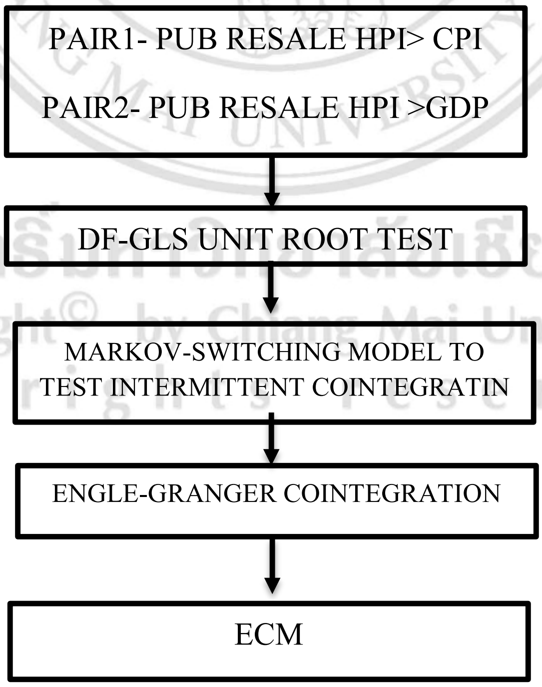
CHAPTER 3

Methodology

3.1 Research Design

As we provided in the literature review, many authors studied in related housing price field with different areas. We decided to do something fit in with model and methodology as well. We will do research on public resale price index and private price index separately. We would like to examine the cointegration between public housing price, private housing price, GDP and CPI of Singapore from 1990 to 2015. We would use Engle-Granger two steps approach and ECM to estimate the cointegration between variables. Additionally, we would use Markov switching model to test if there is intermittent cointegration between variables.

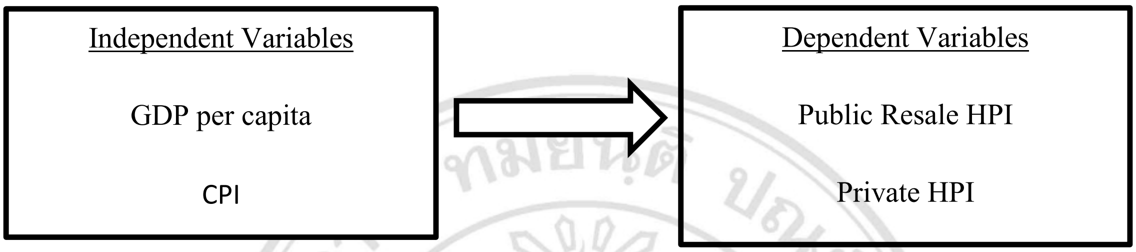


Source: Created by author

Figure 3.1: Procedure of Research

3.2 Conceptual Framework

There are many other variables that affect housing price. We focused on the impact of GDP per capita and CPI on public housing price index and private housing price index of Singapore from 1990 to 2015.



Source: Created by author

Figure 3.2: Conceptual Framework

3.3 Variables Used in the Model and Sources of Data

In this paper, we used dependent data as housing price index, which separated into public and private. Independent variables are GDP per capita and CPI. The data of these variables are collected from World Bank and Singapore Statistics. The period of data set is from 1990 to 2015. The variables chosen are complied with theories and previous research. Table below descried detail of the variables.

Brooks and Tsolacos (1999) show that the term structure and unexpected inflation explain changes in the UK property market. Using a life-cycle model of the property market, Ortalo-Magne and Rady (2006) show that house price dynamics are driven by disposable income.

