「한국 시스템다이내믹스 연구」 제8권 제2호(2007.11) : 25-49

The Regional Homogeneity in the Presence of Heteroskedasticity

Chung, Kyoun-Sup* · Lee, Sang-Yup**

Abstract

An important assumption of the classical linear regression model is that the disturbances appearing in the population regression function are homoskedastic; that is, they all have the same variance. If we persist in using the usual testing procedures despite heteroskedasticity, what ever conclusions we draw or inferences we make be very misleading. The contribution of this paper will be to the concrete procedure of the proper estimation when the heteroskedasticity does exist in the data, because the quality of dependent variable predictions, i.e., the estimated variance of the dependent variable, can be improved by giving consideration to the issues of regional homogeneity and/or heteroskedasticity across the research area. With respect to estimation, specific attention should be paid to the selection of the appropriate strategy in terms of the auxiliary regression model. The paper shows that by testing for heteroskedasticity, and by using robust methods in the presence of with and without heteroskedasticity, more efficient statistical inferences are provided.

Keywords: Heteroskedasticity, Chow & Wald Tests, OLS & WLS Estimates

* Pukyong National University (Main Author, kschung39@pknu.ac.kr)

** Hanseo University (Corresponding Author, leesy598@hanmail.net)

Ⅰ. Introduction

When it usually comes to the heteroskedasticity, it will cause the estimated variance of the coefficient to influence both direction, upward and downward. In other words, the estimated of variance of the coefficient can be over-estimated or under-estimated depends on the nature of causing factors in the model. In short, if we persist in using the usual testing procedures despite heteroskedasticity, whatever conclusions we draw or inferences we make be very misleading. As a result, we can no longer rely on the conventionally computed confidence intervals and the conventionally employed t and F tests (Gujarati, 1988). Put it another way, it is not easy to know the estimated variance of the coefficient, in fact, which is improved, even though the proper method is applied without knowing the exact causing nature of heteroskedasticity. Hedonic housing price regression has been an indispensable tool in econometric studies of urban housing markets. Its theoretical basis is well articulated within a traditional urban economics framework, but not enough attention has been paid to its specification in empirical applications, especially with respect to problems that might due to the spatial nature of data sets and the nature of urban space. The value of this study will be to the tax officers who must reassess property at full market value each year, since the quality of housing price predictions can be improved by giving consideration to the issues of regional homogeneity and/or heteroskedasticity across market areas. With respect to estimation, specific attention should be paid to the selection of the appropriate econometric strategy. The paper shows that by testing for spatial effects, and by using robust methods in the presence of heteroskedasticity, more realistic statistical inferences are provided. There is feedback from the results to urban theory, too. Heretofore this theory has assumed homogeneous urban markets, an assumption that clearly needs to be modified: more efficient estimates of housing prices are provided by calibrating hedonic models for each significantly different neighborhood submarket.

Section II explains the research problem due to the nature of heteroskedasticity. Section III explains the literature review. In the section IV, the operationalization of the research is presented, Section V illustrates the result of analysis, and section VI concludes.

Ⅱ. Research Problem

A substantial part of the econometric analysis in urban or regional economics is based on data collected for spatial units with irregular and arbitrary boundaries. Nonetheless, the interpretation of the various models and the implications for policy are often made with respect to a general notion of space. This would imply that there is some unique and identifiable spatial structure, with clear statistical properties, independent from the way in which the data are organized in spatial units. Unfortunately, matters are not this straightforward. The modifiable areal unit problem pertains to the fact that statistical measures for cross-sectional data are sensitive to the way in which the spatial units are organized.

Specifically, the level of aggregation and the spatial arrangement in zones (i.e., combinations of contiguous units) affects the magnitude of various measures of association, such as spatial autocorrelation coefficients and parameters in a regression model.

As is well known, aggregate models are only meaningful if the underlying phenomenon is homogeneous across the units of observation. Unless there is a homogeneous spatial process underlying the data, any aggregation will tend to be misleading. If this is not the case, both the heterogeneity and structural instability that are present should be accounted for in any aggregation scheme. This aspect of the modifiable area unit problem should be considered as a specification issue related to the form of spatial heterogeneity, and not solely as an issue determined by the spatial organization of the data (Anselin, 1988).

In the conceptualization offered by the classic hedonic price function, $p = f(S, Age, \alpha, \beta) + \varepsilon$ where P is a vector of observed market expenditures on housing at the date of sale, S is the vector of structural characteristics, and Age is the year built, used to construct an age depreciation price effect. α and β are the corresponding parameter vectors, and ε is the vector of random error terms. It has usually been assumed that the effect of structural housing characteristics on property values, a sindicated by α 's, is "fixed" - that is, in variant across neighborhoods.

But this begs the question of where there is a uniform housing market within a given study area, or whether the area is segmented. If stability of the parameters across neighborhoods is revealed, this implies the presence of a single competitive market in the long run since there will be only one price schedule. On the other hand, if neighborhood differentials lead to varying attribute prices, this will indicate the presence of independent price schedules, and thus the existence of segmented markets. The presence of a statistically significant difference between the estimated coefficients for neighborhood sub samples and those obtained using the entire sample, usually based on a Chow test(Chow, 1960), it may provide evidence for the occurrence of market segmentation, and thus may indicate the existence of independent hedonic price schedules.

