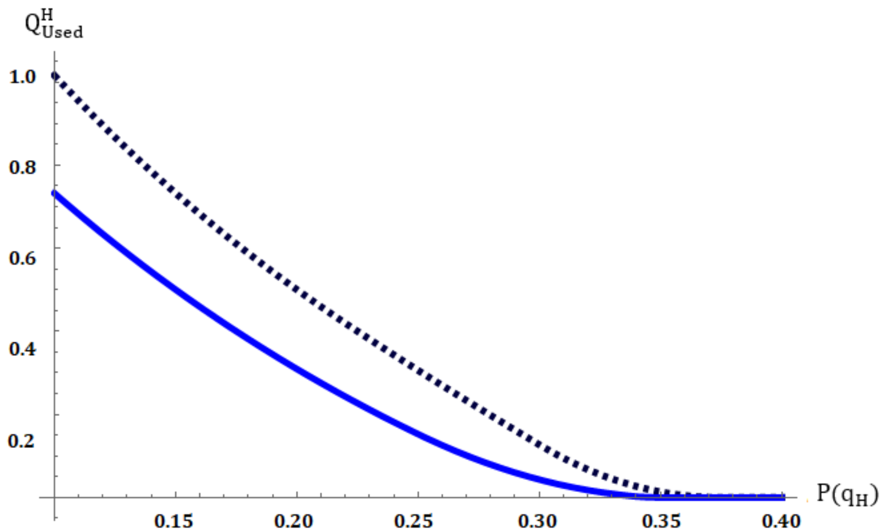
From Figure 3.5 and 3.6 which show the probability of buying each book according to V_i, we can obtain the demand function for each book by integrating probability function as following:

- Demand for a j quality used book

$$\int_0^1 f_{pr}^{q_j}(V_i)dV_i \tag{15}$$

- Demand for a new book

$$\int_0^1 f_{pr}^{new}(V_i)dV_i \tag{16}$$



NOTE: Graphs are drawn by using the parameter values as following: $C_{new} = 0.65$, $q_H = 0.7$, $q_L = 0.4$, $P(q_L) = 0.15$, and $\tau(N^s) = 0.1$. The dashed line shows the effect of decreased $\tau(N^s)$ from 0.1 to 0.05.

Figure 3.7 Demand Curve of High Quality Used Book

Figure 3.7 shows the demand curve of a high-quality used book and the dashed line describes the change of demand when the transaction cost ($\tau(N^s)$) decreases.

3.2.3 Supply of Used Book Market

Type of Sellers

As I discuss in the introduction, this paper assumes there are two types of sellers in a linear city. Figure 3.8 describes the location of each type of seller.

(Type 1 Seller)

This type of sellers sells its own used book via online platform directly. In the result of the transaction, they obtain the utility as follows:

$$U^{seller}_{i,Type1} = P(q_j) - C - V_i(q_j) - \tau_i(N^d) \tag{17}$$

where $P(q_j)$ is the price $j \in [L, H]$ quality used book, C is commission fee they need to pay for using online platform and $V_i(q_j)$ is individual value for own used book of quality j. Although the quality is the same, the individual value for the used book can be different among people. The assumption that $V_i(q_j) \sim U[0, \overline{v_j}]$ considers individual's heterogeneous values on own used book with quality j. That is, the owner's value is distributed over uniformly between 0 and $\overline{v_j}$ ($\overline{v_H} > \overline{v_L}$). $\tau_i(N^d)$ is seller's transaction cost or a user cost (except the commission fee C) at online platforms such as a waiting cost for selling a book or the cost for uploading used book information at online. The transaction cost also has a network effect like the same with $\tau_i(N^s)$ in the demand side and also has the same properties. It decreases as the number of buyers (N^d) increase because of network effect$\Big(\dfrac{d(\tau_i(N^d))}{d(N^d)} < 0\Big)$ and the degree of network effects become smaller as the number of seller increase$\Big(\dfrac{d^2(\tau_i(N^d))}{d(N^d)d(N^d)} > 0\Big)$.

(Type 2 Seller)

This type of sellers sell own used book with quality j to a local offline store at a price $P(q_j) - \gamma$ and the local offline store sell again the used book at an online platform at $P(q_j)$. Thus, the γ become the income of the offline bookstore that acts as a sales agent. For the Type 2 seller, there is a transaction cost $t|X_i - 1/2|$ to visit the local offline store. The utility function of the Type 2 seller is given as follows:

$$U^{seller}_{i,Type2} = P(q_j) - \gamma - t|X_i - 1/2| - V_i(q_j) \tag{18}$$

The Location and Individual Values of Each Type Seller

The seller with $(X_i, V_i(q_j))$ becomes (Type 2) seller if

$$U^{seller}_{i,Type2} > Max[U^{seller}_{i,Type1}, 0]$$

From the condition $U^{seller}_{i,Type2} > U^{seller}_{i,Type1}$, the location of (Type 2) seller is derived as

$$X_i \in \Big(\frac{1}{2} - \frac{\tau_i(N^d) + C - \gamma}{t}, \quad \frac{1}{2} + \frac{\tau_i(N^d) + C - \gamma}{t}\Big)$$

and the condition $U^{seller}_{i,Type2} > 0$ restrict the individual value for own used book (vertical values) as in Figure 3.8. At the same way, we can obtain the location and range of

