Empirical Analysis on Influence of Education Performance on Housing Prices:
A Case of Daegu

教育성과가 아파트가격에 미치는 영향에 대한 실증적 분석:
대구시 사례를 중심으로

Yoon-San Kim (김윤산)**

< Abstract >

Most studies that empirically test whether education performance actually correlates with housing market outcomes focus on Seoul Metropolitan Area, which necessitates a study that sheds light on other regions where education fever is equally high. Therefore, this paper deals with a non-capital region housing market, the Daegu Metropolitan City. To study the correlation appropriately, employed as empirical methodologies are the Box-Cox power transformation model and the panel regression model. For the empirical analyses, a couple of indicators of education performance as well as other six variables are tested based on alternative econometric models of the price of condominium apartments. In doing so, this study elaborates to reveal the difference in the urban spatial structure between Seoul and Daegu by comparing several factors with each other. According to the analyses results, the most decisive factor that determines the condominium prices in this city is education performance of high schools, the number of students admitted to Seoul National University.

Meantime, this fact, incorporated with a study from the Korea Housing Institute, engenders concerns over the housing market of Daegu: the residential real estate in this city is already oversupplied; and the accumulation rate of vacancy stands at very high value; moreover, these problems will be hard to solve in a short period of time; synthesizing these results simply, inappropriate growth of the housing market, biased to a certain district, would be accelerated so that a more prudent policy considering the unique site-specific characteristics of this city should be served.

Keywords: Real Estate, Educational Fever, Housing, Education Performance, Daegu, Seoul National University, Hedonic Price, Panel Data, the Box-Cox Transformation

* 이 논문은 2011년도 KDI 국제정책대학원 석사학위 논문을 토대로 수정한 것으로, 논문심사 중 존중하는 저자와 가르침을 아끼지 않으신 이명의 세분 심사위원님에게 감사드립니다.
** KDI 국제정책대학원 석사 졸업, 한국도로공사 제작 중, yoon.s.kim1972@gmail.com
I. INTRODUCTION

When people talk about the Korean educational enthusiasm, the modifier “excessive” follows, implying that Korean households put special emphasis on education of their children in making consumption and location decisions. The reason for that can be explained in a well-known cultural expression that personal as well as societal success in Korea depends on academic achievement. Notwithstanding the issue of whether such enthusiasm distorts resource allocation in Korean society, it works as a strong demand shift factor in local housing markets where middle and high schools are deemed to have good performance track records. Specifically, it is argued that, in consequence of the educational fever, those locations where high schools have higher admission rates to top-class universities tend to experience too much growth in population (Rhee 2007). As empirical evidence for such demand-side effect, the Survey on Actual Condition of Housing Finance, an annual report announced by the Kookmin Bank, reports that Korean households with middle- or high-school age children put the school district as a top priority when purchasing their residence. 1).

Meantime, a number of previous studies have researched the relationship between education performance and housing market outcomes. This study also aims to address the issue of the previous studies in two ways: first, although most of the studies in this vein focus on Seoul Metropolitan Area (hereafter Seoul), this paper deals with a non-capital region housing market, in particular the Daegu Metropolitan City (hereafter Daegu) where education fever is equally high; this is from having an eye on a media report stating that Susong district in Daegu is known for such a prestigious high-school district so that it is often regarded as eighth-high-school (3) district of Seoul; second, a couple of indicators of education performance, among all eight independent variables, are tested based on alternative econometric models of local housing prices, specifically the price of condominium apartments; the number of student admitted to Seoul National University and the admission rate to four-year-course colleges are shown to be outstanding in terms of their impacts on condominium prices in Daegu. With all eight explanatory variables, tested as empirical methodologies are the Box-Cox power transformation model and the panel regression model,

1) According to the survey hold in 2009, among several major conditions for buying a home, 35.4% of home buyers whose ages are in their 40s took education conditions into account as the first priority. Source: the Survey on Actual Condition of Housing Finance (Kookmin Bank 2009)
2) Source: http://biz.heraldm.com/common/Detail.jsp?newsMId=20090416000290
3) The eighth-high-school district in the Seoul Metropolis is famous the for high admission rate to top-class universities in Korea and for the nation-highest house price.
Unlike the conventional linear or non-linear regression model in the hedonic literature, in order to have efficient parameter estimates. In doing so, pooled time-series and cross-section data for the price of 60 condominium apartments complexes (hereafter condominiums) are compiled and used for empirical analyses.

Through the analyses, astoundingly revealed by the result of this paper is that the education performance of high schools in Daegu is the most decisive factor to determine the condominium prices. Moreover, according to the result of a study from the Korea Housing Institute, the residential real estate in this city is oversupplied and the accumulation rate of vacancy stands at very high value. Such being the case, this study warns that these problems will be hard to solve in a short time period. Synthesizing these results simply, there could be a concern over the fact that inappropriate growth of the housing market, biased to a certain district, would be accelerated (for more detail, see Section V).

Meanwhile, the rest of the paper consists of four sections. After looking over the recent status of Daegu Metropolitan City in Section II, a brief explanation on the hedonic price model, literature review and opinions are provided in Section III. Then, Section IV affords the empirical analyses on the influence of education performance on the housing prices in Daegu and then finally, in Section V, this paper is concluded with discussion on distinct characteristics of the housing market in Daegu and a policy implication.

II. OVERVIEW ON DAEGU METROPOLITAN CITY

Daegu Metropolitan City is located in south-eastern parts of Korea. Its GRDP (Gross Regional Domestics Product) was KRW 32,714 billion at the current prices as of 2008, with holding the proportion of 3.2 percent of total GRDP in this country, following Seoul, Busan, Incheon and Ulsan in order. Introducing recent socioeconomic situations of this city, several factors related to the issue can be introduced: the area of this city was 884.07km² and 798.47km² out of total area, roughly 90 percent, was used for urban areas; among several dwelling patterns, 50.8 percent of total households in Daegu was residing in condominiums; the population stood at 2,509,187 as of 2009. Meanwhile, when the population density of Daegu was compared with that of Seoul, great difference could be detected. As of 2010, the population density of Daegu was 2,750 thousands persons/km², while that of Seoul was 16,593 thousands persons/km²; the population density of Seoul was almost six times greater than that of Daegu. Table 1.2 shows the expected population growth rate of two cities. Following the table, the expected decrease in
population of Daegu is far greater than that of Seoul.

<table>
<thead>
<tr>
<th>City</th>
<th>Yr 2015-2020</th>
<th>Yr 2020-2025</th>
<th>Yr 2025-2030</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seoul</td>
<td>-0.22</td>
<td>-0.41</td>
<td>-0.59</td>
</tr>
<tr>
<td>Daegu</td>
<td>-0.70</td>
<td>-0.78</td>
<td>-0.83</td>
</tr>
</tbody>
</table>


As of 2009, the number of high school in Daegu was 91 and, among them, only 60 high schools sent more than one student to Seoul National University during the time period from 2005 to 2009. These high schools had total 966 students who were admitted to the top-class university and 465 students, or 48.1%, came from Suseong district.