Segmentation may, of course, be contextual: parameter stability should not be assumed, but rather should be searched for. The α 's in the equation may take different values across space corresponding to socio-economic and environmental variations across neighborhoods. In such a case, interactions between S and neighborhood differentials should be sought to bring housing price estimation back into the framework of a single model. The major difficulty in testing for market segmentation is the arbitrary nature of the delineation of geographic submarkets. Proof that S and neighborhood differentials interact and that the interactions account for the submarket differences identified by the Chow test would obviate the difficulty.

Violation of the homoskedasticity assumption with respect to the error term also is a problem, leading to overestimation or underestimation of the variance of the coefficients of the classical hedonic housing price regression, even though it does not affect bias.

To test for heteroskedasticity of the hedonic housing regression, a Breusch Pagan test should be implemented. If heteroskedasticity is detected, Weighted Least Squares (WLS) estimation should be utilized, using the results of auxiliary regressions following the Breusch Pagan Test. The F statistic of market homogeneity based on the Chow test is sensitive to the assumption both of normal error terms and of homoskedasticity. In order to relax these restrictions, a Wald test can be used. The Wald statistic has a chi-squared distribution asymptotically and can be used instead of the F ratio.

Ⅲ. The Literature Review

All of the applications of the hedonic price function are rooted in Lancaster's consumer behavior theory(Lancaster,1966). In this framework, housing is a multidimensional good differentiated into a bundle of attributes that vary in both quantity and quality. Accordingly, the hedonic housing price regression becomes an operational tool that functionally links housing expenditures to some measures of attributes of houses. The traditional econometric approach has been to regress the housing value on a function of various structural and neighborhood attributes of dwellings. The estimated coefficients provide "hedonic prices" which are also called marginal (implicit) prices of the attributes considered.

In the theoretical framework provided by Rosen (1974), the hedonic price function is the market clearing price produced by the interaction of bid functions of households and offer functions of suppliers. In other words, within conventional utility maximization theory, Rosen clarified the essence of the hedonic price function and provided the conceptual basis for the interrelationships of the offer functions of suppliers, the bid price functions of consumers, and hedonic prices. According to Rosen, in the presence of different producers and of consumers who are dissimilar in tastes and income, the hedonic price function is neither a demand nor a supply function, but a market-clearing function determined by consumers' bids and suppliers' offers. Hence, the estimated hedonic regression simply exhibits the market-clearing marginal attribute prices. In this conceptual framework, the hedonic price function equates the observed market price of housing expenditures to the housing attributes and reveals the marginal prices of attributes. At equilibrium in a single competitive market, the formal relationship between the observed household expenditures on housing, P(H), and the level of characteristics contained in vector H, letting housing be a heterogeneous commodity differentiated into a bundle of attributes, is described as follows:

$$P(H) = f(h_1, h_2, h_3, \ldots, h_k) \quad \text{---} \quad (1)$$

From this equation, we can refer to the price of any attribute k contained in H, $P_k (\equiv \partial P(H) / \partial h_k)$, as the equilibrium marginal (implicit) price of that attribute. Upon the proper functional specification of the hedonic price function, the estimated coefficients will provide the estimated marginal prices of attributes.

Within the hedonic framework, housing market expenditures are functionally related to various housing attributes by the hedonic price function. Generally, the housing attributes are classified into two or three major groups. The first comprises the structural characteristics of the dwelling, such as its style, its lot size, the number of bedrooms, and the structural integrity of the building. The second group consists of neighborhood characteristics, including socioeconomic characteristics, availability of urban amenities, and the level of public services. The third group, sometimes, referred to as locational effects, are externalities associated with the geographic location of the dwelling, both its absolute location and the neighborhood in which it is located in the urban area.

Empirical studies applying hedonic functions to housing prices are becoming quite numerous but their essential similarities obviate a complete review. Instead, what I will highlight in this section are those studies that bear directly on the analysis proposed here. According to Berry(1976), the city of Chicago experienced a large degree of minority neighborhood expansion during 1968-1972. His study, investigating the differences in price levels and rates of price increase in six distinctive submarkets, was one of the first to address the question of submarket variability. The Chicago housing market was segmented into white peripheral and white adjacent to minority neighborhoods, the black expansion zone and ghetto, and the Spanish expansion zone and ghetto. The study applied a series of hedonic price models both in linear and logarithmic form.

Can (1990) addressed a different issue: that when cross-sectional data (either areal or point) are used in estimating the hedonic housing linear regression model, the standard assumptions are likely to be violated, resulting in the occurrence of spatial dependence (autocorrelation) and spatial heterogeneity; i.e., the assumption under normality is that ε_i and ε_j are uncorrelated so that the covariance between them is equal to zero - independence and homoskedasticity; these terms are commonly referred to as i.i.d. normal error terms. In terms of spatial econometrics, spatial dependence refers to the possible occurrence of interdependence among observations that are viewed in geographic space, and violates the assumption of uncorrelated error terms in model estimation. On the other hand, spatial heterogeneity refers to the systematic variation in the behavior of a given process across space, and usually leads to heteroskedastic error terms, thus violating the assumption of homoskedasticity in the classical regression model.

Spatial dependence or spatial autocorrelation has been treated in the seminal work of Cliff and Ord (1973). Others, such as Upton and Fingleton (1985), Griffith(1987), and Anselin(1988), discuss the subject at varying levels of complexity and from different perspectives.