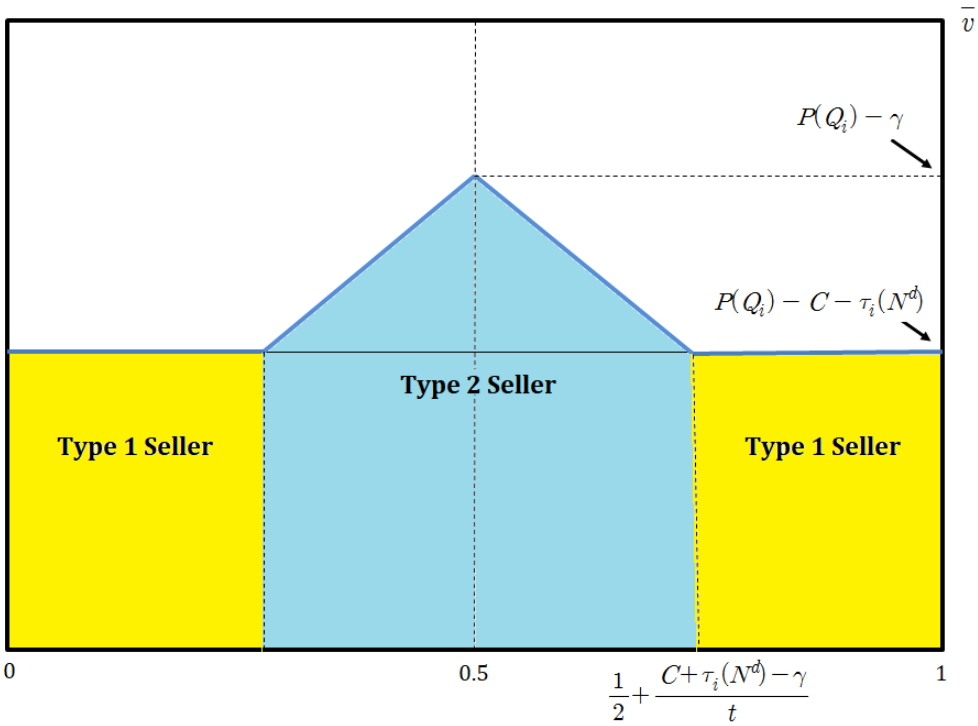


Figure 3.8 Locations and Individual Values of Each Seller

individual value $(X_i, V_i(q_j))$ of (Type 1) seller.

(Supply of Used Book)
The supply of used book with quality j is the area of both (Type 1) and (Type 2) seller in Figure 3.8. Thus, we can derive the supply of quality j used book as follows:

$$Supply = [P(q_j) - C_j - \tau_i(N^d) + \frac{(\tau_i(N^d) + C_j - \gamma)^2}{t}] \cdot \overline{v_j} \tag{19}$$

The online platform commission fee (C) and offline store user cost (γ) are also decided endogenously by the profit maximization of online platform company and offline store. This paper derives the optimal C^* and γ^* in the following section.

Online Platform Firm

According to the model this paper analyzes, an online firm operates the platform for the used book market. Additional cost for operating the used book market platform is almost zero because the firm was already operating the webpage for selling a new book. Thus, it only cost some fixed costs at one time to set up a platform, and the marginal cost of operating the platform for a used book is almost zero. This paper set the profit function of an online firm from operating the online platform as follows:

(Profit function of online firm)

$$\Pi_{on}^{used} = C_L \cdot Q(C_L) + C_H \cdot Q(C_H) \tag{20}$$

where C_j, is an online platform commission fee for selling a j quality used book on the platform. $Q(C_j)$ is used book quantities of quality j traded at the used book market platform which is calculated from equation (19) as follows:

$$Q(C_j) = \frac{1}{2} \cdot [P(q_j) - C - \tau_i(N^d) + \frac{(\tau_i(N^d) + C - \gamma)^2}{t}] \cdot \overline{v_j}$$

When two online platform firms do Bertrand price competition, the commission fee goes down to zero, that is, its marginal cost. On the other hand, if the collusion between two platform firms is possible, they choose the optimal commission fee (C_j^*) from the following equation which is the first order condition of the profit function.[95]

$$\begin{aligned} Q(C_j) + C_j \cdot \frac{d(Q(C_j))}{d(C_j)} &= 0 \\ C_j &= \frac{Q(C_j)}{-d(Q(C_j))/d(C_j)} \end{aligned} \tag{21}$$

where

$$\frac{d(Q(C_j))}{d(C_j)}(< 0) = \frac{\dfrac{2(C_j + \tau_i(N^d) - \gamma)}{t} - 1}{1 - \dfrac{d\tau_i(N^d)}{dQ(C_j)} \cdot \left[\dfrac{2(C_j + \tau_i(N^d) - \gamma)}{t} - 1\right]} \tag{22}$$

When the firm decides the optimal commission fee (C_j^*), there is an additional network effect affected by the fee. The commission fee changes the number of demand (N^d), and it will also affect the transaction cost ($\tau_i(N^d)$). Thus the transaction cost is also

95 Equation(21) and (24) consist of reaction functions of an online firm and an offline store. Thus, the intersection of two curves on (C_j, γ_j) plane becomes an equilibrium.

the function of the commission fee. In equation (22), $\dfrac{d\tau_i(N^d)}{dQ(C_j)}$ captures the indirect network effect considered by the firms when they decide the optimal commission fee. If they do not internalize the indirect network effect, equation (22), is simplified to $\dfrac{2(C_j+\tau_i(N^d)-\gamma)}{t}-1$. The optimal C_j^* is lower when they internalize the indirect network effect.

