EVALUATING SUBDIVISION CHARACTERISTICS ON SINGLE FAMILY HOUSING VALUE USING HIERARCHICAL LINEAR MODELLING

Contact Author:
Woo-Jin Shin, Ph.D.
Korea Real Estate Research Institute
4F, 275 Yangjae-dong, Seocho-gu,
Seoul, 137-130, Korea

and

Jesse Saginor, Ph.D.
Texas A&M University
College Station, TX 77843-3137

and

Shannon Van Zandt, Ph.D.
Texas A&M University
College Station, TX 77843-3137
EVALUATING SUBDIVISION CHARACTERISTICS ON SINGLE FAMILY HOUSING VALUE USING HIERARCHICAL LINEAR MODELLING

Abstract
This research quantifies the financial effects of value creation concepts in relation to housing values in a single-family residential development using hierarchical linear modeling (HLM). As a result, value creation concepts such as sense of arrival, greenway connectivity, and the median length of cul-de-sac have positive effects on single family housing values while the number of accessible entrances and the median length of block variables have negative effects on single family housing values. These results indicate that higher values in a subdivision may result from smaller blocks, interconnected greenways, and a single entrance which provides a sense of arrival.
1. INTRODUCTION

In real estate development, developers increase real estate value by utilizing financial support and/or public incentives from the creation of various financing vehicles to maximize the value and return on the development\(^1\). In order to secure financing for the development, developers often require permission from public partners, private partners, or possibly both depending on financing venture structure. Financing vehicles focus mainly on the financial structure of the deal, but these vehicles do not focus on the actual physical development of the project. An alternative way to create real estate value is to incorporate value creation concepts in the design of real estate development products. Value creation concepts are defined as weighing the trade-offs of design in creating real estate value by maximizing design principles while minimizing financial costs (Sharkawy, 1994). Value creation concepts include the effects on the overall schematic design of residential subdivision development.

In a single family housing development, the value creation concepts are defined as variables that increase housing value at the subdivision level. For example, all homes in one subdivision may share a common value (e.g. the number of entrances, the presence of fountains, etc.) when the subdivision is designed and built by a developer. In this case, although such common values in a subdivision may influence individual housing values, the financial effects of the common values on the single family homes may not be easily identifiable using traditional modeling methods such as the hedonic price model. Until recently, the effects of value creation concepts on single family homes within subdivisions have not been extensively examined. Only a few recent peer-reviewed
papers discuss a few variables at the subdivision level that evaluate common value effects on single family housing. For example, Thorsnes (2002) found a positive relationship between the value of homes and the preserved area attached to the subdivisions. Guttery (2002) showed the negative effects of an alleyway in a subdivision on single family housing values.

In conventional housing valuation studies, structural, locational, environmental, and neighborhood attributes are considered as characteristics affecting single family housing values. It is important to note that previous research rarely showed the effects of subdivision characteristics on housing values, for neighborhood attributes differ from subdivision characteristics. The neighborhood attributes are generally measured on the basis of the neighborhood’s geographical boundaries. For example, neighborhood variables are measured on the basis of Census Tracts/ Block Groups/ Blocks, which are geographic units defined by the United States Census Bureau. These geographic units are used as proxies for neighborhood attributes. On the other hand, subdivision boundaries are defined by developers. Both census tracts and block groups are normally not equal to subdivision boundaries. Hence, neighborhood attributes may not reflect a developer’s unique design values or represent the precise value of the subdivisions. More importantly, a neighborhood may have multiple subdivisions that differ greatly, diluting the possible significance of subdivisions and the impact on real estate valuation. As a result, a subdivision should be considered the appropriate unit of analysis in the study of the effects of the unique value in each subdivision on single family housing values.
Traditional Hedonic Price Model (HPM) is only valid with one level of data. However, the HPM violates the independence of observation assumption with nested data. Unlike HPM, the purpose of this article is to demonstrate that the Hierarchical Linear Model (HLM) is an appropriate method that overcomes the independence of observation violation when the data are constructed with two levels rather than one level. The HLM accomplishes this task by making sub-models of the same number of the unit number of higher level, and determines the effects of both levels of variables on a dependent variable with the within-level effects, between-level effects, and the effects of interactions across levels. The primary purpose of this research is to determine the effects of value creation concepts on the single family housing values in single-family residential development at the subdivision level.

2. LITERATURE REVIEW

2.1 Value Creation Concepts and Subdivision Effects

The concept of value creation applies to the characteristics that distinguish each subdivision from other subdivisions. Sharkawy (1994) tracked several value creation concepts, such as sense of arrival, sense of tradition, security, etc., through 126 different projects built during 1978-1989 and documented in the Urban Land Institute’s project files. Among these concepts, sense of arrival, circulation system, walkability, and greenery can be extracted at the subdivision level. These four components could also be identified as “subdivision effects”. The peer-reviewed real estate research often overlooks the effects of the sense of arrival on single family housing values. The vista of
a subdivision entrance, which may lead neighbors to feel a sense of arrival, can be composed of several wide-ranging characteristics such as signage, divided curbing, a gate, walls, or landscaping. Several peer-reviewed papers used photographs to evaluate participants’ positive perceptions of landscape sites by assigning a score to each scene, even if the photographs were not related to housing value or subdivision. Buhyoff et al. (1982) evaluated perceptions of forest vista landscapes using picture scoring and found that the negative visual impact of insect damage was diminished by the presence of long viewing distances, thick forests, and hilly terrain. Yamashita (2002) examined the perception of water in the landscape and results showed that adults like to see dynamic aspects of water more often than do children. Tunstall et al. (2004) investigated children’s perceptions on a river landscape and found that children recognized the aesthetic appeal of trees in the river landscapes.

The circulation system and walkability characteristics are related to street design and pedestrian mobility including nodes, street lengths, cul-de-sacs, and connectivity. Song and Knaap (2003) found that single family housing values rise when the length of streets are longer, there are fewer street nodes in the neighborhood, and the neighborhood block size is smaller. Other researchers found that cul-de-sacs generated a value premium of nearly 30 percent compared to traditional grid street patterns (Asabere, 1990). Several papers also showed the importance of pedestrian-friendly environments to encourage physical activity and active life-styles (Randall and Baetz, 2001; Saelens et al., 2003). The researchers argue that walking is encouraged by continuous sidewalks and bike route systems, fewer dead-ends, and smaller blocks.
The greenery characteristics are related to park and greenway connections. A number of papers showed the positive effects of parks and the proximity to an open space or the types of open spaces on single family housing values (Bolitzer and Netusil, 2000; Lutzenhiser and Netusil, 2001; Geoghegan, 2002; Irwin, 2002).

### 2.2 Structural, Locational, and Neighborhood Characteristics

Structural characteristics include the characteristics of a house itself, such as the number of bedrooms, the number of bathrooms, the number of fireplaces, garage size, square footage of house, lot size, age of the building in years, pool, and number of stories. All of these variables are often statistically significant and positively related to the single family housing price except for the age of the building in years (Song and Knaap, 2003). In general, the building age variable is a proxy of the variable regarding the quality of the construction of the home because the building age variable is negatively related to sale price, since older homes have experienced depreciation over a longer period of time.

Several papers show the negative relationships between single family housing values and the geographic distance from amenities such as shopping centers, churches, highways, elementary public schools, and central business districts, except hazardous waste sites (Kiel, 1995; Diaz et al., 2008; Seo and Simons, 2009). The locational characteristics also include the impact of the distance from parks, open spaces, golf courses and/or greenways on single family housing values. Many papers show the positive effects of parks on single family housing values (Crompton, 2005). Even though
there is a positive effect on the value of properties backing up to a park, it is lower than the impact on single family properties a block or two away. Properties abutting a park were subjected to many nuisances such as noise and lights, etc. (Crompton, 2000). Li and Brown (1980) pointed out the negative effects of facing heavily-used park facilities on single family housing values. Others show that a single family house with the amenity of a view of parks or water features has a higher value than a house having an obstructed view of parks or water features (Benson et al., 1998). Golf courses also show positive effects on the value of single family homes. Golf course frontage or specific golf course types have a positive premium (Shultz and Schmitz, 2009).

Neighborhood characteristics reflect on characteristics such as the socio-economic status of neighboring residents, the quality of neighboring structures, the median income of the block group, population density in the block group, crime and vandalism, and the percent of individuals in the block group with a bachelor’s degree (Li and Brown, 1980; Simons et al., 1998; Ding et al., 2000; Seo and Simons, 2009). From the literature, it is clear that variables related to the degree of homogeneity of the neighborhood maintain a consistent relationship with housing values. Proportions of African-American households, poverty levels, population densities, and crime variables have negative relationships with single family housing values. On the other hand, median income and education variables show positive relationships with residential sales prices.

