

CHAPTER 3: RESERACH METHODOLOGY

3.1 Introduction

Researches in hedonic price indexes for housing have used long periods and wide geographical regions to cover their temporal and geospatial studies. Gourieroux and Laferrere.(2009) used a 4-years period and 296 strata, Hill and Melser (2008) used a 6-years period and 14 regions, and Dorsey et al. (2010) used 9-years period and 471 zip codes. Of course, it will good to include a wide geographical coverage but bearing in mind the above mentioned researchers have received assistance or sponsorship in their projects

The multiple regression model is used to determine the relationship of the dependent variable with its independent variables. The regression coefficient is the constant in a linear regression equation where $y = a + bx$ that represents the rate of change in the dependent variable (y) as a function of changes in the independent variable or predictor (x); it is the slope of the regression line. We must keep in mind that each coefficient is influenced by the other variables in the model. As predictor variables are nearly always associated, two or more variables may give rise to the same variation in Y. Therefore, each coefficient does not explain the total effect on Y of its corresponding variable, as it would if it were the only variable in the model. Rather, each coefficient represents the additional effect of adding that variable to the model, if the effects of all other variables in the model are held constant or already accounted for. Therefore, each coefficient will change when other variables are added to or deleted from

the model. The mean and standard deviations will be looked at since the data range covers a wide span with multiple locations. The skewness is either a positive or negative value which characterizes the symmetrical measure of the distribution. A large kurtosis value showed the heavy tailed distributions for these values.

3.2 Sampling

This study uses stratified sampling to select from the JPPH quarterly Residential Property Stock Report and the BNM Statistics. This method is chosen as to ensure subgroups within the population are represented in the sample where previous published researches have validated the importance of the factors affecting the hedonic multiple regression. The period covered spanned from Q1 2004 to Q4 2010 forming a time series of twenty eight dataset for each location or independent variable.

3.2.1 Unit of analysis

Currency in Ringgit Malaysia and supply of residential units are measured in continuous number. Dummies for types are in binary where 1 equal to present if transaction occurred in period t or else 0. Log refers to the value at time t over the value of $t - 1$. Natural log is the logarithm to the base e , where e is a constant to approximately 2.718281828. Natural log is used to transform big numbers, like transaction value in hundreds of thousands Ringgit, into more manageable continuous numbers.

3.2.2 Sampling procedure

In the JPPH report they have classified the regions according to Malaysia's official geographical boundary of 13 states and 3 federal territories. However, due to incomplete and inconsistent data over the period, one state (Sarawak) and two federal territories (Putrajaya and Labuan) were dropped from this analysis. In each state or federal territory, it is further sub-divided into districts, however, the number of districts in each state will depend on the size.

Table 3.1 below refers.

Table 3.1
Number of districts in each state.

No.	State/Federal Territory	No. of Districts
1	Perlis*	1
2	Kedah	13
3	Penang	6
4	Perak	10
5	Selangor	10
6	Kuala Lumpur	9
7	Negeri Sembilan	8
8	Malacca	4
9	Johore	9
10	Pahang	12

11	Kelantan	11
12	Kuala Trengganu	8
13	Sabah	21
Total		122

Perlis - Supply is listed for 19 districts but sales price is listed for one Perlis alone.

In the Malaysian context, the type of house is closely related to the number of rooms and bathrooms and hence by selecting the types of house which is wider in scope than just the rooms this research should fully cover the effects of number of rooms in the regression. The JPPH report provided comprehensive analysis of the types of houses available in Malaysia ranging from 1-1.5 storey terrace, 2-3 storey terrace, 1-1.5 storey semi-detached, 2-3 storey semi-detached, detached, cluster, townhouse, low cost house, low cost flat, flat, service apartment, condominium/apartment and SOHO. Cluster home and townhouse are limited to certain district and state only while service apartment and SOHO concepts are quite new in Malaysia. Due to limited transactions data these four categories are dropped. Nevertheless, the research paper still covers nine types for analysis.

3.3 Definition of the variables.

3.3.1 Dependent variable: micro

The dependent variable is the response that is measured whenever there are changes to any of the independent variables. In this regression model it is the mean price.

3.3.2 Independent variables: micro

The independent variable is the variable that changes on its own through the volume of transactions over a time period. The micro independent variables selected for this study are the supply, change in price from one quarter to another and the types of residential units transacted. These variables are expected to have strong relationship to the dependent variable of price with a significant level of observation that 95% of the time the sample is true to its properties and can be reliable for predictions.