Offline Store

Offline book store purchases used books from individuals ((Type 2) seller), and sell it via an online platform. It occurs transaction cost ($\tau_L(N^d)$ or waiting and matching cost) for offline books store when they use an online platform to sell the used books. I assume that the transaction cost of the offline book store is lower than that of individuals. That is $\tau_L(N^d) < \tau_i(N^d)$. The reasons for this assumption are: First, an offline book store sells used books frequently; thus they can have a higher reputation and a longer selling history – which make selling easier and decrease a transaction cost or a waiting cost – than an individual seller at an online. Moreover, offline book stores are more specialized in selling used books using the online platform because it is their job. Second, because of the economy of scale, offline book stores can spend less transaction cost per used book. Importantly, this paper assumes that the transaction cost of individual sellers (Type 1 Seller) changes more sensitively than that of offline book stores as the number of buyers increase. That is, $\left|\dfrac{d(\tau_L(N^d)}{d(N^d)}\right| < \left|\dfrac{d(\tau_i(N^d)}{d(N^d)}\right|$. The Figure 3.9 illustrates this assumption.

(Profit Function of Offline Store)

The profit function of an offline store in the used book market is calculated as follows:

$$\begin{aligned}\Pi_{off}^{used} &= [(P(q_H)-C_H)-(P(q_H)-\gamma_H)-\tau_L(N^d)]\cdot Q_{off}^H(\gamma_H)\\ &\quad+[(P(q_L)-C_L)-(P(q_L)-\gamma_L)-\tau_L(N^d)]\cdot Q_{off}^L(\gamma_L)\\ &=[\gamma_H-C_H-\tau_L(N^d)]\cdot Q_{off}^H(\gamma_H)+[\gamma_L-C_L-\tau_L(N^d)]\cdot Q_{off}^L(\gamma_L) \qquad (23)\end{aligned}$$

where $Q_{off}^j(\gamma_j)$ is the used book quantity of quality j which is traded through the offline store (Type 2 Seller) at an online platform, and it is calculated from the area of

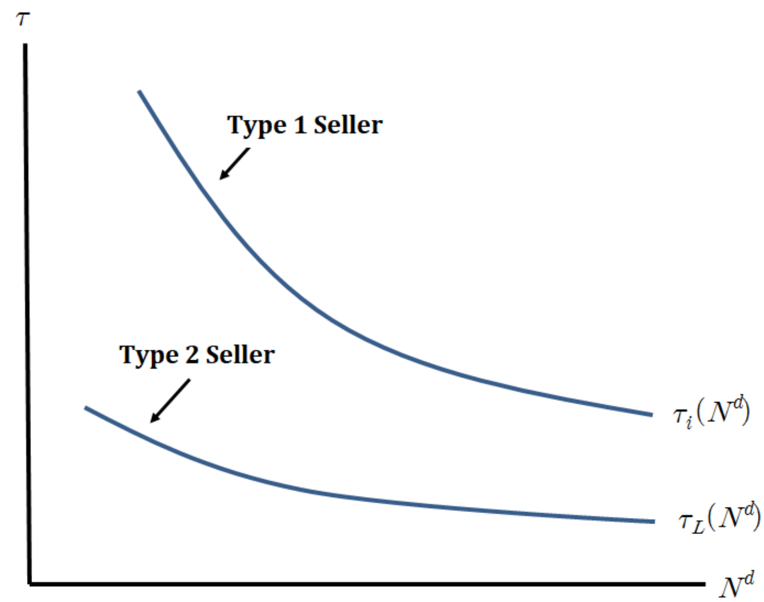


Figure 3.9 Network Effect in Transaction Cost

(Type 2) seller in Figure 3.8 as follows:

$$Q^j_{off}(\gamma_j) = [(C_j + \tau_i(N^d) - \gamma_j)^2 + 2(C + j + \tau_i(N^d) - \gamma_j)(P(q_j) - C_j - \tau_i(N^d)] \cdot \frac{\overline{v_j}}{t}$$

The offline store decides the optimal commission fee for each quality used book γ_H and γ_L to maximize the profit from the equation (23).

From the first order condition of profit function, γ_j^* is calculated from the following equation:[96]

$$\gamma_j = C_j + \tau_L(N^d) - \frac{Q^j_{off}(\gamma_j)}{\frac{d(Q^j_{off}(\gamma_j))}{d(\gamma_j)}} + \frac{d(\tau_L(N^d))}{d(Q^j_{off})} \cdot Q^j_{off} \tag{24}$$

[96] This is the second order equation with respect to γ_j. However, γ_j has unique solution given the condition which is $C_j + \tau_L < \gamma_j < C_j + \tau_i$

where

$$\frac{d(Q^j_{off}(\gamma_j))}{d(\gamma_j)} = \frac{2(P(q_j) - \gamma)}{2t \cdot \dfrac{d(\tau_i(N^d))}{d(Q^j_{off})} \cdot [P(q_j) - C_j - \tau_i(N^d)] - t} \tag{25}$$

When the offline store internalize the indirect network effect, there are two opposite effects. $\frac{d(\tau_L(N^d))}{d(Q^j_{off})}(< 0)$ lowers the optimal γ^*_j and $\frac{d(\tau_i(N^d))}{d(Q^j_{off})}(< 0)$ increases it compared to the situation when they ignore the indirect network effect.[97] Intuitively, $\frac{d(\tau_L(N^d))}{d(Q^j_{off})}$ lowers the optimal γ^*_j, because τ_L is the cost of an offline store and decreasing γ increases the demand (N^d) and lowers its cost (τ_L). Thus, we can interpret this effect as the indirect cost saving effect of decreasing γ. However, decreasing γ also decreases the $\tau_i(N^d)$, and it increases (Type 1 Seller) while decreasing (Type 2 seller). Thus, we can interpret this effect as the competitiveness effect of rivals which increases the optimal γ^*_j.