As for the public transportation, only two subway lines are available and the third subway line is under construction. The subway line 1 and 2, however, does not run the whole area in Daegu. But in the case of the bus, 90 bus lines are on service: three rapid lines, two belt lines, 58 trunk lines and 27 branch lines so that Daegu citizens highly depend on buses. This implies that the driving condition in the respect of road congestion\(^4\) in Daegu is not so much terrible when compared with that in Seoul. The description on this matter is following:

1) traffic congestion cost
   - Seoul: \(\text{KRW} ~ 7.03\text{ billion}\)
   - Daegu: \(\text{KRW} ~ 1.32\text{ billion}\)

2) average driving speed\(^5\) in the downtown
   - Seoul: 15.7 km/hr
   - Daegu: 26.6 km/hr.

Not only that, the better road condition in Daegu can be detected by observing the rate of public transportation use: as of 2006, 62.3% for Seoul and 38.6% for Daegu, respectively. As noted by the theory of the monocentric city model and Bertaud (2004)\(^6\), the road traffic condition in a city is a

\(^4\) The traffic congestion cost is a social cost incurred by the increase of the traffic demand. This consists of the car operating cost and the time value cost. Meantime, the traffic congestion cost above is calculated as of 2007 (Source: [http://car.daegu.go.kr/db/all/all_01_3.jsp](http://car.daegu.go.kr/db/all/all_01_3.jsp)).

\(^5\) The average driving speed in the downtown is the value averaged over five-year time period from 2002 to 2006 (Source: [http://www.index.go.kr/egame/stts/jsp/potal/stts/PO_STTS_OdxMain.jsp?dx_cd=1247&bbs=INDX_001](http://www.index.go.kr/egame/stts/jsp/potal/stts/PO_STTS_OdxMain.jsp?dx_cd=1247&bbs=INDX_001)).
determinant factor for the urban spatial structure. Therefore, the road traffic conditions in Daegu is much better than those in Seoul and this fact is considered as a structural difference in the urban spatial structure when compared with Seoul.

The reason that this paper compare the status of population and road condition in Daegu with those of Seoul is to identify the structural difference between the two cities. By doing so, this paper empirically analyzes the most decisive factor, different from the case of Seoul, that influences the condominium prices and problems of the housing market in Daegu.

III. LITERATURE REVIEW

To examine which and how much internal and external factors contribute to pricing the condominium in Daegu, this study employs the hedonic price model. Through this hedonic pricing procedure, this study attempts to define the most decisive factor that influences the condominium prices. Meanwhile, the term, hedonic, originated from the ancient Greek hedonistic philosophies which formed the foundation of utilitarianism (Lee 2008; DiPasquale et. al 1996) and the term, the hedonic price, is called as implicit price or characteristic price. Because housing is a heterogeneous commodity, its price is determined by characteristics a housing itself contains. Meanwhile, after Ridker (1967) initiatively introduced the hedonic price concept as an analysis model to identify that air pollution affected property values by regressing median census tract property values on a measure of sulfate air pollution, Rosen (1974) provided the foundation for the hedonic pricing applications in his paper that researched the demand for public schooling and calculated cross-price elasticity with a two-stage hedonic technique. Since then, the hedonic price model has extended its study fields and nowadays this model is widely used to estimate the implicit price not only in the housing market studies but also in a variety of study fields such as the studies on the automobile market, the IT and the software market, the telecommunications services market and etc.

Based upon the hedonic price model, lots of previous foreign and Korean researches revealed the relationship between education and the housing price. In those previous researches, various analysis models and a variety of variables were introduced. The followings are brief reviews on them: first of all, a research found that individuals, when purchasing houses, did appear to consider the current test performance of students in the local school. More interesting findings from this research was the necessity of including measures of

neighborhood quality in addition to the school characteristics just to prevent the coefficient estimates for the school characteristics from being biased (Thomas and Zabel 1997); Brasington (1999), meanwhile, found that parents appeared to choose school systems based on peer group effects. Through the result, he suggested not employing value-added measures such as the graduation rate and teacher education levels because they were not consistently positively related to housing prices (see Hayes and Taylor (1996), Downes and Zabel (2002) and Brasington and Haurin (2006)); another supporting research by Bogart et al. (2000) found that disrupting neighborhood schools reduced house value by 9.9 percent, ceteris paribus; also, the proficiency test score was found as the most consistently valued measure of school quality (Brasington and Haurin 2006). An increase in the test score by one standard deviation raised house prices by 7.1 percent, other things being equal; interestingly, Chiodo et al. (2010) and Hanushek (1986, 1997) found that school inputs such as per-pupil spending had no apparent impact on student achievement and were therefore inappropriate as measures of school quality, although various studies in the traditional hedonic analysis had used so-called input-based measures of education quality. In his study, he also revealed that a house associated with higher-quality schools commanded a much higher price premium; Jung (2006) found that, in Seoul, educational variables such as the admission rate into Seoul National University and four-year-course colleges, the rate of private institutes and the eighth-high-school district affected condominium prices; moreover, Eom et al. (2006) found that the average condominium prices in region A was 64.0 percent higher than in region B, if the education system of region A was twice as good as that of region B. Interestingly, their study revealed that regional characteristics were more important than individual ones in determining condominium prices.

Encapsulating lessons from the previous researches, and considering that the hedonic price is an implicit price determined by the characteristics of housing, this study needs to inevitably include factors representing well the characteristics of condominiums: first, in the respect of the characteristics of education factors, it is advised to employ output-characterized ones; for more information, see Chiodo et al. (2010) and Hanushek (1986, 1997). This implies that when parents decide to move just for the children’s education, they generally take the school performance, output-characterized factors, into account. In other words, when it comes to education factors, the value of real estate is decided by apparent school performance not by inputted educational quality because such input-characterized value-added measures cannot stand for accurate school performance; second, as noted by Eom et al. (2006), regional (here in this paper, external) characteristics are more effective than individual (here in this paper, internal) ones in determining condominium prices. Not only that, this study endeavors to mainly focus not on the price of an individual apartment but on the average
price of a condominium. This is because this study attempts to specify whether educational factors, among external ones, most strongly affect the price; for these reasons, according to the lessons from previous researches, this paper restrictively employs two internal-characteristic factors and the output-characterized factors especially for education performance are selected.

IV. EMPIRICAL ANALYSIS ON INFLUENCE OF EDUCATION PERFORMANCE ON HOUSING PRICES IN DAEGU

1. Data Description

As noted previously, the purpose of this study is to empirically analyze the fact that how much education performance influence the prices of condominiums in Daegu. Therefore, this paper attempts to clarify the degree of its influence and then identify the structural problems of the housing market in this city. Finally, a suggestion on the housing market policy is discussed later on. On that purpose, this paper employed total nine variables as in Table 4.1.

To specifically reiterate the purpose of this study, it is to clarify that which external factor most strongly influences the prices of condominiums on average, not the prices of each individual apartment. Therefore, as advised by Eom et al. (2006), this study restrictively employs two internal-characterized variables, while eight external-characterized variables that represent the living environment well are employed.