In most empirical settings, spatial contextual variation is the major cause of spatial heterogeneity. Thus the use of the Casetti's (1972) expansion method inherently allows for the presence of spatial heterogeneity, which is one of the fundamental properties of spatial data. The method eliminates the part of heterogeneity resulting from the 'drift' of parameters across submarkets. Even so, the part resulting from missing variables or other forms of misspecification that lead to heteroskedastic error variance will still be present. There thus is a need to conduct diagnostic tests for heteroskedastic error terms.

The presence of spatial dependence in the geographic structure exhibited by housing prices will violate one of the major properties of regression analysis, i.e., independence across observations. For example, if the prices of nearby houses are similar only because they share common locational factors, then since contextual drift captures this locational effect, spatial autocorrelation will diminish. If the prices of nearby houses have an absolute effect on each other, for example when appraisers assess a given house price in terms of the prices of nearby houses, there will be a need to incorporate an autoregressive term in the model specification in addition to the contextual drift. This autoregressive term will capture spatial spill over effects and thus measure the absolute price effect on nearby dwellings for a given house.

In some cases, a particular form of contextual drift modeling may eliminate spatial autocorrelation if it accounts for the spatial variation in the housing prices. On the other hand, in addition to inclusion of contextual drift, there may be a need for an autoregressive term in the model specification to capture the full scope of spatial dependence in the residuals of the estimated models, to determine whether or not an autoregressive term is needed in the model specification.

Waddell, Berry, and Hoch (1993), knowing that earlier hedonic regression models of housing have ignored or underspecified the locational characteristics of housing, explored potential improvements in the hedonic model based strictly on dwelling attributes by successively adding spatial externalities and a range of variables designed to capture the benefits or disbenefits of neighborhood and environment characteristics to residential property in addition to develop a housing price index for the Dallas housing market. To illustrate spatially autocorrelated errors, they mapped residuals from an initial model that included only dwelling attributes, and showed how the clustering of residuals could be virtually eliminated by successive addition of neighborhood, locational and environmental variables to the equation. At no stage, however, did they test for spatially autocorrelated residuals, a gap that needs to be filled.

Ⅳ. Operationalization of the Research

One of the purposes of this study is to search for better techniques for the estimation of housing prices. A key issue is whether the research housing market area is homogeneous, so that a single estimating equation may be used, or whether they are significant submarket differences in hedonic prices, in which case multiple models may be used and homogeneous submarkets may be identified.

The data for this study were made available through various research projects at the Bruton Center for Development Studies in the University of Texas at Dallas. DCAD provided the original data set, comprising the housing stock records in their certified tax appraisal files, together with the completed sales transactions of single-family dwelling units in Dallas County during the period of 1979-1993. Each sale record contains the date of sale, market price, tax ID code, etc. From this data set, I chose the sales transactions occurring in 1993 to investigate cross-sectional market homogeneity. The number of observations is 1,759. Selling prices that were reported ranged from $ 22,000 to $ 1,092,623. Dwelling characteristics used in the hedonic regression include the log of living area, the age of dwelling, dummy variables for the number of full and half baths, the number of fireplaces, and the presence of a pool, sauna, and wet bar in the Table1. These independent variables are listed in Table2.

Table 1. Descriptive Statistics

	N	Min	Max	Mean	Standard Deviation
LGPRICE	1759	9.999	13.904	11.71305	.450777
FBATH2	1759	0	1	.76	.429
FBATH3	1759	0	1	.18	.384
FBATH4	1759	0	1	.04	.202
HBATH1	1759	0	1	.28	.450
HBATH2	1759	0	1	.00	.063
WETBAR	1759	0	1	.33	.470
FIREPL1	1759	0	1	.85	.361
FIREPL2	1759	0	1	.11	.318
SAUNA	1759	0	1	.01	.086
POOL	1759	0	1	.20	.403
LGLIVA	1759	6.534	8.872	7.63497	.325330
AGE	1759	2	65	11.06	6.150
PRICE	1759	22000	1092623	137869.70	87551.918
Obs	1759				

Table 2. List of Independent Variables

Variable	Description
FBATH2	Full baths = 2
FBATH3	Full baths = 3
FBATH4	Full baths = 4 or more
HBATH1	Half baths = 1
HBATH2	Half baths = 2
WETBAR	Wet bar present
FIREPL1	Fireplace = 1
FIREPL2	Fireplace = 2 or more
SAUNA	Sauna present
POOL	Swimming pool present
LGLIVA	Log of living area in square feet
AGE	Age of construction
AGESQ	Age times Age
AGECUBIC	Age times AGESQ

Among these independent variables, I chose the log of living area (i.e., LGLIVA) as the functional form; accordingly, the coefficient of LGLIVA is the elasticity itself, the percent change in price of housing with respect to percent change in the living area, because the dependent variable is also transformed to logarithms. The rationale for this transformation may be seen in Figures1 and 2. In Figure1, it is obvious that the relationship between the log of price of housing and living area is non-linear. In Figure2, the relationship between the dependent variable and log of living area is, however, linear.