2.3 Hedonic Price Model (HPM) and Hierarchical Linear Model (HLM)
Hedonic Price Model (HPM) is a very popular method to examine the relationships of specific variables on housing value. However, HPM with hierarchical or nested data would not capture the multi-level nature of the relationships in two reasons (Garner and Raudenbush, 1991; Osborne, 2000; Raudenbush and Bryk, 2002; Uyar and Brown, 2007). First, an assumption of HPM, independence of observation, was violated when the hierarchical data were used as each house has a unique and fixed location. Moreover, houses within a subdivision would have very similar characteristics and homes within the same subdivisions cannot be independent of the other homes. When the assumption is violated, the HPM makes too small of a standard deviation and causes a higher probability of null hypothesis rejection (i.e., type I error). Next, HPM cannot be operated with two different units of analysis. In general, HPM is made by assigning subdivision level characteristics to all homes nested in the subdivision. In this case, the results violate the independence of observation. Another way is to aggregate housing level variables up to the subdivision level. Using the subdivision as the level of observation leads to difficulty in determining the variability of housing amenities and its effect on value. An additional issue is that the value of each house is changed to the average housing value of each subdivision, failing to account for differences with the subdivision.

These problems associated with the traditional HPM can be solved when the Hierarchical Linear Model (HLM) is applied with nested data (Osborne, 2000). Lower level units nested in a higher level unit share the same characteristics of the higher level unit. However, HLM does not violate an assumption of independent observation, for HLM decomposes the variance of the dependent variable into lower level effects (i.e.,
individual homes) after controlling for the effects of characteristics of the higher level (i.e., subdivision characteristics) in which each unit in the lower level is nested. Moreover, unlike HPM, different units of multi-level analysis can be used in HLM, for HLM is operated with regression models of the same number of the unit number of higher level. The HLM was originally utilized by researchers in the field of education and is widely used in the field of educational psychology (Raudenbush and Bryk, 2002). Even though the use of HLM is growing in a number of other fields, it is seldom used in evaluating the effects of various housing or neighborhood related variables on single family housing values. There are only a few papers in the peer-reviewed real estate literature which show how HLM can be used to evaluate housing prices. Brown and Uyar (2004) wanted to demonstrate how the HLM could be used to explain the inherent hierarchy in determining housing prices with only two independent variables. They used only one variable for each level: lot size in the housing level and median travel time to work in the neighborhood level. As a result, the HLM model showed that 1) neighborhoods with higher travel times have lower mean housing value; 2) the change in mean housing value associated with increases in land size is the same across neighborhoods; and 3) neighborhoods with higher travel times have a higher rate of increase in housing values associated with increases in land size. Unlike the traditional hedonic price model (HPM), researchers can create a model for each level separately and decide the portions of explained variance that occurs at each level.

Recently, Uyar and Brown (2007) tried to find out the effects of neighborhood zones and school zones on housing value. They made an affluence variable with six
socio-economic variables such as the percentage of owner-occupied, white, education, poverty, median income, and median housing value. The affluence variable was cross-classified with neighborhood zones and school zones. Affluence value of neighborhood zones and school-achievement scores of school zones were used as higher level variables. As a result, the affluence value of neighborhood zones and school-achievement scores of school zones account for statistically significant portion of variation of housing value. Even though the HLM model has only one neighborhood variable, this paper showed the way to apply HLM when a study area is covered with cross-classified area such as school zones and neighborhood zones.

3. METHODOLOGY

3.1 Study Sample

The study site was the City of College Station, Texas. The city is located in the east central part of Texas (See Exhibit 1,) and is the home of Texas A&M University. The study population included all single-family homes in College Station, totaling 10,617 single family homes. Of these 10,617 single family homes, 6,669 single family homes were nested in 122 subdivisions. These subdivision homes were selected as the sample population because they contained all the necessary parcel and subdivision information to be analyzed. Additionally, the parcels lacking information on the number of either bedrooms or bathrooms were excluded from the analyses. Hence, 6,562 single family homes nested in 85 subdivisions with a minimum sample size of 8 were used in this
study as they were the maximum number of data which fit the required number of observations for each level to get the statistical power of over .90.4

3.2 Variables

3.2.1 Dependent Variables

The dependent variable is the appraisal value of 6,562 single family homes in 2008, which was obtained from the Brazos County Appraisal District. Ideally, the market sale value would be the value used, but Texas state laws prevent the acquisition of sales data for single family homes from being made publicly available.5 In College Station, sales data are only accessible through either the Bryan/College Station Association of Realtors or the Real Estate Center at Texas A&M University but this data is prohibited to be provided to a person who does not have a real estate license. The issue of systematic
assessment error could arise when the assessor appraisal value is used for analysis instead of real sales data. A time lag between the time of the sales referenced by the assessor and the publishing date of assessor appraisal value causes the assessor appraisal value reflects a past sale price (Fisher et al., 1999). The systematic assessment error can cause underestimation or overestimation of appraisal value (Clapp and Giaccotto, 1992). The appraisal value is the best available data for this research, in spite of the systematic assessment error in using appraisal value. First, appraisal value is approximately 95% of the sale price based on a sampling of limited available data and is almost perfectly correlated with sales data\(^6\). The reference date of the appraisal value was December 2008. Second, sales data is not publicly available. Third, the appraised value of each house is commonly used as a proxy for its market value to identify the marginal effect of a particular characteristic on housing value using the hedonic pricing model when sales data is unavailable (Hendon, 1972; Berry and Bednarz, 1975; Seiler et al., 2001). Finally, a large portion of the systematic assessment error with a dependent variable of appraisal values can be reduced to negligible proportions by using a large sample size.

### 3.2.2 Independent Variables: Subdivision Level Variables

At the subdivision level, the sense of arrival variable was subjectively evaluated, and all other variables were objectively measured with ArcGIS v. 9.3. The sense of arrival was measured by photo evaluation. The entrance pictures of 85 subdivisions were evaluated sixty-one graduate students in the College of Architecture at Texas A&M University. The average age of participants was 27 years old and their majors were land
development, landscape architecture, urban planning, construction science, and architecture. The evaluation was conducted in a classroom by showing photographic slides for about 5 seconds to the participants. Each entrance was evaluated by assigning a sense of arrival rating number between 1(very low) and 5(very high) on a Sense of Arrival Rating Response Sheet. When a subdivision had more than one entrance, an entrance picture was randomly selected. After collecting evaluation scores from students, the sense of arrival values were calculated as mean values of all scores for each entrance photo (See Exhibit 2). In general, the entrance pictures with relatively high scores had wide entrances, visible signage, a large amount of trees, and a gate. As the subdivision entrance pictures were evaluated by more than two students, the consistency of the evaluations was tested by calculating the Intraclass Correlation Coefficient (ICC)\(^7\). In this study, the intraclass reliability had an ICC statistic of 0.951 (p<.001) at a 95% level of confidence and the ICC was close to 1.0, and larger than .80.\(^8\)

Exhibit 2 | Comparison of Sense of Arrival Scores of Four Subdivisions

(a) Castle Gate (4.27)  
(b) Windwood (4.09)
Circulation system & walkability variables were measured using Environment and Physical Activity: GIS Protocols Version 2.0 (University of Minnesota, 2005). The accessible entrance variable was measured by counting the number of accessible points through the boundary of the subdivision. Sidewalk connectivity was quantified by pedestrian lane length divided by road length. Bike-lane connectivity was calculated by bike lane length divided by road length. The Median length of cul-de-sac was measured by the median length of cul-de-sac, and the median length of a block was measured by median length of blocks in a subdivision.

Greenery characteristics refer to park connectivity and greenway connectivity variables. The park connectivity (or greenway connectivity) variable was encoded with a dummy variable where the value “1” meant that a park (or greenways) was located within the subdivision or a park (or greenways) was either attached directly to, or located across from, a subdivision. In this case, greenways did not include any parks. Greenway means narrow linear open space with connecting persons with places and providing a
pedestrian path, bicycle path, or recreational use (Lindsey, 2003). The definitions for all variables belonging to the five categories in the subdivision level are summarized in Exhibit 3.