3.3.3 Dependent variable: macro

The dependent variable for the macro factors is the Housing Price Index (HPI) which measures the impact of economic changes over the time period.

3.3.4 Independent variables: macro

The macro independent variables selected for this study are the GDP, KLCI, CPI, BLR and FDI.

These variables are expected to have a relationship to the dependent variable of HPI with a significant level of observation at 95% support.

In all economic measurement the GDP is the foremost component to be selected, (Glindro et al., 2008; Pashardes and Savva, 2009).

As capital market and property market are different in liquidity, transaction volume and expected return, investors are keen to know the performance of each of these markets and how they relate to each other. Keogh (1994) monitored and suggest that investors may switch investment between these two forms as economic circumstances change from cycle to cycle. As such, the KLCI is chosen to be included as a macro variable, (Mulok and Kogid, 2008).

CPI and HPI are expected to move in the same direction, although the changes may vary in intensity. This is the effect of inflationary pressure in which mindset may fuel each other to hedge and cause inflation concurrently (Pashardes and Savva, 2009).

BLR has a direct relationship to property transactions. Since it is a major purchase either for home ownership or as investment, buyers tend to borrow from financial institutions and hence have to incur interest cost, (Zhou and Haurin, 2010).

As Malaysia is a developing nation it needs and draws foreign direct investments to grow. The net inflow of investments is expected to generate and raise the income level of its citizens who then have the capability to invest in housing. Hence, it is expected to have a relationship to the HPI.

3.4 Model

Multiple regression analysis is a technique to explore the relationship between one continuous dependent variable and a number of independent variables or predictors (usually continuous), (Tabachnick and Fidell, 2007). It is based on correlation but explore deeper into the interrelationship among a set of variables. In this study, the first group of model uses the micro factors for state level while the second group uses the macro factors for national level.

3.4.1 Micro factors

Let the hedonic housing price models for each state be:

$$\text{LN}(\text{Pr}_{it}) = \alpha_0 + \beta_1 \text{LNS}_{it} + \beta_2 \text{LN}\Delta\text{Pr}_{it} + \sum \beta_3 D_{type} + \varepsilon_i \dots\dots\dots(1)$$

where $\text{LN}(\text{Pr}_{it})$ denotes the natural logarithm of transaction price of residential unit i at time t ; LNS is the natural log of supply of residential unit i at time t in each location; $\text{LN}\Delta\text{Pr}_{it}$ is the natural log of price change of residential unit i at time t in each quarter; D_{type} is a dummy variable for types of residential units that is one if residential unit i transacted at time n (i.e., $n=t$), otherwise zero; and ε_i the error term of the residential unit. The α is the intercept of the regression and the β is the estimate of regression coefficient of each respective independent variable which we are trying to determine. The dummies are further clarified as follows;

DummST = Single Storey Terrace=1, otherwise=0

DummDT = Double Storey Terrace=1, otherwise=0

DummSSD = Single Storey Semi-Detached=1, otherwise=0

DummDSD = Double Storey Semi-Detached=1, otherwise=0

DummDet = Detached=1, otherwise=0

DummLCH = Low Cost House=1, otherwise=0

DummLCF = Low Cost Flat=1, otherwise=0

DummFlat=Flat=1, otherwise=0

3.4.2 Macro factors

The macro model is for the national level where it may be stated as;

$$HPI_t = \alpha_2 + \beta_1 LNGDP_t + \beta_2 LNKLCI_t + \beta_3 LNCPI_t + \beta_4 BLR_t + \beta_5 LNFDI_t + \varepsilon_{i,t} \dots (2)$$

Where the HPI at time t, LNGDP is the natural log of GDP, LNKLCI is the log of KLCI, LNCPI is the natural log of CPI, BLR is the Base Lending Rates from commercial banks, LNFDI is log net FDI.

3.5 Hypotheses

This study is divided into two panels of micro variables and macro variables. The sample period chosen spanned from Q1 2004 to Q4 2010. The

micro variables will look into the mean prices of housing transactions and the physical characteristics bundle of the house in accordance to the JPPH classification, i.e. single storey terraced, 2-3 storey terraced, single storey semi-detached, 2-3 storey semi-detached detached house, low-cost house, low-cost flat, flat and condominium/apartment.