<Table 4.1> Variable Description

<table>
<thead>
<tr>
<th>Name</th>
<th>Characteristic</th>
<th>Explanatory Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHH</td>
<td>Internal</td>
<td>The number of household in each condominium</td>
</tr>
<tr>
<td>AGE</td>
<td></td>
<td>The age (yrs) of each condominium</td>
</tr>
<tr>
<td>SNU</td>
<td>External</td>
<td>The number of student admitted to Seoul National University</td>
</tr>
<tr>
<td>CLG</td>
<td>- Education</td>
<td>The admission rate (%) to four-year-course colleges</td>
</tr>
<tr>
<td>SBW</td>
<td>External</td>
<td>The number of subway station nearby each condominium</td>
</tr>
<tr>
<td>BUS</td>
<td>- Transportation</td>
<td>The number of bus run nearby each condominium</td>
</tr>
<tr>
<td>PRK</td>
<td>External</td>
<td>The number of city park nearby each condominium</td>
</tr>
<tr>
<td>STR</td>
<td>- Living</td>
<td>The number of department store and discount store in the vicinity of each condominium</td>
</tr>
<tr>
<td></td>
<td>Condition</td>
<td></td>
</tr>
</tbody>
</table>

* The number of observations: 60, respectively
To decide the dependent variable, *Price*, 60 condominiums\(^7\) nearby 60 high schools in which more than one student was admitted to Seoul National University were examined first. In doing so, the condominiums aged more than five years\(^8\) as of 2009 were targeted. They were surveyed based on the website, the Daegu Life Geographic Service. The data for the prices, meanwhile, from 2005 to 2009 were obtained from the website, the Budongsan114.

In the mean time, this paper used two internal-characterized variables which represent the traits of each individual condominium as follows: the number of household in each 60 condominium was surveyed and then the age of a condominium was calculated by subtracting the year building completed from every December from 2005 to 2009.

Meantime, a cluster of data for the number of student\(^9\) admitted to Seoul National University of the 60 high schools was obtained from the newspaper website of the DongA Ilbo\(^10\). Besides, the admission rate to four-year-course colleges of 60 high schools was obtained from the website, the Korea Education and Research Information Service.

Moreover, another variables that represent the public transportation conditions were employed: the number of subway station and the number of bus run nearby condominiums. This was from taking notice of a common idea saying that people prefer condominiums located at adjacent area to subway or bus stations. Meanwhile, the subway stations and the bus stops within a radius of about 667 meters of the 60 condominiums were surveyed based on the website, the Daegu Life Geographic Service; the criterion, a radius of about 667 meters, is ascribed to the fact that a

---

7) When the number of high school that sent more than one student to Seoul National University during the time period from 2005 to 2009 was examined, 60 schools out of total 91 schools in Daegu were included. Furthermore, the selected condominiums in this paper are the nearest ones from each 60 high school; the 60 condominiums are decided by being matched to the 60 high schools one by one.

8) This study is based on the time period from 2005 to 2009. This is because the time period of main education-factor data, the number of student admitted to Seoul National University, is limited to that period; those data for that period are only ones that the author could get from the Internet. Therefore, more than 5-year-old complexes were targeted and thus newly-built ones were excluded just to prevent the effect of their name value from improperly influencing the analysis result.

9) The reason this paper takes the number, not the rate, of student admitted to Seoul National University is from taking notice of an idea saying that parents, when they select a high school for their children, consider the education performance of a high school; in doing so, what they generally take into account is not the admission rate to top-class universities but the number of student who got the admission. Put another way, the decisive factor for home buyers’ purchasing decision is not the admission rate to top-class universities but the number of student admitted. In this regard, this paper takes the number of student admitted to the top-class university, Seoul National University, as an educational factor.

human being can walk four kilometers for an hour on average. Besides, this study endeavored to arrange the second transportation variable, the number of bus run. This variable was also from the website just above, and all bus stops nearby the 60 condominiums were examined and then a cluster of data was processed\textsuperscript{11}). The buses examined in this study ran and stopped at bus stops from 301 to 2,279 times a day.

Meantime, the other factors for living conditions were employed. The number of city park and the number of department store and large discount store are those. For the former variable, the city parks whose area were larger than 10,000 m\textsuperscript{2}, middle-sized ones, were observed. All middle-sized city parks also within 10 minutes on foot, a radius of about 667 meters, from each condominium were counted. For the latter variable, total eight brands\textsuperscript{12}) of stores within a radius of about one km\textsuperscript{13}) from each condominium were all counted.

Interestingly, one characteristic of this paper is that the author endeavors to identify whether individual heterogeneity in cross-sectional and time-series data change the result of the empirical analysis. Such being the case, the data transformed by the Box-Cox power transformation is linearly regressed first; and here it is tested to see whether assumptions of normality and homoskedasticity are improved. After the multiple linear regression, the panel data analysis is performed to reduce biases resulting from aggregation over identities. By doing so, this study examines whether the decisive factor for the condominium prices of Daegu is changed. The analysis flow is as shown in Figure 4.1.

\textsuperscript{11}) Firstly, all bus stops within a radius of roughly 667 meters of the 60 condominiums were examined and then all bus numbers stop at each bus stops were obtained; secondly, the intervals of all bus runs were observed from the website, the Daegu Metropolitan City Bus Information, and then the number of bus run was calculated. The way of the calculation is in the following: assuming that buses run for 15 hours a day from 7a.m. to 10p.m., 15 hours multiplied by 60 minutes equals 900 minutes. Then, dividing the 900 minutes by the interval (in minutes) of each bus, the number of bus run per day could be obtained; third, by totaling up the number of each bus run at a bus stop, the total number of bus run at a bus stop can be gained; lastly, again by adding up the total number of bus run at bus stops nearby the condominium, the number of bus run for following empirical analyses was completed finally.

\textsuperscript{12}) The eight brands of stores are as follows: Lotte department store, Daegu department store, Debec plaza, DongA department store, DongA shopping, Homeplus, E-mart and Lotte mart.

\textsuperscript{13}) The reason one km is determined as the radius for selecting stores is that people, when going shopping, usually use their cars and therefore the range people move around gets wider. Specifically, supposing that when people, in the middle of a city, drive their cars at the speed of 20 km/hr, and also supposing that they willingly spend three minutes in a car for their usual shopping, they can reach an one-km-distant discount store. In this regard, this paper decides one km as its range for selecting department stores and discount stores.
Flow of Empirical Analyses

Model Specification
The Box-Cox power transformation

Multiple Regression (1)
Ordinary least square
with NOT transformed data

Estimation Results Comparison (1)
To see how much the transformation
improves the goodness of fit

Multiple Regression (2)
Ordinary least square
with power transformed data

Panel Data Analysis
Another model specification and
an empirical analysis

Estimation Results Comparison (2)
To see the difference between OLS (transformed) and Panel Data Analysis

(Figure 4.1) Flow of Empirical Analyses

2. Model Specification

1) Basic model

This paper specifies the basic model of the hedonic price function with the data of condominium prices and eight internal- and external-characteristic factors as in the following:

\[ P_{it} = f(S_{it}, E_{it}, T_{it}, L_{it}) \]  \hspace{1cm} (4.1)

where \( P_{it} \) is the price of \( i^{th} \) condominium at time \( t \), \( S_{it} \) is the internal structural traits of \( i^{th} \) condominium at time \( t \), \( E_{it} \) is the external educational factors of \( i^{th} \) condominium at time \( t \), \( T_{it} \) is the external transportation factors of \( i^{th} \) condominium at time \( t \) and \( L_{it} \) is the also external factors for living conditions of \( i^{th} \) condominium at time \( t \).