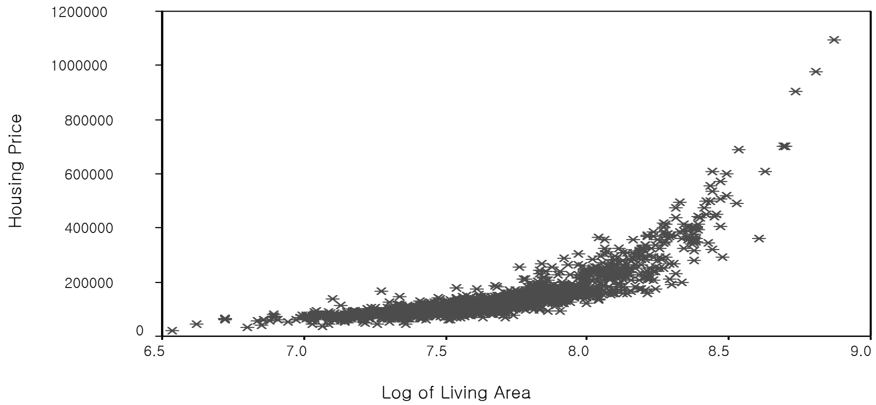

Figure 1. The Relationship between The Log of price of Housing and Living Area

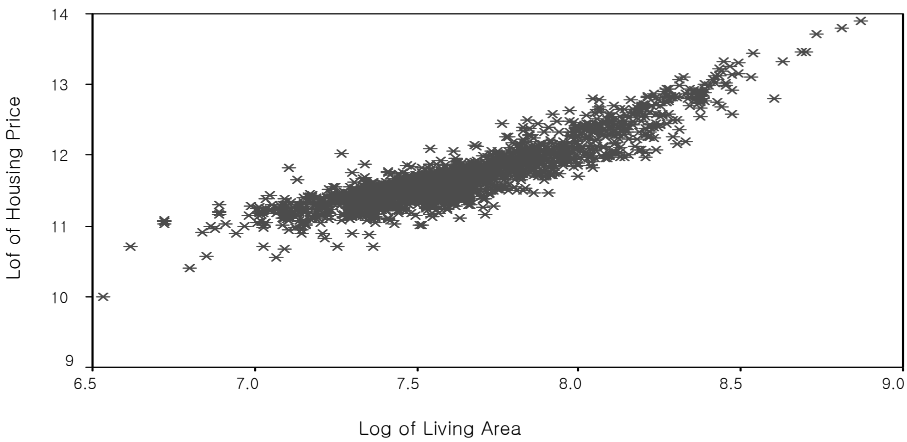

Figure 2. The Relationship between The Price of Housing and Living Area

The initial model that was estimated can be expressed as P = f (S, Age) where P is the logarithm of the total sales price of single family housing at the date of sale; S is a vector of dwelling characteristics, and Age is designed to capture depreciation by way of a cubic spline function. In greater detail, we thus have:

$$LGPRICE = \alpha_0 + \alpha_1 FBATH2 + \alpha_3 FBATH3 + \alpha_4 FBATH4 + \alpha_5 HBATH1 + \alpha_6 HBATH2 + \alpha_7 WETBAR + \alpha_8 FIREPL1 + \alpha_9 FIREPL2 + \alpha_{10} SAUNA + \alpha_{11} POOL + \alpha_{12} LGLIVA + \alpha_{13} AGE + \alpha_{14} AGESQ + \alpha_{15} AGEGCUBIC + \varepsilon$$

where α_0 - α_{15} are the corresponding coefficients on structural variables and ε is the error term, assuming i.i.d. with normal distribution.

Before the homogeneous housing submarkets are determined that any pair of areas (i) and (j) should be combined into a larger submarket on the basis following rules: (1) Contiguity Rule: Even if the P-value was below the critical value, the evidence for homogeneity was ignored unless the areas also were contiguous(see Figure3, Figure4); (2) Absolute Homogeneity Rule: If sets of areas were combined on the basis of the P- value based on Chow and Wald tests, it was required the P- value for every pair of areas within aputative aggregation be beneath the critical level(see Figure5, Figure6); and (3) Lowest P-value Rule: If two sets of area, such as (A & B) & (A & D) are homogeneous and but B & D areas are not homogeneous, then the lowest P-value area will be segmented(see Figure7, Figure8).

In the following figures, suppose that A & B areas are homogeneous submarkets by way of Chow test or Wald test, and H & I areas also are homogeneous submarkets. Then how to divide the homogeneous market areas. In this case, even though they are homogeneous, they can not be merged because the areas are not contiguous. Therefore, the market areas are segmented.

A	B	C
D	E	F
G	H	I

Figure3. Market Areas

<table>
<tr><td colspan="2">A B</td><td>C</td></tr>
<tr><td>D</td><td>E</td><td>F</td></tr>
<tr><td>G</td><td colspan="2">H I</td></tr>
</table>

Figure4. The Outcome of Contiguity Rule

In the following figures, suppose that A & B & D & E areas are homogeneous by way of Chow and Wald tests. Then the pair-wise P-value, such as A & B, A & D, A & E, B & D, B & E, D & E, must be below the critical value separately and simultaneously.