The macro variables will look into some major economic statistics that have an impact on the Housing Price Index (HPI). The housing loan interest rate is an important element to influence an investor to decide on selecting a house. This element is deemed intervening to the housing price model.

The theoretical framework for this project may be expressed in Figure 3.1.

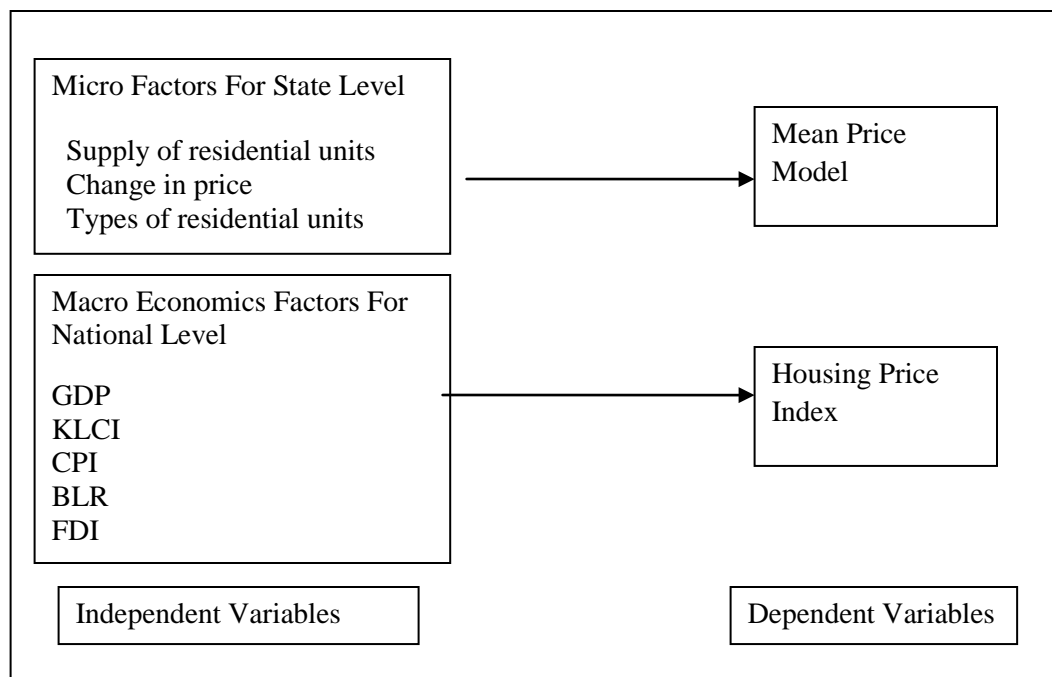


Figure 3.1
Theoretical framework of variables affecting hedonic housing price

3.5.1 Hypotheses with micro variables

The hypotheses for regression are;

H_1 : There is positive and significant relationship between housing price and supply of houses.

Supply of residential units is expected to have a positive relationship to the price in each state. This is due to the fact that as stock of supply increases in the property market it will be well received since the population is increasing and the workforce consists of young couples. Hui and Gu (2009), quoted Case and Shiller (2003); and Hui and Yue (2006), as having selected supply-side and demand-side factors as the fundamental price factors.

H_2 : There is positive and significant relationship between housing price and change in price from one quarter to another.

Change in price is expected to have a positive relationship since the higher value will signify a growing property market. However, should the economy turn into recession the change in price will reflect a negative relationship, (Clayton, Miller and Peng 2010).

H_3 : There is positive and significant relationship between housing price and single storey terrace houses transacted.

H_4 : There is positive and significant relationship between housing price and double storey terrace houses transacted.

H₅ : There is positive and significant relationship between housing price and single storey semi-detached houses transacted.

H₆ : There is positive and significant relationship between housing price and double storey semi-detached houses transacted.

H₇ : There is positive and significant relationship between housing price and detached houses transacted.

In general, the landed types of residential units transacted will lead to a positive relationship to the housing price as these are deemed to be superior. Bostic, Longhofer and Redfearn (2007) noted that houses with “land leverage” will have relatively more volatile prices. Terrace houses provide some land in front and back while semi-detached have land space on the side of the building. Of course, the detached house is the most exclusive as it provided privacy with land surrounding it. The value will also fluctuate depending on size and location.

H₈ : There is negative and significant relationship between housing price and low cost houses transacted.