In the mean time, detailed explanations on these variables are well elucidated in the previous subsection.

2) Model Specification by the Box-Cox Power Transformation

When empirical analyses are performed not only in the econometric field but also in many other fields, researchers usually confront the matter of model specification. This is probably because, specifically in hedonic price studies, a potentially serious source of bias is associated with functional form misspecification, as we were noted by Linnenman (1980). In this regard,
data transformation techniques are commonly used in empirical analyses to improve the
goodness of fit of a model; by doing so, researchers endeavor to meet assumptions such as
normality, homoskedasticity and linearity etc.; Hossain (2011) cited Tukey (1957) arguing that
"a transformation of variables may lead to a more nearly linear model, may stabilize the
error variance, and/or may lead to a model for which a symmetrically, perhaps normally
distributed error term is acceptable". Those transformation techniques, meanwhile, can be
categorized into several forms at large as follows: linear, quadratic, semi-log and log-log model
and the Box-Cox power transformation.

In the respect of transforming data to serve an appropriate function in a quantitative
analysis, the purpose of such transformation techniques is all the same. The Box-Cox power
transformation\footnote{14}, nonetheless, affords several advantages in its use as Sakia (1992) maintained
that the Box-Cox transformation has found more practical utility in the empirical
determination of functional relationships in a variety of fields, especially in econometrics.
Followings are those several strengths: 1) Osborne (2010) argued that most common
transformations reduce positive skew but may exacerbate negative skew unless the variable is
reflected prior to transformation. Box-Cox eliminates the need of this; 2) the Box-Cox
transformation is able to present an approximately normally distributed error term which is
prerequisite for conducting hypothesis tests (Miner 1982); 3) Peltier et al. (1998) employed
the Box-Cox transformation algorithm to test hypotheses in their study and found that
transformed data is very useful to meet the assumption of ANOVA\footnote{15}. Such being the case,
the Box-Cox power transformation has been introduced in a variety of study fields and the
field of hedonic price study cannot be an exception.

To begin with, Linneenman (1980) suggested a systematic statistical methodology for the
analysis of the urban housing market with employing the Box-Cox power transformation and
in his study he used the functional form in which both dependent and independent variables
are all transformed as follows\footnote{16}:

\footnote{14} Box and Cox (1964) originally envisioned this transformation as a panacea for simultaneously correcting
normality, linearity and homoskedasticity. In the procedure, the maximum likelihood estimation (MLE) and
Bayesian methods are suggested for the parameter estimation. Between them, many previous studies used
the former and therefore this paper also employed the MLE to obtain the power parameter.

\footnote{15} The assumption of ANOVA is that the variance in a cell is not significantly different from that of other
cells, which indicates homoskedasticity of variance.

\footnote{16} After Turkey (1957) had introduced a family of power transformations, some alternative versions of the
Box-Cox transformation followed. Those alternative versions, designed to best fit to certain cases, are to
overcome the limitation of the initial one in its use; for more information, see Sakia (1992); a brief
explanation on this limitation will be discussed in this paper later on. This implies the fact that this
\[
\frac{\lambda^2 - 1}{\lambda} = b_0 + \sum_{j=1}^{k} b_j n_j + \sum_{j=k+1}^{n} b_j s_j + e_i
\] (4.2)

where \( n_j = ((N_i j)^{\gamma - 1}) v_j^{-1} \), \( s_j = ((S_i j)^{\gamma - 1}) v_j^{-1} \), \( e_i \) and \( u_i \sim N(0, \sigma^2) \), \( \lambda \) is the power transformation factor for the dependent variable while \( v_j \) is the power transformation factor for the \( j^{th} \) locational trait.

On the other hand, Goodman (1978) introduced another form of the Box-Cox transformation; only dependent variable was transformed. In his paper, he extended hedonic price analysis to the formation of housing price indices measuring variation within a metropolitan area. The Box-Cox procedure is applied to the analysis like the form below:

\[
\frac{1}{\lambda_i} (P_{\lambda_i} - 1) = \beta_{0i} + \sum_{i}^{k} \beta_{ij} C_{ij} + e_j
\] (4.3)

The equation below, meanwhile, is a general form of the power parameter of the Box-Cox power transformation:

\[
y_i^{(\lambda)} = \begin{cases} 
(y_i^{\lambda} - 1)/\lambda; & \lambda \neq 0 \\
\log(y_i); & \lambda = 0 \\
y_i; & \lambda = 1 
\end{cases}
\] (4.4)

and for unknown parameter \( \lambda \),

\[
y^{(\lambda)} = (y^{(\lambda)}_1, y^{(\lambda)}_2, \ldots, y^{(\lambda)}_n)^' = X\theta + \varepsilon ;
\]
where \( X \) is a matrix of known constants, \( \theta \) is a vector of unknown parameters associated with the transformed values and \( \varepsilon \sim N(0, \sigma^2 I_n) \) is a vector of random errors.

Following the discussions on the Box-Cox power transformation above, this paper employs the only-dependent-variable transformed procedure and the parameter is estimated by maximum likelihood (ML) method. Accordingly, when Equation (4.1) and (4.2) are combined, the hedonic price function for the condominiums in this study can be calculated as

transformational technique offers a room for its various use as well as it requires meticulous care for its use from its operators. Put another way, unlike the Linnenman’s case above, a researcher can employ the type where only the dependent variable is transformed. Otherwise, the more important point here is that the researcher needs to test whether the transformed data satisfies assumptions such as normality and homoskedasticity.
\[
\frac{(P_i^{\lambda} - 1)}{\lambda} = \beta_0 + \beta_1 NHI_{it} + \beta_2 AGE_{it} + \beta_3 SNU_{it} + \beta_4 CLG_{it} + \beta_5 BUS_{it} + \beta_6 SBW_{it} + \beta_7 PRK_{it} + \beta_8 STR_{it} + \epsilon
\]

(4.5)

where \( P_i \) is the price, \( NHI_{it} \) is the number of household, \( AGE_{it} \) is the age of condominiums, \( SNU_{it} \) is the number of student admitted into Seoul National University, \( CLG_{it} \) is the admission rate into 4-year-course colleges, \( BUS_{it} \) is the number of bus run, \( SBW_{it} \) is the number of subway station, \( PRK_{it} \) is the number of park, and \( STR_{it} \) is the number of department store and discount store, \( i \) indicates \( i^{th} \) condominium; \{\( x_1, ..., x_8, P_i \): \( i = 1, 2, ..., 60 \)\} for the year 2005 to the year 2009 respectively; \( \epsilon \) is the error term for each year (\( \epsilon \sim \text{N}(0, \sigma^2) \)); \( \beta_2, ..., \beta_8 \) are estimators of the multiple linear regression.