A	B	C
D	E	F
G	H	I

Figure 5. Market Areas

<table>
<tr><td colspan="2" rowspan="2">A B
D E</td><td>C</td></tr>
<tr><td>F</td></tr>
<tr><td>G</td><td>H</td><td>I</td></tr>
</table>

Figure 6. The Outcome of Absolute Homogeneity Rule

In the following figures, suppose that A & B areas are homogeneous submarkets by way of Chow test or Wald test, and A & D areas also are homogeneous submarkets. But B & D are not homogeneous. Then how to divide the homogeneous market areas. In this case, the lowest P-value areas are merged. For example, by way of Chow test, if A & B(F-score 36, P-value 0.001) and A & D(F-score 16, P-value 0.05), then the area A & B should be merged at first.

A	B	C
D	E	F
G	H	I

Figure 7. Market Area

A B		C
D	E	F
G	H	I

Figure8.TheOutcomeof LowestP-valueRule

The Chow test statistic that is computed is as follows:

$$F = \frac{(ESS_{pool} - (ESS_{mkt(i)} + ESS_{mkt(j)}))/q}{(ESS_{mkt(i)} + ESS_{mkt(j)})/(N_{mkt(i)} + N_{mkt(j)} - 2K)}$$

for mkt(i) is not equal to mkt(j).

where mkt(i) and mkt(j) are the initial market areas into which the study area is divided.

(i) = 1,2,3, ⋯; (j) = 1,2,3, ...

ESS = Error Sum of Squared

ESS_{pool} =ESS for pooled sample comprising mkt(i) and mkt(j)

$ESS_{mkt(i)}$ =ESS for mkt(i)

$ESS_{mkt(j)}$ =ESS for mkt(j)

q = the number of restrictions

K = the number of explanatory variables

N = the number of observations

df = degree of freedom, (q, $N_{mkt(i)} + N_{mkt(j)} - 2K$).

As the Chow tests proceeded, it became apparent that there was violation of homoskedasticity with respect the error term in the regression models. The estimated parameters therefore were inefficient (OLS estimation, of course, does not affect bias). As a result, depending on the relationship between the error term and the independent variables, the variance of the coefficients can be overestimated or underestimated. If there is heteroskedasticity, proper estimation is in order, using Weighted Least Squares (WLS), with proper weights sought by Breusch Pagan auxiliary regression. In this procedure, the weight by the natural log of $\hat{e}^2$ from the auxiliary regression is used in the WLS estimation rather than using $\hat{e}^2$ itself (which is used for Breusch Pagan heteroskedasticity test).

The Breusch Pagan test procedure can be summarized as follows: (1) obtain the residuals of the estimated regression equation; (2) use the $\hat{e}^2$ as the dependent variable with all the variables suspected of being related to the variance of the error term of the original equation included as independent variables in an auxiliary regression;

$\hat{e}_i^2 = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \ldots + \alpha_k Z_{ki} + \varepsilon_i$,and (3) $L = \dfrac{RSS}{2(\sum \hat{e}_i^2 / N)^2}$, where RSS is regression sum of squares from the auxiliary regression and N is the number of observations. The L statistic is distributed as chi-square asymptotically with the number of slopes, except the intercept in the auxiliary regression degrees of freedom.

Without using the normal error term and homoskedastic error assumptions of the Chow test, the Wald test provides the asymptotically efficient estimation of the homogeneity of submarket regressions. The Wald test is superior in this sense to the Chow test, which assumes that normal error terms and homoskedastic error terms, i.e., $\sigma_i^2 = \sigma^2$. The Wald test is also to be preferred for other reasons, because it tests for homogeneity of individual structural coefficients rather than providing a single F ratio for pairs of market area models. In this way, it is not only free of restrictive assumptions; it provides a far more sensitive evaluation of market homogeneity/heterogeneity.

The Wald statistic is calculated in the following manner:

$$W = (\hat{\beta}^{(i)}_{WLS} - \hat{\beta}^{(j)}_{WLS})\ \{Var(\hat{\beta}^{(i)}_{WLS}) + Var(\hat{\beta}^{(j)}_{WLS})\}^{-1} (\hat{\beta}^{(i)}_{WLS} - \hat{\beta}^{(j)}_{WLS})$$ for (i) not equal to (j).

where (i) = 1,2,3,...; (j) = 1,2,3,...

( )' = the transpose of the matrix,

$\{\ \}^{-1}$ = the inverse of the matrix

$\hat{\beta}^{(i)}_{WLS}$ = estimated parameter vector from WLS based on Breusch-Pagan regression for the mkt(i)

$\hat{\beta}^{(j)}_{WLS}$ = estimated parameter vector from WLS based on Breusch-Pagan regression for the mkt(j)

$Var(\hat{\beta}^{(i)}_{WLS})$ = estimated variance of the parameter vector from WLS based on Breusch-Pagan regression for the mkt(i)

$Var(\hat{\beta}^{(j)}_{WLS})$ = estimated variance of the parameter vector from WLS based on Breusch-Pagan regression for the mkt(j).

The W statistic assumes a chi-squared distribution asymptotically with the number of restrictions on the degrees of freedom.

Ⅴ. Results of the Analysises

The study area is can be divided into 4 counties in the beginning stage, such as Carrollton, Dallas, Coppell, and Garland; they are located at north of the city of Dallas. The average of housing price of the study area is above the mean of the Dallas city. I don’t think it is necessary to draw the exact map of the study area, because the result of analysis is the exactly same in this conventional diagram as Figure9.