### Exhibit 3 | Summaries of Variables in the Subdivision Level

<table>
<thead>
<tr>
<th>Variables</th>
<th>Definition</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense of Arrival</td>
<td>Average score of the subdivision entrance</td>
<td>Average score of the subdivision entrance’s scenic quality.</td>
</tr>
<tr>
<td>Sidewalk Connectivity</td>
<td>Average sidewalk system</td>
<td>Total sidewalk length / total street length (miles)</td>
</tr>
<tr>
<td>Bike-Lane Connectivity</td>
<td>Average bike lane system</td>
<td>Total bike-lane length / total street length (miles)</td>
</tr>
<tr>
<td>Median Length of Cul-De-Sac</td>
<td>Median length of cul-de-sac</td>
<td>Median length of cul-de-sac (miles)</td>
</tr>
<tr>
<td>Median Length of Block</td>
<td>Median length of block</td>
<td>Median length of block (miles)</td>
</tr>
<tr>
<td>Accessible Entrance</td>
<td>The number of accessible entrances to the subdivision</td>
<td></td>
</tr>
<tr>
<td>Park Connectivity</td>
<td>Accessible to near park</td>
<td>(Dummy Variable) Get “1” if a park is in, attached to, or across a road to the subdivision</td>
</tr>
<tr>
<td>Greenway Connectivity</td>
<td>Accessible to near greenways</td>
<td>(Dummy Variable) Get “1” if any greenways are in, attached to, or across a road to the subdivision (Not include parks)</td>
</tr>
</tbody>
</table>

### 3.2.3 Independent Variables: Housing Level Variables

The structural variables of each house were obtained from the Brazos County Appraisal District (BCAD). Most geographic information system (GIS) data, such as parcel, zoning, road, park, and geographic variables, were obtained from the city of College Station, Texas. Neighborhood characteristics were measured based on the 2000 U.S. census data under the assumption that U.S. census block data (e.g. population) are
evenly distributed in each U.S. census block. All homes within the same U.S. census unit had the same value of a neighborhood variable. The definitions for all variables belonging to the three categories in the housing level are summarized in Exhibit 4.

Exhibit 4 | Summaries of Variables in Housing Level

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Structural</strong></td>
<td></td>
</tr>
<tr>
<td>Lot Size</td>
<td>Area of parcel (square feet)</td>
</tr>
<tr>
<td>Bedroom</td>
<td>The number of bedrooms</td>
</tr>
<tr>
<td>Bathroom</td>
<td>The number of Bathrooms</td>
</tr>
<tr>
<td>Total Main Area</td>
<td>Area of 1st and 2nd floor (square feet)</td>
</tr>
<tr>
<td>Attached Garage</td>
<td>Area of attached garage (square feet)</td>
</tr>
<tr>
<td>Detached Garage</td>
<td>Area of detached garage (square feet)</td>
</tr>
<tr>
<td>All Porches</td>
<td>Area of open, glassed, and screened porch (square feet)</td>
</tr>
<tr>
<td>Building Age</td>
<td>Age of Single Family House (2008 – Built Year) (year old)</td>
</tr>
<tr>
<td>2nd Floor</td>
<td>(Dummy variable) The home is a 2 story house</td>
</tr>
<tr>
<td>Swimming Pool</td>
<td>(Dummy variable) The home has a swimming pool</td>
</tr>
<tr>
<td><strong>Locational</strong></td>
<td></td>
</tr>
<tr>
<td>Attach Golf</td>
<td>(Dummy variable) The home is adjacent to a golf course</td>
</tr>
<tr>
<td>Attach Park</td>
<td>(Dummy variable) The home is adjacent to a park</td>
</tr>
<tr>
<td>Across Park</td>
<td>(Dummy variable) The home is the opposite side of a park across a road</td>
</tr>
<tr>
<td>Cul-De-Sac</td>
<td>(Dummy variable) The home is on Cul-De-Sac</td>
</tr>
<tr>
<td>Corner</td>
<td>(Dummy variable) The home is on corner</td>
</tr>
<tr>
<td>Sport Facility</td>
<td>The number of sport facilities in the closest park</td>
</tr>
<tr>
<td>Sport Lighted Facility</td>
<td>The number of lighted sport facilities in the closest park</td>
</tr>
<tr>
<td>Net Dist_School</td>
<td>Network distance from the nearest elementary school</td>
</tr>
<tr>
<td>Net Dist_TAMU</td>
<td>Network distance from the nearest entrance of Texas A&amp;M University</td>
</tr>
<tr>
<td>Net Dist_Park</td>
<td>Network distance from the nearest park</td>
</tr>
<tr>
<td><strong>Neighborhood</strong></td>
<td></td>
</tr>
<tr>
<td>Population Density</td>
<td>(Census Block) Population per hectare</td>
</tr>
<tr>
<td>Income</td>
<td>(Census Block) Average income in 1999</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>(Census Block) Ratio of white alone on population</td>
</tr>
<tr>
<td>Tenure</td>
<td>(Census Block) Ratio of rental homes on occupied homes</td>
</tr>
<tr>
<td>Workable-age</td>
<td>(Census Block) Ratio of over 20 years old on population</td>
</tr>
<tr>
<td>Employment</td>
<td>(Census Block Group) Ratio of over 16 year employers on population</td>
</tr>
<tr>
<td>Education</td>
<td>(Census Track) Ratio of people with bachelor / grad-professional degree</td>
</tr>
</tbody>
</table>
3.3 Hierarchical Linear Model (HLM)

The Hedonic Price Model (HPM), mainly used to investigate variables influencing housing values, is not able to explain the inherent hierarchy. In the HLM, the basic principles underlying the hierarchical order for entry are the removal of confounding (or spurious relationships) and the causal priority of the research characteristics. Hence, each variable should be entered only after other variables have been input, which may cause a spurious relationship or compounding (Cohen et al., 2003).

3.3.1 Centering Procedure

A centering procedure is used to decide the location of variables in the housing level in HLM. There are two major centering procedures in HLM: grand-mean centering versus group-mean centering. The grand-mean centering is used to center the housing level variables around the grand-mean. For example, the number of bedrooms variable is centered by subtracting the mean number of bedrooms of all 6,562 homes. In this case, the housing level variables have a form of Equation 1.

\[
(X_{ij} - \bar{X}_j);
\]  

[1]

where i means i\textsuperscript{th} single family house;

j means j\textsuperscript{th} subdivision, \(X_{ij}\) means housing level variables;

\(\bar{X}_j\) represents the grand-mean.
The group-mean centering is used to center the housing level variables around the corresponding subdivision unit mean. For example, the number of bedroom variable is centered by subtracting the mean number of bedrooms from each subdivision. By the group-mean centering procedure, the intercept of the level 1 equation in HLM becomes the mean housing value of subdivision \( j \). In this case, the housing level variables have a form of Equation 2.

\[
(X_{ij} - \bar{X}_j); \quad [2]
\]

where \( \bar{X}_j \) represents the mean for subdivision \( j \).

Because this research is focusing on the subdivision level variables, the group-mean centering is adopted in this research with two reasons. First, by applying the group-mean centering procedure, the results in HLM can be interpreted easily, for each housing level variables could be interpreted relative to its subdivision mean (Campbell and Kashy, 2002). Next, the group-mean centering procedure has useful statistical advantage by reducing non-essential multicollinearity in this research. Aiken and West (1991) announced two types of multicollinearity. Essential multicollinearity exists when there are substantial correlations among independent variables; on the other hand, non-essential multicollinearity exists when there are higher order terms, such as the interaction term among independent variables. In general, the non-essential multicollinearity exists because HLM operates with the effects of interactions across levels as well as the within-level and between-level effects. However, when group-mean
centering procedure is conducted, a centered housing level variable score is no more correlated with the subdivision mean variable score (Hox, 1995).

### 3.3.2 Three Types of Hierarchical Linear Models

For this study, three types of HLMs were applied. The models were made with HLM v. 6.06 and SPSS v. 16 using the MIXED procedure; the HLM v. 6.06 was used to make the three hierarchical linear models, and SPSS v. 16 was used to identify the statistical significance of variances and covariances between housing level variables.

The analysis started with fitting the Random-Effect ANOVA Model (“REA model”) to determine the total amount of variability in the appraisal values within and between subdivisions. The REA model showed whether or not the HLM was necessary for analyzing the data in this study or if only the hedonic price model was enough. The REA model is the simplest possible hierarchical linear model and can be explained as:

\[
\text{Level 1: } \ln(\text{Appraisal})_{ij} = \beta_{0j} + e_{ij} \tag{3}
\]

\[
\text{Level 2: } \beta_{0j} = \gamma_{00} + u_{0j} \tag{4}
\]

\[
\text{Combined Model: } \ln(\text{Appraisal}) = \gamma_{00} + u_{0j} + e_{ij}; \tag{5}
\]

where
- \(i\) is the \(i^{\text{th}}\) single family house;
- \(j\) is the \(j^{\text{th}}\) subdivision;
- \(\beta_{0j}\) represents the mean appraisal value of the \(j^{\text{th}}\) subdivision;
- \(\gamma_{00}\) represents the mean appraisal value of all single family homes in College Station;
- \(u_{0j}\) represents a subdivision (level-2) effect;
- \(e_{ij}\) represents a house (level-1) effect.
The variance of the dependent variable can be explained as:

$$\text{Var} \left( \ln(\text{Appraisal}_{ij}) \right) = \text{Var}(u_{ij}) + \text{Var}(e_{ij}) = \text{Var}(u_{ij}) + \text{Var}(e_{ij}) = \tau_{00} + \sigma^2;$$  \[6\]

where $\sigma^2$ is the within-subdivision variability, and $\tau_{00}$ represents the between-subdivision variability.