H₉ : There is negative and significant relationship between housing price and low cost flats transacted.

H₁₀ : There is negative and significant relationship between housing price and flats transacted.

Low cost houses, low cost flats and flats are developed as part of the government's social economic policy to ensure house ownership. There is no land to it, generally developed near main roads for convenience to public transport system but attracting noise and dust, and usually of the leasehold type. Adair A. et al. (2000,pp. 699-716), reported "... accessibility is of little significance in explaining variation in house prices at a city-wide scale but at a sub-market level, particularly in lower-income areas, accessibility can be an important influence." Maintenance of buildings and upkeep in surrounding areas are usually below standard. These types of housing are viewed as inferior and thus having a negative relationship towards the housing price.

3.5.2 Hypotheses with macro variables

The hypotheses are;

H₁₁ : There is positive and significant relationship between HPI and GDP.

The GDP is expected to have a positive relationship to the HPI since the national growth will expand the citizens' income and affordability to acquire a home, (Glindro et al., 2008).

H₁₂ : There is negative and significant relationship between HPI and KLCI.

The KLCI is expected to be in inverse relationship with the HPI since a growing stock market draws funds away from the housing market, (Pashardes and Savva 2009).

H₁₃ : There is positive and significant relationship between HPI and CPI.

As consumer price index moves up people are willing to pay more and invest in housing as a hedge against inflation, (Pashardes and Savva 2009).

H₁₄ : There is negative and significant relationship between HPI and BLR.

The BLR is expected to have a negative relationship since as interest rates rise buyers' financing increases leading to a slow down in the property sector resulting in a lower HPI, (Glindro et al., 2008).

H₁₅ : There is positive and significant relationship between HPI and FDI.

The inflow of funds from FDI creates opportunity and employment leading to wealth for the people to invest in housing.

H₁₆ : There is positive and significant relationship between HPI and its one quarter lag.

The current HPI is expected to be influenced by the one term lag of the previous HPI due to human behaviour's expectation of the trend to continue.

3.6 Data collection and preparation

3.6.1 Sources : micro

The micro data are downloaded from the JPPH or the Valuation and Property Department website. This study has used all the available public information from Q1 2004 to Q4 2010 for the micro factors. This covers a time series of four continuous quarters for seven years period.

All micro data are collections from secondary source from each quarter's Residential Property Stock Report. The report is encrypted in Adobe PDF format. According to its technical notes, supply of residential stock excludes (i) institutional quarters, (ii) abandoned projects, and (iii) squatter units. The mean is the simple average of sales prices and median is the sales price for which one-half of the observations will lie above that price and one-half will lie below it.

The data is extracted from the Adobe file and transferred into an Excel file. At this stage problems are encountered, (please refer to the Problems Encountered Section). The data is segregated into each state.

Data preparation for analysis involved the matching of supply of residential units by types and sales prices of residential properties in each quarter to form a continuous time series for the period Q1 2004 to Q4 2010. The types of residential properties are coded as dummy equals 1 if a transaction occurred during the quarter or else 0. The combined data for the four quarters and seven years period are sorted into the order of district, year and quarter to facilitate the computation of price change in mean or median. Next, the data is sorted by the

types. Upon completion, physical inspection of the table is done for consistency. The supply, mean and median prices were then stacked in each column, respectively. Within each type and each district, the change in mean and median prices were computed in log manner whereby,

$$\text{LN}\Delta\text{Pr}_{it} = \frac{\text{Pr}_t}{\text{Pr}_{t-1}}$$

Where Pr_t = price at time t.

3.6.2 Sources: macro

All relevant macro data are downloaded from Bank Negara website <http://www.bnm.gov.my> which is in Excel format and maintained in time series of either daily, monthly, quarterly or annually. Fortunately, BNM had maintained a consistent approach in its tables. Also, since these tables are in Microsoft Excel spreadsheet, it allowed for easy extraction of relevant data for this project. The extracted tables from BNM are;

BNM Table Reference	Description
2.1	Interest Rates : Banking Institutions
2,12	Bursa Malaysia Securities Berhad : Selected Indicators
3.3.1	Gross National Income (GNI) by Expenditure Components in Constant (2000=100) and Current Prices