To draw the most appropriate hedonic price function\(^{17}\), the transformation parameter \( \lambda \) must be specified so that this study, as noted above, uses maximum likelihood estimation (MLE). The value of the best fitting parameter, \( \lambda \), for the year 2009 turned out to be -0.561 when the value of log-likelihood was -418.55. This result is well described in Figure 4.2. Therefore, only the dependent variable is transformed with \( \lambda \) and then regressed with independent variables in the next subsection.

\[\text{Box-Cox Profile Log-likelihood}\]

\(<Figure 4.2>\) Box-Cox Profile Log-likelihood for the year 2009

\(^{17}\) The Box-Cox transformation has limitations in its use: 1) the transformation is valid only for \( y > 0 \); 2) positively skewed variables are easily dealt with via this transformation but negatively skewed variables should be reversed; 3) basically, data transformation can introduce complexity into the interpretation of the results (as they change the nature of the variable, and any \( \lambda < 0 \) has the effect of reversing the order of the data, and thus care should be taken when interpreting results); for more information, see Box and Cox (1964), Sakia (1992), Osborne (2010) and Hosain (2011). Meanwhile, the independent variables in this study include several discontinuous ones whose values of some identities are 0. Therefore, this study uses the only-dependent-variable-transformed Box-Cox procedure; in fact, when this study made an attempt to introduce the form that both dependent and independent variables are all transformed by substituting 0 value with roughly 0 value (e.g., 1-50), the estimation result shows non-normality and heteroskedasticity; this procedure is not included here. Nonetheless, after that procedure, this study tested normality and homoskedasticity.
Table 4.2: Summary Statistics of Transformed Data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Year 2005</th>
<th>Year 2006</th>
<th>Year 2007</th>
<th>Year 2008</th>
<th>Year 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Mean</td>
</tr>
<tr>
<td>Price</td>
<td>5.4517</td>
<td>0.1775</td>
<td>5.1339</td>
<td>0.1486</td>
<td>5.5237</td>
</tr>
<tr>
<td>NHH</td>
<td>648.12</td>
<td>406.60</td>
<td>648.12</td>
<td>406.60</td>
<td>648.12</td>
</tr>
<tr>
<td>SNU</td>
<td>3.6167</td>
<td>3.8050</td>
<td>3.4833</td>
<td>4.2725</td>
<td>3.55</td>
</tr>
<tr>
<td>CLG</td>
<td>58.51</td>
<td>11.425</td>
<td>58.51</td>
<td>11.425</td>
<td>58.51</td>
</tr>
<tr>
<td>BUS</td>
<td>1062</td>
<td>416.59</td>
<td>1062</td>
<td>416.59</td>
<td>1062</td>
</tr>
<tr>
<td>SBW</td>
<td>0.2833</td>
<td>0.5849</td>
<td>0.6167</td>
<td>0.6911</td>
<td>0.6167</td>
</tr>
<tr>
<td>PRK</td>
<td>1.05</td>
<td>1.0803</td>
<td>1.05</td>
<td>1.0803</td>
<td>1.05</td>
</tr>
<tr>
<td>STR</td>
<td>0.6833</td>
<td>0.6507</td>
<td>0.7167</td>
<td>0.6911</td>
<td>0.7833</td>
</tr>
</tbody>
</table>

* The number of observations: 60, respectively

Meanwhile, Table 4.2 above is the summary statistics of transformed data and the results of MLE for the data for the year 2005 to 2009 are included in Table 4.3.

As noticed previously, this study clarifies whether assumptions of normality and homoskedasticity are improved by the Box-Cox power transformation; also in the following subsection, the estimation results of the transformed data only and the comparison of the normality and homoskedasticity tests are served; the estimation results of the untransformed data are excluded in this paper.

3. Estimation Results of Multiple Linear Regression

After the condominium prices of the dependent variable for each year are transformed by the parameter $\lambda$, they are regressed with eight independent variables for each year and thereafter the statistically significant factors influencing the condominium prices are examined.

In this subsection, to examine the statistically significant factors, the ordinary least square (OLS) is employed first and then the panel data analysis is introduced in the next subsection. Table 4.3 below shows the estimation results of the multiple linear regression.

According to the MLR results, statistically significant variables which influence the condominium prices in Daegu are $AGE$, $SNU$ and, between the two, $SNU$ is only the statistically significant variable among six external-characteristic factors. Moreover, this paper can identify the fact that the only variable that turns out to be statistically significant during
whole five-year period is also SNU. The reason, this variable is only statistically significant, is probably attributed to the structural difference of Daegu when compared with Seoul, as anticipated previously. A more in-depth discussion on this matter and the interpretation on estimation results is continued in the next section.

\textit{Table 4.3} Results of Multiple Linear Regression (Enter method (all variables included))

<table>
<thead>
<tr>
<th>Variable</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta$</td>
<td>$b$</td>
<td>$t$</td>
<td>$\beta$</td>
<td>$b$</td>
</tr>
<tr>
<td>NHH</td>
<td>0.0001</td>
<td>0.1528</td>
<td>1.41</td>
<td>0.0001</td>
<td>0.1752</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.007***</td>
<td>-0.2795</td>
<td>-2.57</td>
<td>-0.0038</td>
<td>-0.1639</td>
</tr>
<tr>
<td>SNU</td>
<td>0.024***</td>
<td>0.5162</td>
<td>4.27</td>
<td>0.019***</td>
<td>0.3579</td>
</tr>
<tr>
<td>CLG</td>
<td>0.0063</td>
<td>0.0221</td>
<td>0.20</td>
<td>0.0061</td>
<td>0.0786</td>
</tr>
<tr>
<td>BUS</td>
<td>0.0001</td>
<td>0.1995</td>
<td>1.89</td>
<td>0.0002</td>
<td>0.0644</td>
</tr>
<tr>
<td>SBW</td>
<td>0.0085</td>
<td>0.0281</td>
<td>0.26</td>
<td>0.0217</td>
<td>0.1010</td>
</tr>
<tr>
<td>PK</td>
<td>0.0191</td>
<td>0.1162</td>
<td>1.07</td>
<td>0.0019</td>
<td>0.0141</td>
</tr>
<tr>
<td>STR</td>
<td>-0.0999</td>
<td>-0.0362</td>
<td>-0.34</td>
<td>-0.014</td>
<td>-0.0649</td>
</tr>
<tr>
<td>Constant</td>
<td>5.28***</td>
<td>-43.4</td>
<td>4.98***</td>
<td>-46.9</td>
<td>5.36***</td>
</tr>
</tbody>
</table>

MLE

|          | $\lambda$: | -0.0814 | | $\lambda$: | -0.1018 | | $\lambda$: | -0.0790 | | $\lambda$: | -0.4691 | | $\lambda$: | -0.5610 | |

|          | Log likelihood | | Log likelihood | | Log likelihood | | Log likelihood | | Log likelihood |

Log likelihood: -427.8961

Log likelihood: -432.0343

Log likelihood: -426.0386

Log likelihood: -426.4591

Log likelihood: -418.5535

**: significant at p<0.05, ***: significant at p<0.01, $\beta$: unstandardized coefficient, b: standardized coefficient

\textit{Table 4.4} Results of Multiple Linear Regression (Backward method (optimal model))

<table>
<thead>
<tr>
<th>Variable</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta$</td>
<td>$b$</td>
<td>$t$</td>
<td>$\beta$</td>
<td>$b$</td>
</tr>
<tr>
<td>AGE</td>
<td>-0.007***</td>
<td>-0.2772</td>
<td>-2.66</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SNU</td>
<td>0.024***</td>
<td>0.5191</td>
<td>4.98</td>
<td>0.214***</td>
<td>0.6157</td>
</tr>
<tr>
<td>Constant</td>
<td>4.45***</td>
<td>-127.4</td>
<td>5.05***</td>
<td>-256.4</td>
<td>5.423***</td>
</tr>
</tbody>
</table>

|          | $\lambda$: | -0.0814 | | $\lambda$: | -0.1018 | | $\lambda$: | -0.0790 | | $\lambda$: | -0.4691 | | $\lambda$: | -0.5610 | |

|          | Log likelihood | | Log likelihood | | Log likelihood | | Log likelihood | | Log likelihood |