The next step in this research was to make the Random-Coefficient Regression Model ("RCR model"), which was useful in identifying statistically significant variables at the housing level. The RCR model was consistent with two levels (Equation 7 and 8). The first four coefficients in Equation 7 were specified as random in the subdivision level model; on the other hand, the last three coefficients had only a fixed effect. It just indicated that the effect of building age, attach to a golf course, and the existence of swimming pool variables on housing value did not vary across the 85 subdivisions.

Level 1: \[Y_{ij} = \beta_{0j} + \beta_{1j}(\ln\text{TMA}_C) + \beta_{2j}(\ln\text{BATH}_C) + \beta_{3j}(\ln\text{AP}_C) + \beta_{4j}(\text{BA}_C) + \beta_{5j}(\text{AG}_C) + \beta_{6j}(\text{GOLF}) + \beta_{7j}(\text{POOL}) + e_{ij}\]  \[7\]

Level 2: \[\beta_{qj} = \gamma_{q0} + u_{qj} \text{ for } q = 0, 1, 2, 3, 4,\]

\[\beta_{qj} = \gamma_{q0} \text{ for } q = 5, 6, 7;\]  \[8\]

where $Y_{ij}$ was the log of appraisal value of the $i^{th}$ single family house in the $j^{th}$ subdivision;

$\ln\text{TMA}_C$ was the group-mean centered log of total main area;

$\ln\text{BATH}_C$ was the group-mean centered log of the number of bathrooms;

$\ln\text{AP}_C$ was the group-mean centered log of all porches;

$\text{BA}_C$ was the group-mean centered building age;

$\text{AG}_C$ was the group-mean centered attached garage;
GOLF was a dummy variable for attachment to a golf course; POOL was a dummy variable for the existence of a swimming pool; $\gamma_{q0}$ was the mean value for each subdivision effect; $e_{ij}$ was an error term

$\sigma^2$, the variance of $e_{ij}$, represented the residual variance at level one that remained unexplained after considering the homes’ total main area, the number of bathrooms, all porches, building age, attached garage, attachment to a golf course, and the existence of swimming pool.

The final HLM was the Intercepts- and Slopes-as-Outcomes Model (“ISO model”), which included all statistically significant variables in both the housing level and the subdivision level. The previous RCR model shows that seven housing level variables had a significant relationship with appraisal values. All statistically significant level-1 variables in the RCR model should remain at least at a fixed effect in the housing-level model of the ISO model (Raudenbush and Bryk, 2002). Hence, the group-mean centered log of total main area, the group-mean centered log of the number of bathrooms, the group-mean centered log of all porches, the group-mean centered building age, the group-mean centered attached garage, the attachment to a golf course, and the existence of a swimming pool variables were added in the housing level model (see Equation 9).

It was recommended that the group-means of level-1 variables needed to be reintroduced into the macro level model when the group-mean centering procedure was used (Kreft et al., 1995). The reason was that this action compensated the removed group-mean effects caused by the group-mean centering of the level-1 variables. Hence, three group-means – the log of total main area group-mean, the log of the number of
bathrooms group-mean, and the attached garage group-mean – were added in the subdivision level model (Equation 10). However, the group-means of all porches and building age variables were not included, because the two group-mean variables were not statistically significant.

In the subdivision level model in the final model, five variables were added. The variables were sense of arrival (SENOFARR), the median length of block (lnMEDBLO), the median length of cul-de-sac (lnMEDCUL), the number of accessible entrances (ACCENT), and greenway connectivity (GRECON). The joint effects of three subdivision level variables – sense of arrival, the number of accessible entrances, and the log of total main area group-mean – on two housing level variables – the group-mean centered log of total main area and the group-mean centered attached garage – were modeled. Statistically significant interactions among subdivision level variables were also added. The intercepts- and slopes-as-outcomes model can be explained as:

\[
\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j}(\lnTMA_{C}) + \beta_{2j}(\lnBATH_{C}) + \beta_{3j}(\lnAP_{C}) + \beta_{4j}(BA_{C}) + \beta_{5j}(AG_{C}) + \beta_{qj}(GOLF) + \beta_{7j}(POOL) + \epsilon_{ij} \tag{9}
\]

\[
\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}(SENOFARR) + \gamma_{02}(\lnMEDCUL) + \gamma_{03}(\lnMEDBLO) + \gamma_{04}(ACCENT) + \gamma_{05}(ACCGRE) + \gamma_{06}(\lnTMA_{GM}) + \gamma_{07}(\lnBATH_{GM}) + \gamma_{08}(AG_{GM}) + \gamma_{09}(\lnTMA_{GM} \cdot GRECON) + \gamma_{010}(\lnMEDCUL \cdot ACCENT) + \gamma_{011}(\lnMEDBLO \cdot GRECON) + u_{0j} 
\]

\[
\beta_{3j} = \gamma_{10} + \gamma_{11}(SENOFARR) + \gamma_{12}(ACCENT) + u_{1j} 
\]

\[
\beta_{qj} = \gamma_{q0} + u_{qj} \quad \text{for } q = 2, 3, 4, 
\]
\[ \beta_{5j} = \gamma_{50} + \gamma_{51}(\text{SENOFARR}) + \gamma_{52}(\ln\text{TMA\_GM}) \]

\[ \beta_{qj} = \gamma_{q0} \quad \text{for } q = 6, 7 \quad [10] \]

Combined [Fixed Part]: \( Y_{ij} = \)
\[ \gamma_{00} + \gamma_{01}(\text{SENOFARR}) + \gamma_{02}(\ln\text{MEDCUL}) + \gamma_{03}(\ln\text{MEDBLO}) + \gamma_{04}(\text{ACCENT}) + \gamma_{05}(\text{ACCGRE}) + \gamma_{06}(\ln\text{TMA\_GM}) + \gamma_{07}(\ln\text{BATH\_GM}) + \gamma_{08}(\text{AG\_GM}) + \gamma_{09}(\ln\text{TMA\_GM} \times \text{GRECON}) + \gamma_{10}(\ln\text{MEDCUL} \times \text{ACCENT}) + \gamma_{11}(\ln\text{MEDBLO} \times \text{GRECON}) + \gamma_{12}(\ln\text{TMA\_C}) + \gamma_{13}(\text{SENOFARR}) \times (\ln\text{TMA\_C}) + \gamma_{14}(\text{ACCENT}) \times (\ln\text{TMA\_C}) + \gamma_{20}(\ln\text{BATH\_C}) + \gamma_{21}(\ln\text{AP\_C}) + \gamma_{22}(\text{BA\_C}) + \gamma_{23}(\text{AG\_C}) + \gamma_{24}(\text{AG\_C}) \times (\ln\text{TMA\_GM}) + \gamma_{25}(\text{GOLF}) + \gamma_{70}(\text{POOL}) \quad [11] \]

Due to the five level-2 random effects (\( u_{ij} \) in Equation 10), the variances and covariances matrix can be measured. In Exhibit 4, the random effects find out the difference of intercepts or slopes of between the regression line of a subdivision and the overall model. Hence, the random effects show significant information about the pattern of regression lines of subdivisions through the relationship among dependent variables and four independent variables - total main area, the number of bathrooms, all porches, and building age - in the ISO Model. First, the within-subdivision variability (\( \sigma^2 \)) tests the difference between the appraisal value of the \( i^{th} \) house and the mean appraisal value of a subdivision including the \( i^{th} \) house. \( \sigma^2 \) shows that the appraisal values of homes nested in a subdivision are not the same as the mean appraisal value of the subdivision. Second, in the variance-covariance matrix (\( T \)), \( \tau_{00} \) tests the difference between the intercept value of
the regression line of \( j^{th} \) subdivision and the intercept value of the regression line of the overall model. Because there are 85 subdivisions, 85 regression lines could be identified. \( \tau_{00} \) shows that the intercepts of regression lines of 85 subdivisions are not the same as the intercept value of regression line of the overall model. Third, \( \tau_{11}, \tau_{22}, \tau_{33}, \) and \( \tau_{44} \) shows that the slopes of regression lines of 85 subdivisions are not the same as the slope of regression line of the overall model when the \textit{appraisal value} was on the y-axis and \textit{total main area}, \textit{the number of bathrooms}, and \textit{building age} were on the x-axis. Finally, \( \tau_{10} \) shows that there is no correlation between intercepts and slopes of regression lines of all subdivisions when the \textit{appraisal value} is on the y-axis and \textit{total main area} is on the x-axis. In Exhibit 5, the degree of freedom of each variable was different, for the number of subdivision units that had sufficient data for computation of the variable are different. Even though the degree of freedom of each variable was used to calculate Chi-square statistics, variance components and fixed effects were based on all 85 subdivisions. Hence, the random effects and fixed effects of all data can be interpreted with variance components in Exhibit 5 and coefficients in Exhibit 9 regardless of the difference of the degrees of freedom of variables.