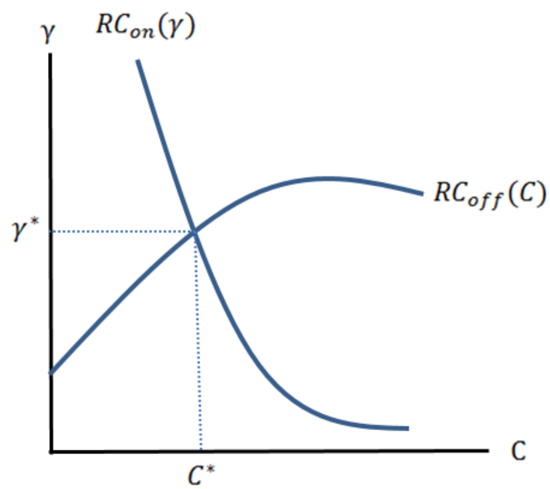


Figure 3.10 Reaction Curves of Online Firm and Offline Store

Equation (21) and (24) consist of reaction functions of the online platform firm and the offline store. $RC_{on}(\gamma)$ is reaction curve of online firm according to offline store's decision (γ) and $RC_{off}(C)$ shows the optimal γ given online firm's commission

97 In this case, equation (24) and (25) are simplified where $\frac{d(\tau_L)}{d(Q^j_{off})} = \frac{d(\tau_i(N^d))}{d(Q^j_{off})} = 0$.

fee (C). The optimal commission fee imposed by the online platform decreases as the γ increases. The reason is that the quantity of used book transacted via the online platform is decreased if the γ increases, as we can check easily from Figure 3.8. That is, increasing C or γ decreases the online firm's quantity, thus C and γ have strategic substitute relation for the online firm's decision. On the other hand, offline store's optimal price increases as the commission of the online firm increases. For offline stores, they compete with the online platform with price γ. Individual sellers become (Type 1 Seller) or (Type 2 seller) by comparing C and γ. Thus, for the offline stores, C and γ have strategic complementary relation. Moreover, note that C also acts as the cost of offline stores. Increasing C means aggravated cost condition of the offline store and push the offline store's price upward.

3.3 Effects of New RPM Regulation and the Equilibrium of Market

3.3.1 Profit of Offline Stores

Because of the new RPM regulation, a part of demand in new book market moves to the used book market. As a result, the number of demand (N^d) in the used book market increases that lowers the transaction cost of the online platform by the network effect. Also, an increased number of buyers at the platform directly affects the profit of an offline store. I investigate two cases when the two platform firms do Bertrand price competition with commission fee (C_j) and when the collusion for C_j is possible.

(Proposition 1)

If the following condition is satisfied, the profit of offline store is decreased as the number of demand increases in used book market[98] when two online platform firms do Bertrand price competition.

$$(1 <)\frac{P(q_j) - \gamma_j}{P(q_j) - \tau_i(N^d) - C_j} < \frac{\left|\dfrac{d(\tau_i(N^d))}{d(N^d)}\right|}{\left|\dfrac{d(\tau_L(N^d))}{d(N^d)}\right|} \tag{26}$$

(Proposition 2)

[98] The proof is in appendix

If we consider the case when the online firms can collude, the condition that the profit of an offline store is decreased as the number of demand increases in the used book market is changed as the follows:

$$(1 <)\frac{P(q_j)-\gamma_j}{P(q_j)-\tau_i(N^d)-C_j} < \frac{\left|\frac{d(\tau_i(N^d))}{d(N^d)}+\frac{d(C_j)}{d(N^d)}\right|}{\left|\frac{d(\tau_L(N^d))}{d(N^d)}+\frac{d(C_j)}{d(N^d)}\right|} \qquad (27)$$