Carrollton	Dallas
Coppell	Garland

Figure 9. The Research Area

The results of estimating the initial model are summarized in Table3. The adjusted - R^2 is 0.908. The overall goodness of fit is plausible. The histogram of the log of housing price with respect residual is presented in the Figure10. The expected signs are reasonable except the Full Bath (i.e., FBATH2, FBTATH3) and Fireplace(FIREPL1); maybe, they don't have enough observations comparing other independent variables.

Table 3. The Result of Initial OLS Model

		Unstandaridized		Standarized		
		B	SE	Beta	t	Sig
1	(상수)	6.019	.130		46.153	.000
	FBATH2	-.142	.029	-.135	-4.915	.000
	FBATH3	-2.253E-03	.033	-.002	-.069	.945
	FBATH4	.188	.039	.084	4.847	.000
	HBATH1	1.263E-02	.009	.013	1.441	.150
	HBATH2	.216	.053	.030	4.071	.000
	WETBAR	6.598E-02	.009	.069	7.379	.000
	FIREPL1	-3.548E-02	.021	-.028	-1.699	.090
	FIREPL2	.105	.025	.074	4.186	.000
	SAUNA	8.029E-02	.039	.015	2.070	.039
	POOL	.125	.009	.112	13.256	.000
	LGLIVA	.780	.018	.563	42.709	.000
	AGE	-1.985E-02	.003	-.271	-7.088	.000
	AGESQ	1.795E-04	.000	.086	1.209	.227
	AGECUBIC	-1.887E-06	.000	-.039	-.971	.331

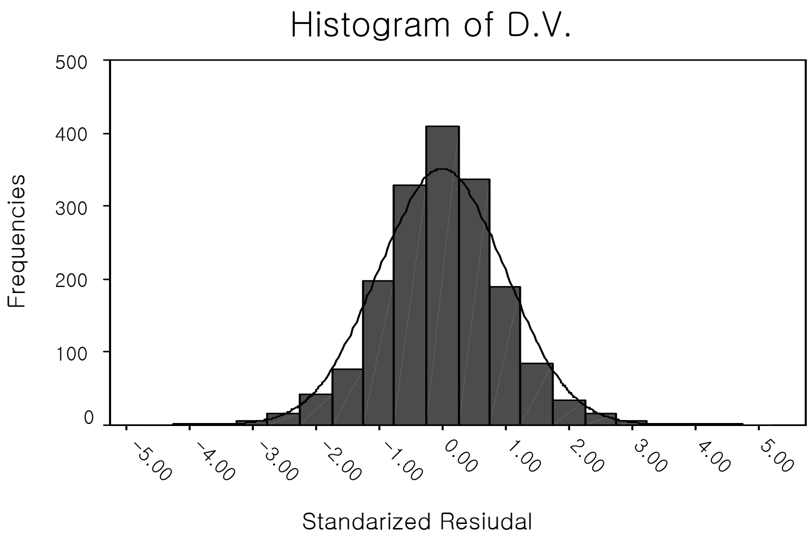

Figure 10. The Histogram of Log of Price in the Initial Model

First of all, the result of Chow test is presented in the table 4 without any restrictions such as heteroskedasticity etc.

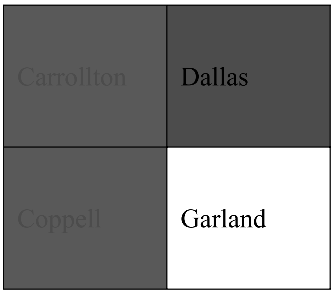

Figure 11. The Result of Chow Test

The Chow statistic and the corresponding critical values are presented in the Table4.

As can be seen in the Figure11 and Table4, the homogeneous submarkets are grouped into

3 different housing submarkets. The homogenous housing submarket can be grouped into 2 submarkets such as Carrollton, Coppell, Garland, vs.Dallas in the first stage. But Coppellvs. Garland cannot be categorized into the homogeneous housing submarket since the F-statistic is 3.39 which is the above the critical value 2.33.

Table 4. The Chow Statistic

α	Critical Value	Carrollton vs. Dallas	Carrollton vs. Coppell	Carrollton vs. Garland	Dallas vs. Coppell	Dallas vs. Garland	Coppell vs. Garland
0.01	2.33	5.48*	1.27	1.62	4.36*	6.10*	3.39*

* At 1% level of significance, the null hypothesis(the parameters are same in two submarkets) is rejected.

As mentioned before, the Chow test does not consider the heteroskedasticity; therefore, the heteroskedasticity test should be in order. Accordingly, the statistic of the Breusch Pagan test can be summarized in the Table5.

Table 5. The Result of Breusch-Pagan Test

	Carrollton	Dallas	Coppell	Garland
_N_	344	469	317	629
B-P	12.78	14.61	57.22	8.81
P-Value	0.01*	0.01*	0.00*	0.07
Null Hypothesis	Reject	Reject	Reject	Fail to be Rejected

* At 5% level of significance, the null hypothesis(the homoskedasticity assumption) is rejected

Except Garland county, the remaining 3 other counties has detected the heteroskedasticity. Therefore, the proper estimation is in order, using Weighted Least Squares(WLS), with proper weights sought by Breusch Pagan auxiliary regression. In this procedure, the weight by the natural log of $\hat{e}^2$ from the auxiliary regression is used in the WLS estimation rather than using $\hat{e}^2$ itself (which is used for Breusch Pagan heteroskedasticity test).