Log likelihood: -427.8961

Log likelihood: -432.0343

Log likelihood: -426.0386

Log likelihood: -426.4591

Log likelihood: -418.5535

**: significant at p<0.05, ***: significant at p<0.01

18) The variance inflation factor (VIF) is an indicator of the multicollinearity diagnostic. If its value is less than seven, it is referred to as that the independent variable changes under the condition of statistical significanc
Meanwhile, the tests on normality and homoskedasticity are performed to make this functional form specification as a more robust one; the former is examined by the Komorogov-Smirnov test and the latter is tested by the Cook-Weisberg test as below.

\[ \text{Table 4.5} \] Results of the Komorogov-Smirnov test and the Cook-Weisberg test

<table>
<thead>
<tr>
<th>Year</th>
<th>Normality Test</th>
<th>Homoskedasticity Test(^19)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Untransformed Data</td>
<td>Transformed Data</td>
</tr>
<tr>
<td></td>
<td>Skewness: 0.186</td>
<td>Skewness: 0.887</td>
</tr>
<tr>
<td></td>
<td>Kurtosis: 0.808</td>
<td>Kurtosis: 0.544</td>
</tr>
<tr>
<td>2005</td>
<td>( P &gt; \chi^2: 0.3888 )</td>
<td>( P &gt; \chi^2: 0.8213 )</td>
</tr>
<tr>
<td></td>
<td>Skewness: 0.054</td>
<td>Skewness: 0.833</td>
</tr>
<tr>
<td></td>
<td>Kurtosis: 0.274</td>
<td>Kurtosis: 0.437</td>
</tr>
<tr>
<td></td>
<td>( P &gt; \chi^2: 0.0881 )</td>
<td>( P &gt; \chi^2: 0.7169 )</td>
</tr>
<tr>
<td>2006</td>
<td>Skewness: 0.010</td>
<td>Skewness: 0.526</td>
</tr>
<tr>
<td></td>
<td>Kurtosis: 0.007</td>
<td>Kurtosis: 0.955</td>
</tr>
<tr>
<td></td>
<td>( P &gt; \chi^2: 0.0036 )</td>
<td>( P &gt; \chi^2: 0.8140 )</td>
</tr>
<tr>
<td>2007</td>
<td>Skewness: 0.008</td>
<td>Skewness: 0.816</td>
</tr>
<tr>
<td></td>
<td>Kurtosis: 0.001</td>
<td>Kurtosis: 0.035</td>
</tr>
<tr>
<td></td>
<td>( P &gt; \chi^2: 0.0010 )</td>
<td>( P &gt; \chi^2: 0.1031 )</td>
</tr>
<tr>
<td>2008</td>
<td>Skewness: 0.001</td>
<td>Skewness: 0.482</td>
</tr>
<tr>
<td></td>
<td>Kurtosis: 0.006</td>
<td>Kurtosis: 0.115</td>
</tr>
<tr>
<td></td>
<td>( P &gt; \chi^2: 0.005 )</td>
<td>( P &gt; \chi^2: 0.2099 )</td>
</tr>
</tbody>
</table>

Interestingly, in the table above, the normality and homoskedasticity test results for both untransformed and transformed data are compared so that how much the Box-Cox power transformation improves the appropriateness of the model is researched. When the test results are examined, it is detected that both normality of a distribution and homoskedasticity of variance of the data transformed by the Box-Cox procedure are improved, while almost all of the untransformed data do not meet the assumptions of normality and homoskedasticity. This in response to the dependent variable; in this study, since all VIFs have the value less than seven, there is no need to concern about the matter of multicollinearity.

19) The Komorogov-Smirnov test for normality examines the cumulative distribution of the residuals against that of the theoretical normal distribution with the chi-square test to see whether a statistically significant difference exists; when \( p < 0.05 \), we must reject \( H_0 \) and the residuals are non-normally distributed; the null hypothesis of the Cook-Weisberg test for homoskedasticity is that the error variances are all equal versus the alternative so that the error variances are a multiplicative function of one or more variables. When the chi-square value is small, this indicates heteroskedasticity is probably not a problem.
implies the fact that the Box-Cox power transformation can be a practical utility in the empirical determination of a functional form in the hedonic price study and therefore this transformation technique helps set up a more robust estimation model.

4. Panel Data Analysis

Baltagi (2008) and Haiso (2003) tell us, “several benefits from using panel data are in the following: 1) As time-series and cross-section studies not controlling the heterogeneity run the risk of obtaining biased results, the panel data analysis is necessary to control for individual heterogeneity; 2) Panel data give more informative data, more variability, less collinearity among the variables, more degrees of freedom and more efficiency; 3) Panel data are better able to study the dynamics of adjustment. That is, the panel data analysis helps a study in detecting a multitude of changes hidden in relatively stable cross-sectional distributions; 4) Panel data are better able to identify and measure effects that are simply not detectable in pure cross-section or pure time-series data and etc”.

Unlike the previous case of the multiple linear regression (MLR), the hedonic price function slightly changes in the panel data analysis. That’s because the pooled data, a time-series one, is used to obtain the estimates. The functional form of the panel data analysis is in the following:

$$P_{it} = \beta_0 + \beta_1 NHH_{it} + \beta_2 AGE_{it} + \beta_3 SNU_{it} + \beta_4 CLG_{it} + \beta_5 BUS_{it} + \beta_6 SBW_{it} + \beta_7 PRK_{it} + \beta_8 STR_{it} + u_{it}$$

(4. 6)

where $P_{it}$ is the price, $NHH_{it}$ is the number of household, $AGE_{it}$ is the age of condominiums, $SNU_{it}$ is the number of student admitted into Seoul National University, $CLG_{it}$ is the admission rate into 4-year-course colleges, $BUS_{it}$ is the number of bus run, $SBW_{it}$ is the number of subway station, $PRK_{it}$ is the number of park, and $STR_{it}$ is the number of department store and discount store; $i$ indicates $i^{th}$ condominium and $t$ does time $t$; $\{(x_{i1} in, ..., x_{in}, P_y): i = 1, 2, ..., 60; t = Yr 2005, Yr 2006, ..., Yr 2009\}$; the form of $u_{it}$, the error term, varies in response to the assumption condition of the selected one among three models at large.

Meanwhile, the author does not strictly assume whether the error term$^{20}$ is correlated with the

---

$^{20}$ The assumptions on three panel data analysis models:

1. the seemingly unrelated regressions (SUR) or the pooled OLS model: $v_{it} = u_t + \varepsilon_{it}$, where $v_{it}$ is observed factors that affect the dependent variable and has a time invariant component ($u_t$: dependent across time and $\varepsilon_{it}$: independent across time) and $\text{cov}(v_{it}, v_{it}) = \sigma^2_{u} + 0$. 