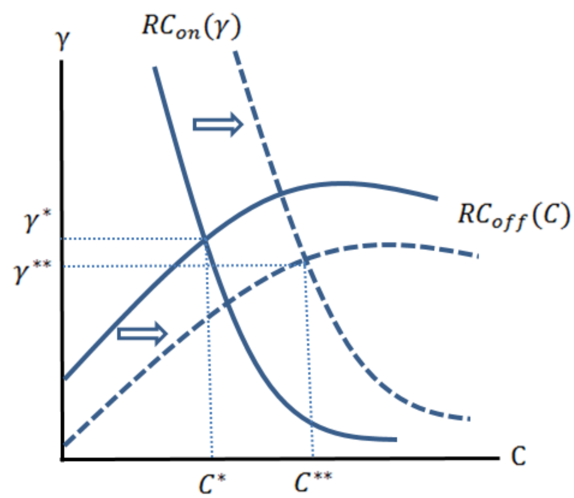


Figure 3.11 Effect of Increasing Demand on Reaction Curves

If an individual's amount of decreasing transaction cost caused by the network effect from increasing demand is greater than that of the offline store and satisfy the above condition, then the profit of the offline store is decreased. The intuitive reason is that γ_j should be $C_j+\tau_L(N^d) < \gamma_j < C_j+\tau_i(N^d)$ to have positive demand ($Q^j_{off} > 0$). In this situation, if the number of demand (N^d) increases, the transaction cost of an individual ($\tau_i(N^d)$) decreases more than that of an offline store ($\tau_L(N^d)$) because of the reason explained in section 3.5.2. This effect aggravates offline firms' profit condition. Lowered γ_j can increase the demand (Q^j_{off}), but this is also restricted by decreasing $\tau_i(N^d)$. Figure 3.11 describes the effects of increasing demand (N^d) on both reaction curves. As a result of increasing demand (N^d), both curves shift to the right. Consequently, the optimal commission fee C^*_j increases and the optimal price of the offline store γ^*_j can be increased or decreased.

3.3.2 Equilibrium of the Used Book Market

To sum up, the equilibrium of the used book market of quality j $\{P(q_j), Q_j, C_j, \gamma_j\}$ is decided by the following equations.

(1) Demand for j quality used book

$$\int_0^1 f_{pr}^{q_j}(V_i)dV_i \tag{15}$$

(2) Supply for j quality used book

$$[P(q_j) - C_j - \tau_i(N^d) + \frac{(\tau_i(N^d) + C_j - \gamma_j)^2}{t}] \cdot \overline{v_j} \tag{19}$$

(3) Reaction function of online firm[99]

$$C_j = \frac{Q(C_j)}{-d(Q(C_j))/d(C_j)} \tag{21}$$

(4) Reaction function of offline store

$$\gamma_j = C_j + \tau_L(N^d) - \frac{Q_{off}^j(\gamma_j)}{\frac{d(Q_{off}^j(\gamma_j))}{d(\gamma_j)}} + \frac{d(\tau_L(N^d))}{d(Q_{off}^j)} \cdot Q_{off}^j \tag{24}$$

$$s.t \quad C_j + \tau_L(N^d) < \gamma_j < C_j + +\tau_i(N^d)$$

Figure 3.12 shows the changes in equilibrium in high quality used book market caused by the new RPM regulation. At first, it increases the number of demand (N^d) for high quality used book which is illustrated in Figure 3.4b. After deriving the probability functions from the (V_i, α_i) plane, we can obtain the new demand curve (dashed negative slope in Figure 3.12) by integrating the changed probability functions. Also, more buyers in the online platform decrease the transaction costs $(\tau_i(N^d))$ of used booksellers, thereby the number of seller increases. These changes also affect the optimal decisions of firms in the used book market described in the reaction curve (Figure 3.10). As a result of the shock from new RPM regulation, sellers and online platform companies change their decisions, which changes the supply curve (dashed

[99] If online firms do Bertrand price competition, $C_j^* = 0$.

upward slope in Figure 3.12.

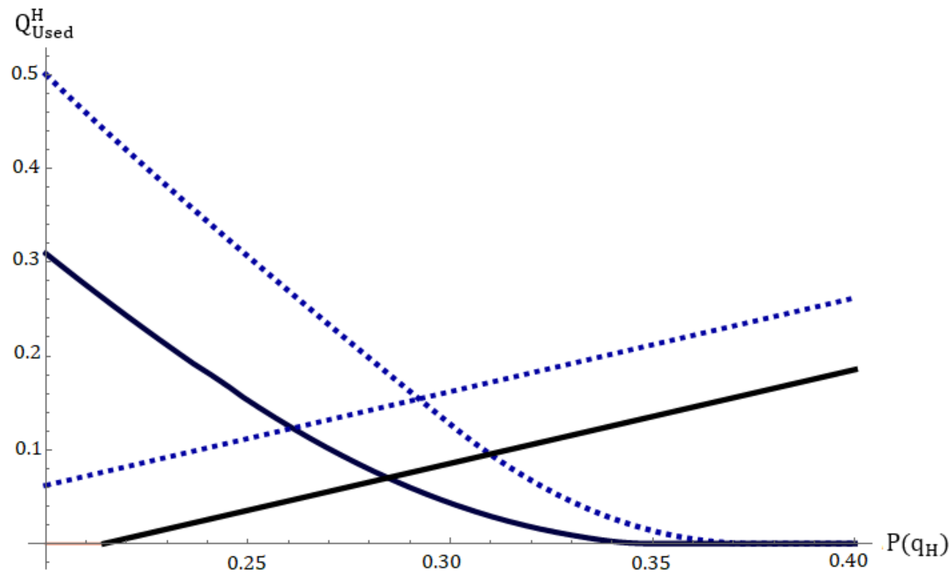


Figure 3.12 Effect of Increasing Demand on Used Book Market

3.4 Conclusion

The reinforced resale price maintenance (RPM) regulation took effect from November 2014 in Korea under the policy purpose to protect a variety in culture and knowledge by supporting the business of small-scale publishing companies and offline bookstores. The biggest changes from the previous RPM regulation is that a non-newly-published book – which is published more than 18 months ago – is also included in the new regulation coverage. As a result, the new regulation naturally causes a massive price increase of non-newly-published books. A non-newly-published book has substitution goods, that is a used book. Thus, the new RPM regulation unintentionally sent customers for a new book to a used book market and triggers the rapid growth of the used book market in Korea.