<table>
<thead>
<tr>
<th>Exhibit 5</th>
<th>Estimated Random Effects on the ISO Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance Component</td>
<td>D.F.</td>
</tr>
<tr>
<td>Mean Housing Value, ( \tau_{00} )</td>
<td>.0065</td>
</tr>
<tr>
<td>LN (Total Main Area_Centered), ( \tau_{11} )</td>
<td>.0152</td>
</tr>
<tr>
<td>LN (Bathroom_Centered), ( \tau_{22} )</td>
<td>.0027</td>
</tr>
<tr>
<td>LN (All Porches_Centered), ( \tau_{33} )</td>
<td>.0000</td>
</tr>
<tr>
<td>(Building Age_Centered), ( \tau_{44} )</td>
<td>.0000</td>
</tr>
</tbody>
</table>
Level-1 effect, $\sigma^2$ .0054

<table>
<thead>
<tr>
<th>Correlation Among Subdivision Effects</th>
<th>Mean Housing Value</th>
<th>LN (Total Main Area_Centered)</th>
<th>LN (Bathroom_Centered)</th>
<th>LN (All Porches_Centered)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LN (Total Main Area_Centered)</td>
<td>$\tau_{10} = .0026$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LN (Bathroom_Centered)</td>
<td>$\tau_{20} = -.0013$</td>
<td>$\tau_{21} = -.0012$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LN (All Porches_Centered)</td>
<td>$\tau_{30} = -.0001$</td>
<td>$\tau_{31} = .0004$</td>
<td>$\tau_{32} = -.0001$</td>
<td></td>
</tr>
<tr>
<td>(Building Age_Centered)</td>
<td>$\tau_{40} = -.0001$</td>
<td>$\tau_{41} = -.0002$</td>
<td>$\tau_{42} = -.0001$</td>
<td>$\tau_{43} = -.0000$</td>
</tr>
</tbody>
</table>

4. RESULTS

4.1 Descriptive Statistics

The characteristics of the continuous variables at the housing level are summarized in Exhibit 6. The mean appraisal value of single family homes was $177,740. On average, single family homes had three bedrooms and two bathrooms, and the age was 19 years. The mean distances from the nearest elementary school and the nearest park to each house were 1.2 mile and 0.4 mile, respectively. The mean number of sport facilities in the nearest park from each house was five.

Exhibit 6: Descriptive Statistics of Continuous Variables in the Housing level

<table>
<thead>
<tr>
<th>Variable</th>
<th>Count</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appraisal Value</td>
<td>6,562</td>
<td>31,080.00</td>
<td>778,060.00</td>
<td>177,740.48</td>
<td>74,553.90</td>
</tr>
<tr>
<td>Lot Size</td>
<td>6,562</td>
<td>1,874.73</td>
<td>60,000.00</td>
<td>11,827.70</td>
<td>7,416.21</td>
</tr>
<tr>
<td>Total Main Area</td>
<td>6,562</td>
<td>702.00</td>
<td>7,056.00</td>
<td>1,995.51</td>
<td>655.95</td>
</tr>
<tr>
<td>Number of Bedrooms</td>
<td>6,562</td>
<td>1.00</td>
<td>6.00</td>
<td>3.48</td>
<td>.59</td>
</tr>
<tr>
<td>Number of Bathrooms</td>
<td>6,562</td>
<td>1.00</td>
<td>5.50</td>
<td>2.24</td>
<td>.54</td>
</tr>
<tr>
<td>Building Age</td>
<td>6,562</td>
<td>1.00</td>
<td>117.00</td>
<td>18.72</td>
<td>13.65</td>
</tr>
<tr>
<td>Attached Garage</td>
<td>6,562</td>
<td>.00</td>
<td>1,685.00</td>
<td>411.81</td>
<td>199.85</td>
</tr>
<tr>
<td>Detached Garage</td>
<td>6,562</td>
<td>.00</td>
<td>1,500.00</td>
<td>43.61</td>
<td>153.25</td>
</tr>
<tr>
<td>All Porches</td>
<td>6,562</td>
<td>.00</td>
<td>2,128.00</td>
<td>144.22</td>
<td>142.77</td>
</tr>
</tbody>
</table>
Next, the characteristics of the continuous variables at the subdivision level were summarized in Exhibit 7. Data showed that the mean number of entrances of subdivisions was four, and, on average, the median length of cul-de-sac and the median length of blocks were 0.01 miles and 0.5 miles, respectively. The mean number of intersections per mile was five in subdivisions in College Station, Texas.

<table>
<thead>
<tr>
<th>Sense of Arrival</th>
<th>Count</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessible Entrance</td>
<td>85</td>
<td>1.00</td>
<td>10.00</td>
<td>4.38</td>
<td>2.46</td>
</tr>
<tr>
<td>Median Length of Cul-De-Sac</td>
<td>85</td>
<td>.00</td>
<td>.16</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>Median Length of Blocks</td>
<td>85</td>
<td>.20</td>
<td>1.88</td>
<td>.51</td>
<td>.17</td>
</tr>
<tr>
<td>Sidewalk Connectivity</td>
<td>85</td>
<td>.00</td>
<td>.42</td>
<td>.06</td>
<td>.10</td>
</tr>
<tr>
<td>Bike-Lane Connectivity</td>
<td>85</td>
<td>.00</td>
<td>6.65</td>
<td>.24</td>
<td>.42</td>
</tr>
</tbody>
</table>

In general, a number of papers in regard to analyzing property values had most commonly used a log transformation when data were not normally distributed. The
transformation of the data is helpful to reduce the impact of outliers and for converting not normally distributed data to normally distributed data. The log transformation is performed by taking the logarithm of the dependent variable or the logarithm of both independent and dependent variables. To examine the data’s normality, the skewness and kurtosis for each variable were used. Compared to the skewness and kurtosis of the original data, the log-transformed variables showed much lower values of skewness and kurtosis. The dependent variable (appraisal value) and eight continuous variables at the housing level (the number of bathrooms, total main area, detached garage, lot size, all porches, the network distance from the nearest park, population density, and ethnicity) and four continuous variables at the subdivision level (the median length of cul-de-sac, the median length of blocks, sidewalk connectivity, and bike-lane connectivity) were log-transformed to fit to normal distribution and to be more easily interpreted. Even though the maximum value of the porch variable looks quite large, the variable shows normal distribution when a logarithmic form was applied. On the other hand, three variables in the housing level - detached garage, lot size, and ethnicity are not normally distributed and show some outliers.

4.2 Hierarchical Linear Model Results

4.2.1 The Random-Effect ANOVA Model (“REA model”)  

The interclass correlation coefficient (ICC) is the most import parameter in the REA model and its expression is as below;

\[ \rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \]  

[12]
In Exhibit 8, the estimate of residual, the within-subdivision variability ($\sigma^2$), was .04, and the estimate of the intercept, the between-subdivision variability ($\tau_{00}$), was .128. The ICC evaluates the proportion of the variance in the dependent variable (the log of appraisal value) that was between the level-2 units (subdivisions). The ICC can be calculated by equation 12 as shown below:

$$\rho = \frac{\tau_{00}}{(\tau_{00} + \sigma^2)} = \frac{0.128}{(0.128 + 0.044)} = 0.744$$

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variance Component</th>
<th>D.F.</th>
<th>Chi-Square</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.128</td>
<td>84</td>
<td>14981.167</td>
<td>.000</td>
</tr>
<tr>
<td>Residual</td>
<td>.044</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to the ICC, it can be said that about 75% of the variance in the dependent variable (log of appraisal value) is between subdivisions. To verify whether the HLM is needed for the dataset statistically, Muthen and Satorra (1995) suggests the Design Effect, which is the ratio of the total number of homes required using subdivision level randomization to the number required using housing level randomization. The design effect can be calculated by equation 13.

$$Design\ Effect = 1 + (\text{average cluster size} - 1) \times ICC = 1 + (77 - 1) \times 0.744 = 57.5$$ [13]
The average cluster size in this research was 77 (= the number of total homes / the number of subdivisions = 6,562 / 85 = 77), and the Design Effect was 57.5. Maas and Hox (2002) mentioned that using single level analysis is likely to lead to biased results if the Design Effect was larger than 2. Because the Design Effect of the data (57.5) was larger than 2, the HLM would give unbiased results instead of a single level model.