---
independent variables or not and thus, first of all, an appropriate model is suggested by the Hausman test. Under the null hypothesis\(^\text{21)}\), as the test result does not reject \(H_0\), it is referred to as that there exists systemic difference between the independent variables and the error term when those of the random effect model and the fixed effect model are compared. Therefore, the former (RE model) turns out to be the more effective one to get consistent estimators.

In the mean time, the estimators from the random effect model rest on assumptions as

\[ \nu_i = u_i + e; \quad E(u_i) = 0, \quad \text{var}(u_i) = \sigma^2_u, \quad \text{cov}(u_i, u_j) = 0; \quad E(e_{it}) = 0, \quad \text{var}(e_{it}) = \sigma^2_e, \quad \text{cov}(e_{it}, e_{jt}) = 0; \quad \text{cov}(u_i, e_{it}) = 0 \]

where \( \nu_i \) is the compounded error term, \( u_i \) is the stochastic difference only across identities and \( e \) is the regression error or residuals.

These assumptions imply several important points in the respect of the functional form specification for the random effect model as follows: 1) Basically, the random effect model assumes that the independent variables are not correlated with the stochastic difference \( (u_i) \) so that the endogeneity problem of independent variables can be assumedly reduced to a great extent or even eliminated\(^\text{22)}\); moreover, since the variances of both the stochastic difference and the regression error are assumed as consistent value \((\sigma^2_u \text{ or } \sigma^2_e)\), under the assumption, heteroskedasticity of the error term also does not becomes a serious problem in this model; nonetheless, when the correlation of one identity \( (i) \) at different time \( (t) \) is considered, we can detect that the covariance of \( \nu_i \) and \( \nu_{i+1} \) are not same \( (\text{cov}(u_i + e_{it}, u_i + e_{i+1}) = \sigma^2_u \neq 0) \). This inconsistent covariance induces the first-order autocorrelation.

Such being the case, this author endeavors to obtain an appropriate model to deal with the autocorrelation problem and, for this effort, a generalized least square (GLS) is introduced. The functional form for this random effect model is specified as below.

\[
(P_{it} - \theta \bar{P}_i) = \beta_0 (1 - \theta) + \sum_{x = NH}^{STR} \beta_x (x_{it} - \theta \bar{x}_i) + [u_i (1 - \theta) + (e_{it} - \theta e_{i})] \quad (4.7)
\]

2. the fixed effect (FE) model: the error term (\( e \sim N(0, \sigma^2_e) \)) is independent for all time \( (t) \) and identities \( (i) \), while the estimators \( (\beta) \) change across time \( (t) \) and identities \( (i) \) but the constant \( (\beta_0) \) does only across identities \( (i) \), \( \text{cov}(x_{it}, u_i) = 0 \).

3. the random effect (RE) model or the error component model: \( \beta_{ii} = \beta_1 + u_i \), where \( \beta_{ii} \) is the constant, \( \beta_1 \) is the fixed population parameter and \( u_i \) is the stochastic difference across identities (random effect), \( \text{var}(u_i) = \sigma^2_u \) for more information, see Baltagi (2008) and Hill et al. (2011).

21) \( H_0: \text{cov}(x_{it}, u_i) = 0; \quad \chi^2 = 11.45, \quad P > \chi^2 = 0.1775; \) to refer to the test procedure, see Min (2009).

22) With this assumption, we can expect more consistent estimates when compared with the fixed effect model.

For more information, see Min (2009) and Hill et al. (2011).
where \( \bar{P}_i, \bar{x}, \bar{e}_i \) are the mean value of each identity and
\[ \theta = 1 - \frac{\sigma^2}{\bar{e}^2_i + \sigma^2_i}. \]

Through the generalized least square\(^{23}\) whose functional form is as shown above, the estimates which have the minimum variance for the random effect model can be obtained. Table 4.6 is the regression result of the random effect model and an interesting point is that, unlike the case of MLR with the Box-Cox transformation, two more variables (CLG and BUS) are included as statistically significant variables. In the table below, the value of theta (\( \theta \)) is 0.6821 and when putting this value into Equation (4.7), the functional form can be specified in detail as shown in Equation (4.8).

\[
\begin{align*}
\text{<Table 4.6> Estimation Result of Random Effect Model} \\
\begin{array}{|c|c|c|c|}
\hline
\text{Variable} & \text{Coefficient} & \text{z-statistic} & \text{P > |z|} \\
\hline
\text{NHH} & 0.0240 & 0.45 & 0.655 \\
\text{AGE} & -16.9522*** & -5.49 & 0.000 \\
\text{SNU} & 39.7450*** & 7.22 & 0.000 \\
\text{CLG} & 4.3723** & 2.12 & 0.034 \\
\text{BUS} & 0.1912*** & 3.26 & 0.001 \\
\text{SBW} & 35.2495 & 1.27 & 0.203 \\
\text{PRK} & -4.9686 & -0.22 & 0.822 \\
\text{STR} & -40.9587 & 1.60 & 0.109 \\
\text{Constant} & 1,087.437*** & 7.40 & 0.000 \\
\hline
\end{array}
\end{align*}
\]

\[
(P - 0.6821 \bar{P}_i) = \beta_0 (1 - 0.6821) + \sum_{x = \text{NHH}}^{SFR} \beta_x (x_{it} - 0.6821 \bar{x}_i) + [u_i (1 - 0.6821) + (e_{it} - 0.6821 \bar{e}_i)]
\]

(4.8)

Meanwhile, the autocorrelation of the error term is examined by the Wooldridge\(^{25}\) test

\(^{23}\) Following Baltagi (2008), cross-sectional dependence and autocorrelation tests mainly apply to macro panels with long time series (e.g., over 20-30 years). Otherwise, this paper tried to address such an autocorrelation problem by introducing the GLS method.

\(^{24}\) The “rho” is the fraction of variance due to \( u_i \): the portion of \( u_i \) to whole error term. And it can be described in the form of an equation as follows: \( \rho = \sigma_u^2/(\sigma_u^2 + \sigma_e^2) \)

\(^{25}\) The Wooldridge test is a first-differencing-based test of serial correlation for (the idiosyncratic component of) the errors by examining the difference between different orders; \( H_0 = \) no first order autocorrelation;
and the result shows that there is no need to concern about the autocorrelation problem because the null hypothesis of this test is not rejected.

With that estimation result above, what the author want to emphasize here is that the influence of SNU becomes greater, when compared with the case of MLR. Besides, the other educational factor, CLG, is also included as a statistically significant variable with not much high degree. Comparing the estimation result of the multiple linear regression (MLR) by the Box-Cox transformed data with that of the panel data analysis, Table 4.6 serves some valuable information to us as follows: first, the panel data analysis is likely to provide us with more information on the effect of omitted unobserved variables indeed. Through the panel data analysis, this paper can include other statistically significant variables such as CLG and BUS. Because the author comprehends the importance of the functional form specification in econometric study fields and the fact that the estimation results highly depends on the precision of the model specification, two empirical analysis technologies are employed and thereby the panel regression model presents a different analysis result, the effect of omitted unobserved variables, when compared with the multiple linear regression model; second, the panel data analysis statistically shows greater influence of educational factors on condominium prices; it is examined by the magnitude of coefficient change of SNU; and the other educational factor becomes a statistically significant variable. Nonetheless, what the author would like to argue here is not how much the influence of educational factors becomes greater in response to the analysis model but the fact that, throughout these empirical analyses, only one educational factor, SNU, remains statistically significant in all cases, while another important external variables for transportational factors turn out statistically significant only once; this result clarifies the author's argument that the most decisive factor to determine the condominium prices in Daegu is educational one indeed.