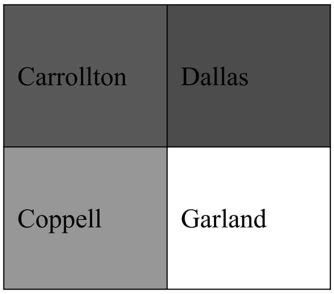

Figure 12. The Result of Wald Test

Finally, the result of Wald statistic and the corresponding critical value are presented in the Figure12 and Table6.

Table 6. The WaldStatistic

α	Critical Value	Carrollton vs. Dallas	Carrollton vs. Coppell	Carrollton vs. Garland	Dallas vs. Coppell	Dallas vs. Garland	Coppell vs. Garland
0.01	29.1	93.95*	31.46*	32.61*	79.94*	78.22*	48.01*

* At 1% level of significance, the null hypothesis(the parameters are same in two submarkets) is rejected.

As can be seen in the Figure12 and Table6, the homogeneous submarkets are grouped into 4 different housing submarkets. That is to say, there are no homogeneous submarkets in the research area, which is the different result of the Chow test. The critical value of chi-statistic is 29.1 at the 1%. All of the study area is above the critical value, which means that there are independent distinctive submarkets separately by themselves.

Ⅵ. Conclusions and Recommendations

It is important to for realtors and planners in their understanding neighborhood dynamics such as whether market segmentation is or not. The information presented in this study will bring systemization into the process of hedonic housing prices assessment by way of allowing to the differential contribution of various housing attributes in different neighborhood structures when the market segments do exist, which this study supports for. Therefore, they will be pricing houses in a much more reliable manner depends on neighborhood submarkets.

From a policy point of view, the estimated models could be used for to the tax officers who must reassess property at full market value each year. By knowing how the monetary contribution of each structural attribute varies across the study area, they can predict the effect of changing neighborhood quality on housing prices. In this way, they can applied to differential property tax rates according to the characteristics of the different neighborhood submarkets suppose if we try to calculate proper tax rate across neighborhood submarkets. Also, we can infer the area to which public service programs should be applied by way of allowing to vary hedonic housing prices across neighborhood submarkets.

The relationship between structural characteristics of housing and the social characteristics of housing should be investigated in more rigorous way by bring the social characteristics of housing into the hedonic equation directly.

Above all, the searching for spatial autocorrelation will be desirable in terms of both 'within' homogeneous boundary and 'between' homogeneous boundaries in order to capture the spatial dependence or spillover effect in near future.

Also, there is no escape the two dimensions in the real world, namely space and time. Accordingly, the estimation with consideration both space and time should be the most desirable estimation if we reflect the real situation.

It should be noted that the problem of heteroskedasticity is likely to be more common in cross-sectional than time-series data. In cross-sectional data, one usually deals with members of a population at a given point in time, such as individual consumers of their families, firms, industries, or geographic subdivision, such as state, country, or city, etc. Moreover, these members may be of different sizes, such as small, medium, or large firms or low, medium, or high income.

Any aggregate models are only meaningful if the underlying phenomenon is homogeneous across the units of observations when it comes to the application of urban housing market. Unless there is a homogeneous spatial process underlying the data, any aggregation will tend to be misleading. If this is not the case, both the heterogeneity and structural instability that are present should be accounted for in any aggregation scheme. This aspect of the modifiable area unit problem should be considered as a specification issue related to the form of spatial heterogeneity, and not solely as an issue determined by the spatial organization of the data (Anselin, 1988: 26-27).

The effect of heteroskedasticity on the test for homogeneity across samples has concentrated on the case of heteroskedasticity between samples but not within samples in the previous research. An assumption of between-sample heteroskedasticity explicitly states that the factor that divides the samples is the only one that generates differences in the spread of the disturbances. Such an assumption may be reasonable in some circumstances after all, the slopes of the equation are presumed to differ as well. The theory usually offers little guidance about the nature of the distribution of the disturbance however: assuming that heteroskedasticity exists but only as a function of sample-dividing factor is rather specific and too restrictive in many circumstances. Indeed, much of the present-day research on heteroskedasticity is focused on allowing generality and enhancing flexibility of estimation and testing procedures.

The most important conclusions of this paper can be summarized as follows: (1) Hedonic models of housing prices must be corrected for heteroskedasticity to ensure greater efficiency in the estimation of hedonic prices; (2) There are significant submarket differences in hedonic prices of housing attributes, implying that great care should be taken in the specification of the geographic units for which hedonic models are estimated; and (3) After careful submarket specification, the 3 distinctive submarkets out of 4 submarkets can be categorized based on the Chow test; and, there is no homogeneous submarkets in the study area based on the Wald test.

From the point of view of urban theory, the most important conclusion is that markets are segmented: hedonic prices differ among market segments, reflecting differences in the bid and offer curves for housing. It is recommended the dimensions of this segmentation be explored further. The numbers and types of neighborhoods into which housing markets should be disaggregated need to be established. The manner in which valuation of housing attributes varies across these market segments also needs to be established. The results then need to feedback into urban theory to generalize the homogeneous market concept is currently in use.

This paper addresses the question of homogeneity across a housing market. Studies often assume that the market is homogeneous within a given geographic boundary. Since such spatial units are drawn up for historical and political reasons, one may question whether it is fair to assume that there is a single economic market for housing. If there is segmentation, assuming homogeneity leads to uninformative estimates of housing price equations and public policy that lacks proper foundation.