4.2.2 The Random-Coefficient Regression Model (“RCR model”)

The RCR model represents the structural, locational, and neighborhood distribution of the appraisal value in each of the 85 subdivisions. It is an important early step in a hierarchical linear model in identifying a range of useful housing level variables for the sequentially final model including both housing level and subdivision level. The statistically significant housing level variables in this model should be used in the next model. The final RCR model was developed using statistically significant seven variables among all the possible housing level variables. As a result, the appraisal value for single family home i in subdivision j was regressed on the log of total main area, the log of the number of bathrooms, the log of all porches, attached garage, building age, attach to a golf course, and the existence of swimming pool. The five continuous variables – the log of total main area, the log of the number of bathrooms, the log of all porches, attached garage, and building age – were all group-mean centered with the form of \( \ln(X_{ij}) - \ln(\overline{X}_j) \).

4.2.3 The Intercepts- and Slopes-as-Outcomes Model (“ISO model”)
The results for the ISO model are presented in Exhibit 8. Goodness of fit of HLM can be identified with the likelihood ratio test which is conducted with the absolute value of the difference between two “deviance (-2 log likelihood)” statistics of two HLMs.\textsuperscript{11} The deviance shows the quality of model fitness. The absolute value has a chi-square distribution with the number of degrees of freedom equal to the difference between the degrees of freedom of the two models. The deviance statistics of REA, RCR, and ISO model are -1,494.70, -14,706.54, and -14,893.98, respectively. The degrees of freedom of REA, RCR, and ISO models are 2, 24, and 39, respectively. The ISO model demonstrated goodness of fit (Chi-square for change in HLM fit =13,439.28, 37 d.f., $p \approx 0$) compared to REA model. It shows that the ISO model has statistically significant explanatory power compared to the REA model. That is, the subdivision level and housing level variables in the ISO model explains significant amount of variance in the appraisal value of houses. In addition, the ISO model showed goodness of fit (Chi-square for change in HLM fit =187.54, 15 d.f., $p \approx 0$) compared to RCR model. In other words, the subdivision level variables in the ISO model give explanation to significant amount of variance in the appraisal value of houses.

In Exhibit 9, the fixed effects can be interpreted in a familiar way with traditional regression model. Each variable is tested whether it is significantly different from zero. Coefficients on the seven variables (total main area, the number of bathrooms, all porches, building age, attached garage, attach to a golf course, and the existence of swimming pool) in the housing level were positive and statistically significant as expected except the building age variable. The number of bathrooms, all porches,
building age, attach to a golf course, and the existence of swimming pool variables were interpreted easily because there were no interaction terms with the subdivision level variables.

Exhibit 9 | Estimated Fixed Effects on the ISO Model

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t Ratio</th>
<th>d.f.</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subdivision Mean Housing Value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BASE, $\gamma_0$</td>
<td>4.2465</td>
<td>0.4616</td>
<td>9.200</td>
<td>73</td>
<td>0.000</td>
</tr>
<tr>
<td>Sense of Arrival, $\gamma_1$</td>
<td>0.0518</td>
<td>0.0198</td>
<td>2.614</td>
<td>73</td>
<td>0.011</td>
</tr>
<tr>
<td>Ln(median length of Cul-De-Sac), $\gamma_2$</td>
<td>0.0444</td>
<td>0.0101</td>
<td>4.393</td>
<td>73</td>
<td>0.000</td>
</tr>
<tr>
<td>Ln(median length of block), $\gamma_3$</td>
<td>-0.1364</td>
<td>0.0326</td>
<td>-4.190</td>
<td>73</td>
<td>0.000</td>
</tr>
<tr>
<td>Number of Accessible Entrance, $\gamma_4$</td>
<td>-0.0536</td>
<td>0.0137</td>
<td>-3.907</td>
<td>73</td>
<td>0.000</td>
</tr>
<tr>
<td>Greenway Connectivity, $\gamma_5$</td>
<td>1.7052</td>
<td>0.7116</td>
<td>2.396</td>
<td>73</td>
<td>0.019</td>
</tr>
<tr>
<td>Ln(total main area group-mean), $\gamma_6$</td>
<td>0.9643</td>
<td>0.0664</td>
<td>14.514</td>
<td>73</td>
<td>0.000</td>
</tr>
<tr>
<td>Ln(number of bathroom group-mean), $\gamma_7$</td>
<td>0.3937</td>
<td>0.0976</td>
<td>4.032</td>
<td>73</td>
<td>0.000</td>
</tr>
<tr>
<td>Attached garage group-mean, $\gamma_8$</td>
<td>0.0003</td>
<td>0.0001</td>
<td>4.082</td>
<td>73</td>
<td>0.000</td>
</tr>
<tr>
<td>Ln(total main area group-mean)* Greenway Connectivity, $\gamma_9$</td>
<td>-0.1999</td>
<td>0.0932</td>
<td>-2.145</td>
<td>73</td>
<td>0.035</td>
</tr>
<tr>
<td>Ln(median length of Cul-De-Sac)* Number of Accessible Entrance, $\gamma_{10}$</td>
<td>-0.0099</td>
<td>0.0025</td>
<td>-4.022</td>
<td>73</td>
<td>0.000</td>
</tr>
<tr>
<td>Ln(median length of block)* Greenway Connectivity, $\gamma_{11}$</td>
<td>0.1952</td>
<td>0.067</td>
<td>2.914</td>
<td>73</td>
<td>0.005</td>
</tr>
<tr>
<td><strong>Total Main Area</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BASE, $\gamma_{10}$</td>
<td>0.3529</td>
<td>0.0963</td>
<td>3.663</td>
<td>82</td>
<td>0.001</td>
</tr>
<tr>
<td>Sense of Arrival, $\gamma_{11}$</td>
<td>0.1001</td>
<td>0.0303</td>
<td>3.306</td>
<td>82</td>
<td>0.002</td>
</tr>
<tr>
<td>Number of Accessible Entrance, $\gamma_{12}$</td>
<td>0.0210</td>
<td>0.0066</td>
<td>3.168</td>
<td>82</td>
<td>0.003</td>
</tr>
<tr>
<td><strong>The Number of Bathrooms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BASE, $\gamma_{20}$</td>
<td>0.0537</td>
<td>0.0104</td>
<td>5.150</td>
<td>84</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Porches</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BASE, $\gamma_{30}$</td>
<td>0.0090</td>
<td>0.0013</td>
<td>7.071</td>
<td>84</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Building Age</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BASE, $\gamma_{40}$</td>
<td>-0.0066</td>
<td>0.0009</td>
<td>-7.775</td>
<td>84</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Attached Garage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BASE, $\gamma_{50}$</td>
<td>0.00133</td>
<td>0.0002</td>
<td>7.037</td>
<td>6,539</td>
<td>0.000</td>
</tr>
<tr>
<td>Sense of Arrival, $\gamma_{51}$</td>
<td>0.00004</td>
<td>0.0000</td>
<td>2.816</td>
<td>6,539</td>
<td>0.005</td>
</tr>
</tbody>
</table>
4.3 Interpretation of differently transformed independent variables

As the dependent variable was log transformed, coefficients on the independent variables would be interpreted differently based on the form of the variables (Asteriou and Hall, 2007). First, when the independent variable was log transformed as well, the coefficient of the variable should be interpreted as elasticity\(^{12}\). Second, when the independent variable was not transformed, the coefficient of the variable should be interpreted as a relative change in dependent variables on an absolute change in the dependent variable\(^{13}\). Finally, when the independent variable was an untransformed dummy variable, the true proportional change in the dependent variable resulting from a unit change in a dependent variable should be calculated with the equation of “100(exp(b1)-1)” (Halvorsen and Palmquist, 1980; Hardy, 1993)\(^{14}\).