The problem is that large online firms have also dominated the used book market. After the new RPM regulation took effect, online firms aggressively enter the used book market. As I discuss in the introduction, online firms are in a better position in the used book market than offline bookstores because they have a much larger variety of books with diverse qualities. Moreover, large online firms have various marketing tools that can legally evade the new RPMregulation as we can see at the case of "Buy-Back" service. According to the survey by KPIPA, 60% of offline bookstore owners answered that the rapid growth of the used book market negatively affected their business.

This paper shows that the profits of offline stores from the used book market can be diminished based on the theoretical model including the used book market, which has properties of a two-sided market and is linked with the new book market. If the diminishing transaction cost of an offline store is relatively smaller than that of an individual seller, when the increasing number of buyers decrease the transaction cost by the network effect of the two-sided market, then the profit of an offline store decrease at the used book market.

Based on the problems of new RPM regulation, I suggest considering an alternative method to accomplish the policy purpose of RPM regulation on the book market. We can think of direct subsidy instead of existing the price regulation.

Nowadays, many people buy books online after browsing the book at local offline stores. From the local offline store, they get some benefit, however, they do not pay anything for that. Offline stores have cost for displaying books and operating the offline store. However, online firms get some benefits indirectly from the service offered by offline local stores. Offline local stores act like a real displaying place for online firms which do not have an actual place for display. Thus, there is an externality in the used book market and the existence of this externality justifies the transfer of profit or direct subsidy from online firms to offline bookstores. The direct subsidy as a corrective tax to fix the externality between online firms and offline bookstores can more effectively achieve the policy goal, less distorting the book market than price regulation such as RPM. In the following research, we can consider the market-friendly incentive design considering the externality that offline stores serve as display places for online firms.

APPENDIX

APPENDIX C

Proof of Proposition 1 and 2

From the profit function of an off-line store in used book market(equation (23))

$$\Pi_{off}^{used} = [\gamma_H - C_H - \tau_L(N^d)] \cdot Q_{off}^H(\gamma_H) + [\gamma_L - C_L - \tau_L(N^d)] \cdot Q_{off}^L(\gamma_L) \tag{23}$$

$$\begin{aligned}
\frac{d\Pi_{off}^{used}}{dN^d} &= \Big[\frac{d\gamma_j}{dN^d} - \frac{dC_j}{dN^d} - \frac{d\tau_L}{dN^d}\Big] \cdot Q_{off}^j(\gamma_j) + \Big[\gamma_j - C_j - \tau_L(N^d)\Big] \cdot \frac{dQ_{off}^j(\gamma_j)}{dN^d} \\
&= \Big[\frac{d\gamma_j}{dN^d} - \frac{dC_j}{dN^d} - \frac{d\tau_L}{dN^d}\Big] \cdot Q_{off}^j(\gamma_j) - \frac{Q_{off}^j(\gamma_j)}{dQ_{off}^j(\gamma_j)/d\gamma_j} \cdot \frac{dQ_{off}^j(\gamma_j)}{dN^d} \\
& where \quad \frac{dQ_{off}^j(\gamma_j)}{dN^d} = \frac{d\gamma_j}{dN^d} \cdot \frac{\partial Q_{off}^j}{\partial \gamma_j} + \frac{dC_j}{dN^d} \cdot \frac{\partial Q_{off}^j}{\partial C_j} + \frac{d\tau_i}{dN^d} \cdot \frac{\partial Q_{off}^j}{\partial \tau_i} \\
&= \Big[-\frac{dC_j}{dN^d} - \frac{d\tau_L}{dN^d}\Big] \cdot Q_{off}^j(\gamma_j) - \frac{Q_{off}^j(\gamma_j)}{dQ_{off}^j(\gamma_j)/d\gamma_j} \cdot \Big[\frac{dC_j}{dN^d} \cdot \frac{\partial Q_{off}^j}{\partial C_j} + \frac{d\tau_i}{dN^d} \cdot \frac{\partial Q_{off}^j}{\partial \tau_i}\Big]
\end{aligned} \tag{28}$$

$$where \quad \frac{dQ_{off}^j(\gamma_j)}{d\gamma_j} = P(q_j) - \gamma_j \quad and \quad \frac{\partial Q_{off}^j}{\partial \tau_i} = \frac{\partial Q_{off}^j}{\partial C_j} = P(q_j) - \tau_i - C_j$$

Thus, the condition that the profit of an off-line store decrease as the number of demand for used book is following :

$$\begin{aligned}
&\frac{d\Pi_{off}^{used}}{dN^d} < 0 \\
&\Rightarrow \Big[-\frac{dC_j}{dN^d} - \frac{d\tau_L}{dN^d}\Big] \cdot Q_{off}^j(\gamma_j) - \frac{Q_{off}^j(\gamma_j)}{dQ_{off}^j(\gamma_j)/d\gamma_j} \cdot \Big[\frac{dC_j}{dN^d} \cdot \frac{\partial Q_{off}^j}{\partial C_j} + \frac{d\tau_i}{dN^d} \cdot \frac{\partial Q_{off}^j}{\partial \tau_i}\Big] < 0 \\
&\Rightarrow \Big|\frac{dC_j}{dN^d} + \frac{d\tau_L}{dN^d}\Big| < \Big|\frac{dC_j}{dN^d} + \frac{d\tau_i}{dN^d}\Big| \cdot \frac{P(q_j) - \tau_i - C_j}{P(q_j) - \gamma_j} \\
&\Rightarrow (1 <)\frac{P(q_j) - \gamma_j}{P(q_j) - \tau_i - C_j} < \frac{\Big|\frac{dC_j}{dN^d} + \frac{d\tau_i}{dN^d}\Big|}{\Big|\frac{dC_j}{dN^d} + \frac{d\tau_L}{dN^d}\Big|}
\end{aligned} \tag{29}$$