As a final note, while this paper uses the housing price equation to define neighborhoods, there is no suggestion that we believe there is a uniquely defined concept of a neighborhood or a housing submarket. There are other social and spatial concepts of neighborhoods with different boundaries that are far more relevant to other behavioral models and to other policy issues, where a housing neighborhood simply would not be informative and therefore would not apply (Sawicki and Flynn 1996). Whenever such neighborhoods are determined by statistical procedures, including regression analysis, the insights gained in this paper apply.

【References】

Anselin, Luc. 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Press. 26-27.

_______. 1988. "Lagrange Multiplier Test Diagnostics for Spatial Dependence and Heterogeneity," Geographical Analysis 20, pp. 1-17.

Berry, Brian J.L. 1976. "Ghetto Expansion and Single-Family Housing Prices: Chicago, 1968-1972, "Journal of Urban Economics 3, pp. 397-423.

_______, and Bednarz, R. 1975. "A Hedonic Model of Prices and Assessments for Single-Family Homes in Chicago: Does the Assess or Follow the Market or the Market the Assessor?" Land Economics L1, pp. 21-40.

Breusch, T. S. & Pagan, A. R. 1979. A Simple Test for Heteroscedasticity and Random Coefficient Variation. Econometrica, 47-5. 1287-1294.

Cain, Glen C. 1976. The Challenge of Segmented Labor Market Theories to Orthodox Theory: A Survey. Journal of Economic Literature, 14-4. 1215-1257.

Can, A. 1990. "The Measurement of Neighborhood Dynamics in Urban House Prices," Economic Geography 66, pp. 254-272.

_______. 1992."Specification and Estimation of Hedonic Housing Price Models," Regional Science and Urban Economics 22, pp. 453-474.

Casetti, E. 1972. "Generating Models by the Expansion Method: Applications to Geographical Research," Geographical Analysis 4, pp. 81-91.

Chow, Gregory C. 1960. Tests of Equality between Sets of Coefficients in Two Linear Regressions. Econometrica, 28-3. 591-605.

Chung, Kyoun-Sup. 1994. Hedonic Models of Housing Prices: A Critical Appraisal of the Assumptions, Ph.D. dissertation, University of Texasat Dallas.

Cliff, A. and J. Ord. 1973. Spatial Autocorrelation, London.

Dubin, R. 1992. "Spatial Autocorrelation and Neighborhood Quality," Regional Science and Urban Economics 22, pp. 433-452.

Follain, James R. & Stephen Malpezzi. 1980. Dissecting Housing Value and Rent. Washington, D.C.: The Urban Institute.

Galster, George. 1996. The Analysis of Housing Sub-Markets and Filtering. Urban Studies, 33-10. 1797-1805.

Glejser, H. 1969. A New Test for Heteroskedasticity, Journal of the American Statistical Association, vol.64, 316-323.

Goldfeld S, & R. E. Quandt. 1972. Nonlinear Methods in Econometrics, North-Holland Publishing Company.

Green, W. 1992. Econometric Analysis: 2nded.

Griffith, D.A. 1987. Spatial Autocorrelation: A Primer. Washington, D.C.: Association of American Geographers.

Gujarati, D. 1988. Basic Econometrics, 2nd Ed. McGraw-Hill Book Company.

Halvorsen, R. and R. Palmquist. 1981. "The Interpretation of Dummy Variables in Semilogarithmic Equations," American Economic Review 71, p. 802.

Lancaster, K. 1966. A New Approach to Consumer Theory. Journal of Political Economy, 74. 132-157.

Muth, Richard F. 1975. Urban Economic Problems. NewYork, NY: Harper&Row.

Park, R. E. 1966. Estimation with Heteroskedastic Error Terms, Econometrica, vol.34, no.4.

Porell, Frank W. 1982. Models of Intraurban Residential Location. Dordrecht: Kluwer Academic Publishing Group.

Rosen, Sherwin. 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy, 72. 34-55.

Sawicki, David S. & Patrice, Flynn. 1996. Neighborhood Indicators: A Review of the Literature and an Assessment of Conceptual and Methodological Issues. Journal of the American Planning Association, 62-2. 165-183.

Schmidt and Sickles. 1977. "Some Further Evidence on the use of the Chow Test under Heteroskedasticity," Econometrica, 45, pp. 1293-1298.

Schnare, Ann & Raymond Struyk. 1976. Segmentation Ii Urban Housing Markets. Journal of Urban Economics, April 1976.

Strange, William. 1992. Overlapping Neighborhoods and Housing Externalities. Journal of Urban Economics, 32-1. 17-39.

Upton, G. and B. Fingleton. 1985. Spatial Data Analysis by Example. NewYork: Wiley.

Waddell, Paul, Brian J. L. Berry & Kyoun-Sup Chung. 1995. Variations in Housing Price Depreciation: the 'Taste For Newness' across Heterogeneous Submarkets. Urban Geography, 16-8. 722-733.

Waddell, Paul, Brian J. L. Berry & Irving Hoch. 1993. Residential Property Values in A Multinodal

Urban Area: New Evidence on the Implicit Price of Location. Journal of Real Estate Finance and Economics, 7. 117-141.

논문접수일: 2007.09.05/ 심사완료일: 2007.10.2