4.3.1 Subdivision Level Variables (Value Creation Concepts)

The sense of arrival variable shows a positive relationship with the housing appraisal value as expected, and was statistically significant. Results showed that the
average appraisal value of homes increased by 5.2% when a unit of the *sense of arrival* score rise. For example, when a house valued at $177,800 (the mean appraisal value) and one unit of *sense of arrival* score rose, the housing value would increase on average by $9,250, other things being constant.

Among circulation and walkability characteristics, *the number of accessible entrances, the median length of cul-de-sac, and the median length of block* variables were statistically significant. The *median length of cul-de-sac* variable had a statistically significant and positive relationship with appraisal value. On the other hand, *the number of accessible entrance* and *the median length of block* variables had negative relationship with appraisal value, as was expected. Because these three variables were related to intersection terms, the variables should be interpreted considering the intersection terms.

*The number of accessible entrances* variable was related to *the log of median length of cul-de-sac* variable. For one more accessible entrance to a subdivision, if the subdivision had the mean value of *the median length of cul-de-sac* of 0.014, the appraisal value of a house, which was nested in the subdivision and had the mean housing value of the subdivision, decreased by 1.09%. For example, a house valued at $177,800 was in a subdivision which had four *accessible entrances* and *the median length of cul-de-sac* was 0.014 miles. If the subdivision had five *accessible entrances*, the housing value would decrease by $1,940, other things being constant.

*The log of median length of cul-de-sac* variable was related to the number of accessible entrances variable. For a 1% increase in *the median length of cul-de-sac* of a subdivision where the subdivision had four *accessible entrances*, then the appraisal
value of a house nested in the subdivision with the mean value of the subdivision, increased by 0.004%. For example, a house valued at $177,800 was in a subdivision with a median length of cul-de-sac of 0.01 miles and four accessible entrances. If the subdivision had 0.0101 miles of the median length of cul-de-sac, the housing value would increase by $7, other things being constant.

The log of median length of block variable should be interpreted with the greenway connectivity variable. For a 1% increase of the median length of block of a subdivision, if the subdivision connected to greenway trails, the appraisal value of a house, which was nested in the subdivision and had the mean housing value of the subdivision, decreased by 0.059%. For example, a house valued at $177,800 was in a subdivision with the median length of block of 1 mile and connected to any greenway trails. If the subdivision had 1.01 miles of the median length of block, the housing value would decrease by $105, other things being constant.

In greenery characteristics, the greenway connectivity variable was statistically significant and positive. The greenway connectivity variable should be interpreted with both the log of total main area group-mean and the log of median length of block variables, and the log of total main area group-mean variable was related to greenway connectivity variable. The mean total main area of subdivisions was about 2,000 square feet, and the mean of the median length of blocks was 0.5 miles. Based on these mean values, it could be said that greenway connectivity had a positive effect on housing value on average.
If a home nested in a subdivision has a mean total main area of 2,000 square feet and a mean median length of block of 0.5 miles, and the subdivision was recently connected to greenway trails, the appraisal value of the house would increased by 5.177% because of the connection to greenway trails. For example, a house nested in a subdivision with disconnection to greenway trails was valued at $177,800. If the subdivision was connected to greenway trails recently, the housing value would increase by $9,205, other things being constant.

4.3.2 Housing Level Variables

The main point of this paper is to highlight subdivision level data, and housing level data are included only for control purposes. First, among structural characteristics, total main area, attached garage, all porches, building age, the existence of swimming pool, and the number of bathrooms variables were statistically significant. All significant variables had a positive relationship with appraisal value except the building age variable. Contrary to expectation, the number of bedrooms variable was not statistically significant; however, it was not an extraordinary result, for several models in the literature only included the main area (living area) variable without the number of bedrooms or the number of bathrooms variables (Mooney, 2001; des Rosiers et al., 2002). Next, in locational characteristics, only attach to a golf course variable was statistically significant. Finally, no variables in neighborhood characteristics were statistically significant. This might be explained by different geographic boundaries of subdivisions and neighborhoods. The neighborhood attributes were measured on the
basis of census tracks/ block group/ block which were not exactly equal to the subdivision boundaries. Each house was nested in a subdivision; that is, homes were considered as units in any corresponding subdivisions. However, homes within the same census unit should have the same scores for neighborhood variables, and each census unit existed over several subdivisions. This geographical confusion might cause the neighborhood variables to not be statistically significant in HLM.

5. CONCLUSION AND DISCUSSION

5.1 Conclusion

This study identified useful variables in the subdivision level to evaluate value creation concepts to maximize single family housing appraisal value. Especially, as a pilot study, Hierarchical Linear Model (HLM) is used to measure the relationship between housing values and variables in both the housing level and subdivision level. The findings highlight that sense of arrival, circulation system and walkability, and greenery characteristics have important and significant positive effects on housing appraisal values. Consequently, the study recommends that five value creation concepts should be considered in maximizing housing values. Sense of arrival, the median length of cul-de-sac, and greenway connectivity variables have positive effects on single family housing values; on the other hand, the number of accessible entrance and the median length of block variables have negative effects on single family housing values. In view of percentage change of a housing value per one percent or one unit change of a
subdivision variable from the priority list in Exhibit 10, developers may consider mainly greenway connectivity (5.2%) and sense of arrival (5.2%) variables in their residential development. Next, the number of accessible entrance (1.5%), the median length of block (0.02%), and the median length of cul-de-sac (0.004%) variables also will be considered.

Exhibit 10 | Summaries of Significant characteristics and Variables

<table>
<thead>
<tr>
<th>Level</th>
<th>characteristics</th>
<th>Variable</th>
<th>Direction of Effect</th>
<th>% Change</th>
<th>$ Change for $177,800 house</th>
<th>Note</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subdivision</td>
<td>Sense of Arrival</td>
<td>Sense of Arrival</td>
<td>+</td>
<td>5.18 %</td>
<td>$9,000</td>
<td>U</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Median Length of Cul-De-Sac</td>
<td>+</td>
<td>0.004%</td>
<td>$7</td>
<td>P</td>
<td>5</td>
</tr>
<tr>
<td>Circulation System &amp; Walkability</td>
<td>Median Length of Block</td>
<td>-</td>
<td>0.059%</td>
<td>$105</td>
<td>P</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Num. of Accessible Entrance</td>
<td>-</td>
<td>1.50%</td>
<td>$2,700</td>
<td>U</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Greenery</td>
<td>Greenway Connectivity</td>
<td>+</td>
<td>5.22%</td>
<td>$9,300</td>
<td>D</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Housing</td>
<td>Structural</td>
<td>Number of Bathrooms</td>
<td>+</td>
<td>0.054%</td>
<td>$96</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Main Area</td>
<td>+</td>
<td>0.74%</td>
<td>$1,300</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Attached Garage</td>
<td>+</td>
<td>0.01%</td>
<td>$18</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>All Porches</td>
<td>+</td>
<td>0.009%</td>
<td>$16</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Building Age</td>
<td>-</td>
<td>0.66%</td>
<td>$1,170</td>
<td>U</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Swimming Pool</td>
<td>+</td>
<td>6.35%</td>
<td>$11,600</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Location</td>
<td>Attach to Golf Course</td>
<td>+</td>
<td>16.25%</td>
<td>$28,900</td>
<td>D</td>
<td></td>
</tr>
</tbody>
</table>

Assume: 1) The value of a house in a subdivision is $177,800
2) The subdivision has an average connection rate of 0.6 to greenway trails
3) The subdivision has four accessible entrances
4) The subdivision has the mean value of median length of cul-de-sac of 0.014 miles
5) The subdivision has the mean value of median length of block of 0.5 miles
6) The subdivision has a mean value of total main area group-mean of 2,000 square feet
7) The subdivision has a mean value of a sense of arrival score of 3

Note: P: % change of housing value for a 1% change of the variable
U: % change of housing value for one unit change of the variable
D: % change of housing value when the variable exists (dummy variable)

Priority: Priority order of value creation concepts for developers based on % and $ changes
5.2 Discussion

The main important application of the research findings is to use the Hierarchical Linear Model (HLM) to measure the relationship between the housing values and variables in both the housing level and the subdivision level. Even though the traditional hedonic price model is extensively used in most research related to housing values due to easy application and interpretation, the results can be slightly biased when the data in the model has a hierarchical structure. The HLM includes interaction terms between housing level variables and subdivision level variables, and among subdivision level variables themselves. This research would be a contribution as being a test of a more robust method for dealing with nested data. Another practical application of this study is to give the opportunity for researchers and real estate developers to examine the role of value creation concepts in terms of the amount of potential change in single family housing values. First, developers are guided to protect or enhance the property values by actively adapting positive value concepts in subdivision development. From the result of the research, it could be concluded that dwellers may pay more money for a house within a subdivision with greenway connectivity, a favorite shape of the subdivisions’ entrance, a fewer number of subdivision entrances, smaller blocks, or longer cul-de-sacs. In addition, it is expected that appraisals predict more accurately the property values and the sale prices by applying weights depending on the number and type of value concepts. To evaluate the more reasonable value of each house, the appraisals have to consider not only structural and locational variables of homes, but also sense of arrival, circulation system & walkability, and greenery characteristics in a subdivision which nest the homes.
This research has several limitations. First, GIS data and U.S. Census data are old-
dated. The GIS data are not frequently updated so that the GIS data used in this study
may reflect past situations, and the only U.S. Census data that could be used are for the
year 2000. However, because all GIS data and appraisal value data are from about same
year, 2008, some portion of this limitation would be reduced. Second, appraisal value is
used as a proxy for market sales data. Finally, the results from this research are limited
to College Station, Texas, or to similar sized and characterized campus towns in Texas.