If we set $\frac{dC_j}{dN^d} = 0$, the case when the on-line firms do Bertrand price competition and there is no market power to set the optimal C^j, above equation is changed to the proposition 1.

BIBLIOGRAPHY

BIBLIOGRAPHY

Armstrong, M. (2006), "Competition in two?sided markets." *The RAND Journal of Economics* 37(3): 668–691.

Bajari, P., C. L. Benkard, and J. Krainer. (2005), "House Prices and Consumer Welfare." *Journal of Urban Economics* 58: 474–487.

Bajari, P., and C. L. Benkard. (2005), "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach." *Journal of Political Economy*, 113 (6): 1239–76.

Bajari, P., Fruehwirth, J. C., Kim, K. I., and Timmins, C. (2012), "A rational expectations approach to hedonic price regressions with time-varying unobserved product attributes: The price of pollution." *The American Economic Review*, 102(5): 1898–1926.

Bajari, P., and M. Khan. (2005), "Estimating Housing Demand With An Application to Explaining Racial Segregation in Cities." *Journal of Business and Economic Statistics*, 23(1): 20–33

Baum-Snow, N. and Lutz, B.F. (2011), "School desegregation, school choice, and changes in residential location patterns by race." *American Economic Review*, 101(7): 3019–46.

Bayer, P., Fang, H. and McMillan, R. (2014), "Separate when equal? Racial inequality and residential segregation." *Journal of Urban Economics*, 82: 32–48.

Bayer, P., Ferreira, F. and McMillan, R. (2007), "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy*, 115(4): 588–638.

Bayer, P., Keohane, N. and Timmins, C. (2009) "Migration and hedonic valuation: The case of air quality." *Journal of Environmental Economics and Management*, 58(1): 1–14.

Bayer, P. and McMillan, R. (2012), "Tiebout sorting and neighborhood stratification." *Journal of Public Economics*, 96(11-12): 1129–1143.

Bayer, P., McMillan, R., Murphy, A., and Timmins, C. (2016). "A dynamic model of demand for houses and neighborhoods." *Econometrica*, 84(3): 893–942.

Bayer, P., N. Keohane, and C. Timmins. (2009) "Migration and Hedonic Valuation: The Case of Air Quality." *Journal of Environmental Economics and Management*, 58(1) 1–14.

Bento, A., Freedman, M. and Lang, C. (2015) "Who benefits from environmental regulation? Evidence from the Clean Air Act Amendments." *Review of Economics and Statistics*, 97(3): 610–622.

Bifulco, R. and Ladd, H. F. (2007), "School choice, racial segregation, and test-score gaps: Evidence from North Carolina's charter school program." *Journal of Policy Analysis and Management*, 26(1): 31–56.

Black, S. E. (1999), "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics*, 114(2): 577–99.

Black, S., and Stephen M. (2011), "Housing valuations of school performance." *Handbook of the Economics of Education*, 3: 485–519

Boustan, L.P. (2010), "Was postwar suburbanization "white flight"? Evidence from the black migration." *The Quarterly Journal of Economics*, 125(1): 417–443.

Boustan, L.P. (2012), "School desegregation and urban change: Evidence from city boundaries." *American Economic Journal: Applied Economics*, 4(1): 85–108.

Boustan, L.P. (2013) "Racial Residential Segregation in American Cities." *The Oxford Handbook of Urban Economics and Planning*

Caetano, G. and Maheshri, V. (2017) "School segregation and the identification of tipping behavior." *Journal of Public Economics*, 148: 115–135.

Card, D., Mas, A. and Rothstein, J. (2008), "Tipping and the Dynamics of Segregation." *The Quarterly Journal of Economics*, 123(1): 177–218.

Card, D. and Rothstein, J. (2007), "Racial segregation and the black-white test score gap." *Journal of Public Economics*, 91(11): 2158–2184.

Case, K.E. and Shiller, R.J. (1989), "The Efficiency of the Market for Single-Family Homes." *American Economic Review* 79 (1): 125–37.

Chay, K.Y. and Greenstone, M. (2005), "Does air quality matter? Evidence from the housing market." *Journal of Political Economy*, 113(2): 376–424.

Cho, S.I. (2014), "Pricing Capability of Retailers and the Effect of e-Commerce: the Case of the Book Retail Market." *KDI Policy Study*

Cho, S.I. (2015), "Study on the Economic Effect of Enhanced Fixed Book Pricing Scheme." *KDI Policy Study* 1–98.

Clapp, J.M. and Ross, S.L. (2004), "Schools and housing markets: An examination of school segregation and performance in connecticut." *The Economic Journal*, 114(499): 425–440.