3.5.8	Consumer Price Index
3.5.11	House Price Indicators
3.6.12	Foreign Direct Investment in Malaysia by Sector

Figure 3.2
Bank Negara Malaysia Reference Table

3.6.3 Problems Encountered

As the project got underway, difficulties and problems were encountered with data collection from the JPPH reports. These quarterly reports were all protected by Adobe encryption. To key in the data for 13 states, 122 districts, 4 quarters, 7 years, 10 types, 1 supply and 2 prices (mean and median) will result in 532,896 entries. With an average of 5 digits for price value, it means a keystroke of 2,664,480. The JPPH was approached for assistance in the hope of obtaining the data in Microsoft Excel format. Unfortunately, they were not accommodating due to their work volume and had specified tough confirmation requirements despite providing them with the UM reference letter confirming that data collection is for academic research purpose and not for commercial use.

Fortunately, my supervisor advised that the Adobe file could be converted using a software. The Adobe software was installed and it allowed file conversion to Microsoft Words. Unfortunately, the conversion process was not smooth due to either file corruption or structural change in different version of the software over the long periods. Some files could be opened while some cannot. Some

files when opened contained jumbled data due to formatting. Whatever data that could be converted into Words format were then saved and copied into Excel file. Any missing data were then manually keyed in. The whole process was taxing and time consuming.

WP Kuala Lumpur	22,210	59,888	633	WP Putrajaya	0	608	0	WP Labuan	588
730	126	Selangor	129,561	288,963	6,575	Johor	145,560	103,024	23,958
Pulau	Pinang	37,232	37,099	7,684	Negeri Sembilan	69,244	27,467	8,516	Perak
102,194	62,097	14,055	Melaka	39,159	22,004	3,822	Kedah	51,066	17,196
26,330	Pahang	32,700	17,935	10,152	Terengganu	8,701	2,993	3,654	Kelantan
13,934	4,157	2,227	Perlis	5,608	894	2,156	Sabah	15,515	27,064
2,518	Sarawak	24,441	35,992	8,398					

Figure 3.3
Text and data within one column when downloaded

The table on supply of residential units by types and sales prices of residential properties are not in the same structure. In addition, while the listing of districts looks similar it did not match perfectly. The supply contains the sum total of all units from the districts in the states and appeared as a last line item while the summary prices for the state appeared as the first line item. There could also be supply listed for a district but that did not appear in the sales prices, perhaps due to none transactions. Usually to maintain table consistency the same district was maintained but listed as N[. Without this consistency, merging of these two tables became a nightmare when data analysis found weird results.

As the JPPH reports must have been compiled from many different locations, it was subsequently found that the handling of keying in varied over the periods and locations. In subsequent files merging for all the periods and sorting it was discovered that spacing was available in some text. For unknown reason, one district in a state could contain sales prices value keyed in as thousand of ringgit whereas all others in unit of ringgit.

Table 3.2
Inconsistent data value

Year	Quarter	District	Single Storey Terraced	2 – 3 Storey Terraced	Single Storey Semi-Detached	2 – 3 Storey Semi-Detached
2007	Q2	Kota Kinabalu	171,500	266,512	275.6	416,313
2007	Q2	Sabah	151,939	228,135	234.75	357,389
2007	Q2	Sandakan	114,000	140,688	130	233,000
2007	Q3	Kota Kinabalu	175,693	282,007	234,333	412,869
2007	Q3	Sabah	153,435	226,291	182,000	367,576

In Table 3.2 the column for single storey semi-detached in Q2 were keyed in thousand as compared to unit in other types and Q3 period.

3.6.4 Conclusion

Ten locations (states) were finally selected for micro factors analysis of which seven have valid cases above 1,000 while the smaller states of Negeri Sembilan, Malacca and Trengganu have less since their numbers of districts are smaller. Upon completion, each case is assigned an i.d. and then each worksheet file is exported to the SPSS programme for data analysis using the SPSS V16. Descriptive tests are conducted on the data collected to ensure

consistency and to identify abnormality. Correlation among the variables is tested for linearity and non-linearity in order to obtain a good fit for the multiple regression model. Finally, linear regression analysis is conducted with collinearity diagnostics report.

The macro data are compiled into respective column and assigned case i.d. for each quarter. It is then exported into SPSS programme for analysis.

The Mean Price, Median Price, Supply and GDP data were transformed into the natural logarithm for easier management. The Change in Mean Price, Change in Median Price, KLCI, CPI and FDI were computed using the logarithm method of value at time t divided by the value at time $t - 1$.