This study is a pilot study to test the legitimacy and robustness of an alternative
method and to verify the effects of subdivision variables on housing value. Future
research would apply the models to other places which have different climates or crime
rates, etc., to generalize the external validity. Until now, there is no research about the
effects of value creation concepts on housing values. Moreover, the study area of this
paper is one place, College Station, Texas, which has specific characteristics such as a
hot climate and is a university town. Second, subjective evaluation of value creation
concepts could be reexamined. This paper tries to develop an objective evaluation of
value creation concepts. If opinions of subdivision developers can be gathered through
interviews or surveys, it may be possible to do a subjective evaluation of value creation
concepts. Each developer may have preferred value creation concepts which he or she
had already applied to prior subdivisions. Third, coefficients of some value creation
concepts in HLM with linear relations instead of constants could be examined in detail.
For example, the coefficient of the median length of cul-de-sac variable shows with an
equation of “0.044 - 0.001 * the number of accessible entrance;” that is, the median
length of cul-de-sac variable of a subdivision would be decided by the number of entrances of the subdivision. In this study, to compare the results in both models with ease, the coefficient was calculated by applying the mean value of the number of entrances in College Station, Texas. Fourth, the effects of components of the sense of arrival variable on housing value could be examined. The sense of arrival variable was evaluated with pictures of subdivision entrances in this study. Each scene has different components such as trees, signage, a gate, sidewalk, and so on, and each component may have different effects on housing value. Fifth, for further study, it would be better to survey homeowners’ opinions rather than students on the sense of arrival since they have already acquired a home and provide a more accurate reflection of sense of arrival in their decision to purchase a home in a subdivision. Finally, this study could be developed with real sales data instead of appraisal values. Comparison of the results from a model with real sale data with the results of this study will be helpful to enhance validity and reliability of this study.
REFERENCES


The authors sincerely thank Dr. Atef Sharkawy, Dr. Woo-Hwa Shin, and the two anonymous reviewers for their valuable comments on earlier drafts of this article.
1 Public Improvement District (PID) or Tax Increment Financing (TIF) can be an example.


3 There is the possibility that some houses in developing subdivisions in 2008 were excluded in the data, and it may cause selection bias in this research.

4 The required sample size for housing level can be calculated using three other variables related to statistical inference: significance criterion \((\alpha)\), population effect size (ES), and statistical power (Cohen, 1988). In general, \(\alpha\) is equal to .05. The statistical power is \(1 – \beta\). \(\beta\) is a Type II error, the probability to accept \(H_0\) (null hypothesis) when \(H_0\) is false. A statistical power smaller value than .90 can cause too big a risk of a Type II error. Finally, ES is the probability that \(H_0\) is false by the discrepancy between \(H_0\) and \(H_1\). Cohen (1992) defined medium ES as “an effect likely to be visible to the naked eye of a careful observer”. For regression analysis, he found that medium ES is .15. If it is assumed that 60 independent variables (IV) are statistically significant, the error degree of freedom (V) and noncentrality parameter \((\lambda)\) can be calculated. The required sample size of 245 is calculated with both the noncentrality parameter \((\lambda)\) and population effect size (ES). Hence, the number of homes for this study should be at least 245. Next, to identify the required sample size for the subdivision level, the Optimal Design software can be used. This software is a freeware program, developed by several researchers to provide general power computations. The required sample size can be calculated using four other variables related to statistical inference: significance criterion \((\alpha)\), standardized effect size \((\delta)\), intra-correlation \((\rho)\), and statistical power. The value of \(\alpha\) is equal to .05, and the statistical power is .90 (Cohen, 1988). The intra-class correlation \((\rho)\) is captured by a ratio of the variability between clusters to the total variability. For U.S. data sets of school achievement, \(\rho\) typically ranges between 0.05 and 0.15. Because there is no
information of ρ value on subdivisions, the minimum value of ρ (.05) is used. Finally, a standardized

effect size between .50 and .80 is a “large” probability that H₀ is false by the discrepancy between H₀ and

H₁; however, the standardized effect sizes between 0.20 and 0.30 are often considered worth detecting.

Among “small” effect sizes, .30, which is close to a medium effect size, is used. As a result, at least 8

single family homes per subdivision are required to get the power of over .90 when there are 85

subdivisions.

5 Texas House Bill No. 2188 was passed by the Texas State House of Representatives in April, 2007, and

formally enacted in June, 2007. The law was heavily supported by the real estate industry and lobbyists.

Based on H.B. No. 2188, it amended Sec. 552.148 in the Government Code, Chapter 552, Public

Information, Subchapter C: “Information Excepted from Required Disclosure” and changed the

availability of real estate data. According to Sec. 552.148, public information is available to the public,

with several exceptions. Sec. 552.148 mentions an exception regarding records of the comptroller or

appraisal district received from a private entity. Sec. 552.148 states that “Information relating to real

property sales prices, descriptions, characteristics, and other related information received from a private

entity by the comptroller or the chief appraiser of an appraisal district remains confidential in the

possession of the property owner or agent; and may not be disclosed to a person who is not authorized to

receive or inspect the information”.

6 To compare sales data and appraisal data, 37 samples were collected. Among them, seven sales values

were gathered from Zillow.com which is an online real estate website (www.zillow.com/). Thirty other

sales data samples were provided by anonymous researchers. The sales data were only used in

comparison with appraisal values. The 37 single family homes were sold in 2003, 2007, and 2008. Each

sale value was compared with the appraisal value of the sold year. The average sale price of the 37

samples was about $160,000, and the average appraisal value was about $152,000. From the statistics of

samples, it can be said that appraisal value is about 95% of the sales value of a single family homes in the

city of College Station. Pearson’s correlation coefficient measures how two variables are related. The

correlation coefficient can be calculated accurately when the data do not have any outliers and have a
linear relationship. Pearson's correlation coefficient (.989) is significant at the 0.01 level. From the correlation, it is clear that sale prices are almost perfectly correlated with appraisal values.

7 There are two ways to assess the intraclass reliability: Cohen's Kappa, and the Intraclass Correlation coefficient (ICC). Cohen's Kappa for intraclass reliability could be used when there are only two raters; hence, an intraclass reliability analysis using the ICC statistic was performed to determine consistency among the sixty-one raters. The intraclass reliability is defined as “the extent to which two or more individuals evaluating the same product or performance give identical judgments” (Leedy and Ormrod, 2001).

8 Because there is no exact rule for the ICC, it would be best way to use the criteria of Cohen’s Kappa statistic. As a rule of thumb, Cohen’s Kappa statistic of 0.40 to 0.59 is moderate intraclass reliability, 0.60 to 0.79 substantial, and 0.80 outstanding (Landis and Koch, 1977).

9 Skewness represents how much data distribution was skewed, and kurtosis shows how peak or flat the graph of the data distribution is. A zero value of both the skewness and kurtosis indicate that the distribution was perfectly normal.

10 If the value for skewness or kurtosis of a variable was greater than +3 or less than -3, the variable was not normally distributed (Princeton University, Online-help, regression_intro.htm, 2009).

11 To do likelihood ratio test, as the method of estimation in HLM, Full Information Maximum Likelihood (FIML) was used instead of Restricted Maximum Likelihood (REML) because of the number of estimated parameter.

12 For instance, in the case of the total main area variable, a 1% increase of the total main area of a single family house led to an average appraisal value increase by 0.9%.

13 For example, in the case of the Attached Garage variable, a square foot rise in the attached garage resulted in a 0.01% rise of the appraisal value.

14 In the case of the swimming pool variable, the true portion change would be 7.10% (100 * (exp(0.0689) -1) = 7.13291). Hence, the expected appraisal value for a house with a swimming pool was 0.071 (7.13%) higher than the appraisal value for a house without a swimming pool.