Clapp, J.M., Nanda, A. and Ross, S.L. (2008), "Which school attributes matter? The influence of school district performance and demographic composition on property values." *Journal of urban Economics*, 63(2): 451–466.

De los Santos, B. and Wildenbeest, M.R. (2017), "E-book pricing and vertical restraints." *Quantitative Marketing and Economics* 15(2): 85–122.

Ellison, G. and Ellison, S.F. (2018), "Match quality, search, and the Internet market for used books" *National Bureau of Economic Research*, (No. w24197)

Epple, D. (1987), "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products." *Journal of Political Economy*, 95(1): 59–80.

Figlio, D. N., and Maurice E. L. (2004), "What's in a grade? School report cards and the housing market." *American Economic Review*, 94(3): 591–604

Fiva, J. H. and Kirkeben, L. J. (2011), "Information Shocks and the Dynamics of the Housing Market." *The Scandinavian Journal of Economics*, 113: 525-552.

Ghose, A., Smith, M.D. and Telang, R. (2006), "Internet exchanges for used books: An empirical analysis of product cannibalization and welfare impact." *Information systems research*, 17(1): 3–19.

Gibbons, S. and Machin, S. (2003), "Valuing English primary schools." *Journal of urban economics*, 53(2): 197–219.

Gibbons, S. and Machin, S. (2008), "Valuing school quality, better transport, and lower crime: evidence from house prices." *Oxford Review of Economic Policy*, 24(1): 99–119.

Gibbons, S., Machin, S. and Silva, O. (2013), "Valuing School Quality Using Boundary Discontinuity Regressions." *Journal of Urban Economics*, Volume 75: 15–28,

Grainger, C.A. (2012), "The distributional effects of pollution regulations: Do renters fully pay for cleaner air?" *Journal of Public Economics*, 96(9-10): 840–852.

Hamilton, T.L. and Phaneuf, D.J. (2015), "An integrated model of regional and local residential sorting with application to air quality." *Journal of Environmental Economics and Management*, 74: 71–93.

Imberman, S. A., and Lovenheim, M. F. (2016), "Does the market value value-added? Evidence from housing prices after a public release of school and teacher value-added." *Journal of Urban Economics*, Volume 91: 104–121,

Kane, T. J., Riegg, S. K. and Staiger, D. O. (2006), "School Quality, Neighborhoods, and Housing Prices." *American Law and Economics Review*, 8 (2): 183–212.

Klaiber, H. A., and Phaneuf, D. J. (2010), "Valuing open space in a residential sorting model of the Twin Cities." *Journal of Environmental Economics and Management*, 60(2): 57–77.

Machin, S. (2011). "Houses and schools: Valuation of school quality through the housing market." *Labour Economics*, 18(6): 723–729.

Machin, S. and Salvanes, K.G. (2016). "Valuing school quality via a school choice reform." *The Scandinavian Journal of Economics*, 118(1): 3–24.

Nickell, S. (1981). "Biases in dynamic models with fixed effects." *Econometrica: Journal of the Econometric Society*: 1417–1426.

Nguyen-Hoang, P. and Yinger, J. (2011), "The capitalization of school quality into house values: A review." *Journal of Housing Economics*, 20(1): 30–48.

Poort, J. and van Eijk, N. (2017), "Digital fixation: the law and economics of a fixed e-book price. " *International Journal of Cultural Policy* 23(4): 464–481.

Reardon, S.F. (2011), "The widening academic achievement gap between the rich and the poor: New evidence and possible explanations." *Whither Opportunity*: 91–116.

Reardon, S.F. (2016), "School segregation and racial academic achievement gaps." *RSF*

Reardon, S.F., Yun, J.T. and Eitle, T.M. (2000), "The changing structure of school segregation: Measurement and evidence of multiracial metropolitan-area school segregation, 1989-1995." *Demography*, 37(3): 351–364.

Reardon, S.F., Grewal, E.T., Kalogrides, D. and Greenberg, E. (2012), "Brown Fades: The End of Court?Ordered School Desegregation and the Resegregation of American Public Schools." *Journal of Policy Analysis and Management*, 31(4): 876–904.

Reback, R. (2005), "House prices and the provision of local public services: capitalization under school choice programs." *Journal of Urban Economics*, 57(2): 275–301.

Ries, J. and Somerville, T. (2010), "School quality and residential property values: evidence from Vancouver rezoning." *The Review of Economics and Statistics*, 92(4): 928–944.

Rochet, J.C. and Tirole, J. (2003), "Platform competition in two-sided markets." *Journal of the european economic association*, 1(4): 990–1029.

Rochet, J.C. and Tirole, J. (2006), "Two-sided markets: a progress report." *The RAND journal of economics*, 37(3): 645–667.

Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy*, 82(1): 34–55.

Rysman, M. (2009), "The economics of two-sided markets." *Journal of Economic Perspectives*, 82(1): 34–55.

Sieg, H., Smith, V.K., Banzhaf, H.S. and Walsh, R. (2004), "Estimating the general equilibrium benefits of large changes in spatially delineated public goods." *International Economic Review*, 45(4): 1047–1077.

Telser, L.G. (1960), "Why should manufacturers want fair trade?" *The journal of law and economics*, 3: 86–105.

Tra, C. I. (2010), "A discrete choice equilibrium approach to valuing large environmental changes." *Journal of Public Economics*, 94(1-2): 183–196.