---

26) When the estimation results are compared, we can see that, when a highschool has one more student admitted to Seoul National University, the price of an adjacent condominium statistically goes up as by 1) Multiple linear regression: KRW 100 thousand/100㎡ (after λ value of the Box-Cox procedure is reversed).
2) Panel data analysis: almost KRW 4 million/100㎡, ceteris peribus. and the result of the panel data analysis shows much higher value.

27) Transportation factors are important elements of determining the hedonic price in housing market studies. There is a compelling evidence to support this argument: in the traditional monocentric model, it is described as that the price difference of real estate is an offset against the change of transportation cost.
V. CONCLUSION

According to the estimation results, only one educational factor, specifically the number of student admitted to Seoul National University, was statistically significant for all estimations. Besides, other variables such as the age of condominiums, the admission rate into 4-year-course colleges and the number of bus run were statistically significant occasionally. Unlike the previous studies targeting Seoul\(^{28}\), the variable relevant to the subway station appeared not to be statistically significant. This estimation result indicates several facts as follows: first, as mentioned in Section I, because not only does the road traffic in Daegu serve satisfactory conditions, but the subway network is not completed yet, transportational factors rarely affect the home buyers' purchasing decision in this city and this differs from the case of Seoul; second, other variables representing living conditions are even excluded from the list of significant variables. This is probably because the citizens in Daegu can use those facilities conveniently owing to relatively low population density; lastly, only the educational factors considerably show statistical significance during empirical analyses. This result is highly attributed to the Korean educational fever (a society based on academic success) and the city's site-specific characteristics that differ from Seoul's characteristics such as inconvenient facility use due to high population density and high demand for the public transportation to avoid terrible road traffic conditions.

Meantime, a research report from the Korea Housing Institute interestingly argued that, when the housing supply track record and the requirement for supply for a five-year time period were compared, the housing was oversupplied to almost all region in this country except the Seoul Metropolitan Area and Daejeon (Kim et al. 2009). Not only that, following the report, in the case of Daegu, the oversupply index stood at 169.6 and the accumulation rate of vacancy\(^{29}\) had the value of 36.0, occupying the second-highest level. Furthermore, this report warned that this phenomenon of Daegu highly resulted from the structural problem caused by the housing types that regional demand attributes were not considered and the high selling prices.

\(^{28}\) Following the results of previous studies targeting on Seoul, Jung (2006) found that not only educational factors but also other traits of condominiums such as the distance to a subway station were included as to be statistically significant at 10% level; Eom et al. (2006) also found that the number of household, the age of condominiums, the distance to a subway station and the number of park statistically affected the price of condominiums at 1% significant level. From the standpoint of statistically significant variables, these analysis results present a different aspect against the analysis result of this study based on Daegu.

\(^{29}\) Oversupply Index: "housing supply track record" over "requirement by demand estimation" from 2004 to 2008, Accumulation Rate of Vacancy: "The number of unsold house" over "the number of supplied house" (Source: http://www.khi.re.kr/info INFO1.php?boardid=board2&mode=view&idx=55&sk=\&sw=a&offset=)
and thus, such structural problem would be hard to solve in a short time period.

This result of the report above, compounded with two facts that the population of Daegu is anticipated to decrease by the degree of more than 0.7% (more than 0.4 million people) until 2030 and the Urban Affairs and Housing Bureau of Daegu City has a plan to supply cumulatively 343 thousands more houses until the year 2020\(^30\), can probably make the structural problem in Daegu worse.

If these conflicting situations above are incorporated with the result of this study, the most decisive factor that determines the housing prices in Daegu is educational fever, a more serious consequence will follow for the following reasons: to begin with, because of the decrease in population, transportation and other living conditions will hold a better position and it goes without saying that the vacancy rate by the structural problem of the housing market in Daegu will worsen. In addition, as planned by Daegu city, more housing supply will make this situation substantially deteriorated; nonetheless, as the result of this study shows, the convergence of population into Suseong district in which the education performance of high schools is excellent will remain still steady; these situations will incur a serious problem that only the suburbs or unpopular zones of this city will suffer from an increasing vacancy rate; that is to say, inappropriate growth of the housing market, biased to a certain district, will be accelerated.

Up to now, this paper explained a predictable serious problem. In this regard, it is necessary to devise a more rational housing plan to prevent the growth of the housing market in Daegu from being biased to a certain district. Without a comprehensive housing plan that deals with the unique site-specific characteristics of this city, additional housing supply in the future will incur another problem of urban sprawl. Conclusively, by recognizing an undeniable fact that there is a limit on demand creation for housing because of a decrease in population and poor conditions of other socioeconomic factors, a more prudent policy considering various social aspects should be served in order for this city to grow smartly. In the case of Daegu, the educational fever can be one of the social aspects.

References


국문요약

교육성과가 아파트가격에 미치는 영향에 대한 실증적 분석:
대구시 사례를 중심으로

본 연구에서는 교육성과가 주택시장(아파트 가격)에 미치는 영향을 분석하였다. 이러한 교육성과와 주택시장의 상관관계를 다루는 이전의 연구들이 대부분 서울시도권을 대상으로 하고 있어, 역시나 교육열이 높은 다른 지역에 대한 연구 또한 필요한 것으로 판단됨에 따라 비수도권 지역인 대구광역시 사례를 중심으로 그러한 상관관계를 분석하였다.

대구광역시의 교육성과와 아파트가격의 상관관계를 실증적으로 분석하기 위한 적정 모형설정을 위해 "박스-콕스 변환모형(the Box-Cox Power Transformation Model)"과 "패널 데이터 회귀모형(the Panel Regression Model)"을 이용하였으며, 2005년에서 2009년까지 5 개년간의 자료를 이용하여 모형을 추정하였다. 추정결과, 대구시 아파트가격 결정에 통계적으로 가장 큰 영향을 미치는 변수는 교육관련 요소(특히 서울대합격자수)이었으며, 주거환경과 관련된 나머지 변수(아파트 개별특성, 교통조건 등)는 그 영향성이 미미한 것으로 나타났다. 이러한 추정결과는 서울시를 대상으로 한 이전 연구와는 또 다른 양상을 보이는 바, 이는 도시구조 특성이 서로 다른 것에 기인하는 것으로 판단되었다.

한편, 대구시의 이러한 교육성과와 아파트가격이 가지는 특수성은, 현재 파업 공급된 주거용 부동산 및 높은 공성비라는 대구시 주택시장의 구조적 문제가 함께, 기타 행후 사회적 변화를 고려하였을 때, 우수한 교육성과를 내는 고등학교들이 많이 있는 특성지역(수성구)으로의 인구집중을 더욱 가속화 할 우려가 있어, 변두리 지역의 공성비 악화와 함께, 한 도시 내 편중된 지역발전이라는 악영향을 초래할 수도 있을 것으로 판단되는 바, 대구시 지역특성을 고려한 신중한 주택정책이 요구된다.