Table 3.1: Definition and Data sources of the variables

VARIABLES	DEFINITIONS	DATA SOURCES
Public Housing Price Index	It is the measurement of the price changed of residential housing	Singapore Statistics
Public Housing Price Index	It is the measurement of the price changed of residential housing	Singapore Statistics
GDP per capita	It is a measurement of a country's economic output that accounts for population.	World Bank
CPI	It is a measurement that examines the weighted average of prices of a basket of consumer goods and services	World Bank

Source: Created by author

Table 3.2: Expected sign of the variables

Macro Variables	Expected Signs
GDP per capita	+
CPI	+

Source: Created by author

According to the previous literature reviews and theories, the variables have expected sings are in figure below. According to the study of Tien-Foo Sing, I-Chun Tsai, Ming-Chi Chen (2006), Shanmuga Pillaiyan(2015), Lei Feng, Wei Lu, Weiyan Hu, Kun Liu(2010) and Ogonna Nneji, Chris Brooks and Charles Ward (2003), GDP had a positive relationship with housing prices. CPI had a positive correlation with housing price due to the study of Tien-Foo Sing, I-Chun Tsai, Ming-Chi Chen (2006), Nicholas

Apergis(2003), Ogonna Nneji, Chris Brooks and Charles Ward (2003), Nicholas Apergis and Anthony Rezitis(2000) and Baffoe-Bonnie(1998). That is why GDP and CPI are expected to be positive.

3.4 Model Specification

This study falls into two parts, public resale housing price index and private residential price index. The four equations are as follow:

$$PUB\text{Resale HPI}_{it} = \alpha_0 + \alpha_1 lnGDP_{it} + \mu_{it} \quad (21)$$

$$PUB\text{Resale HPI}_{it} = \alpha_0 + \alpha_1 CPI_{it} + \mu_{it} \quad (22)$$

$$PRI\_HPI_{it} = \alpha_0 + \alpha_1 lnGDP_{it} + \mu_{it} \quad (23)$$

$$PRI\_HPI_{it} = \alpha_0 + \alpha_1 CPI_{it} + \mu_{it} \quad (24)$$

where,

$PUB\text{Resale HPI}_{it}$= Public Resale housing price index

PRI_HPI_{it}= Private housing price index

$lnGDP_{it}$= Natural log of GDP per capita

CPI= Consumer Price Index

We used Engle-granger two steps approach and ECM as econometric techniques to test above equations. We added Markov-Switching Engle-Granger cointegration model to test if there is an intermittent cointegration or not between variables.

3.5 Research Methodologies

This paper investigates the long-run and short-run linkages of four pairs of variables, which are (1) public resale HPI and CPI, (2) public resale HPI and GDP per capita, (3) private HPI and CPI and (4) private HPI and GDP per capita. To examine the relationships between housing price indices, CPI and GDP per capita, this study applied the Engle-Granger 2-step Error Correction Model (ECM) to study the short-run and long-run relationships between each pair of random variables. Since the data present the possibility of different trends and relationships across time, this study also adopts the Markov-Switching Engle-Granger (MS-EG) cointegration test to allow the cointegration to be intermittent. Therefore, for the methodology part, section (3.1) first discusses the standard ECM model and section (3.2) then discusses the MS-EG cointegration test

3.5.1 DF-GLS Unit root tests (ERS test)

DF-GLS test makes stronger Augmented Dickey-Fuller test when the sample size is small. Dickey-Fuller test statistics was modified into better version by El- liott, Rothenberg and Stock (ERS) by using a generalized least square, which was published in 1996. That modified test has the conclusively ordinary Dickey-Fuller test and the best performance in sample size.In particular, Elliott et al. find that their “DF-GLS” test “has substantially improved power when an unknown mean or trend is present.” (1996, p.813).

Just as the standard Dickey-Fuller test may be run with or without a trend term, there are two forms of DF-GLS: GLS detruding and GLS demeaning. With GLS detrending, the series to be tested is regressed on a constant and linear trend, and the residual series is used in a standard Dickey-Fuller regression. With GLS demeaning, only a constant appears in the first stage regression; the residual series is then used as the regress and in a Dickey- Fuller regression.

The estimation based on the generalised least square (GLS) using the transformation:

$$y_1 = y_1, \widetilde{y_t} = y_t - \rho y_{t-1}, \text{ t=2,..,T} \tag{25}$$

$$x_1 = 1, x_t = 1 - \rho, \text{ t=2.....,T} \tag{26}$$

where $\rho = 1 + \frac{\bar{c}}{T}$ and $\bar{c} = -7$, based on the equation

$$\widetilde{y_t} = \beta_0 x_t + \varepsilon_t \tag{27}$$

Parameter $\widetilde{\beta_0}$ is estimated by the least squared method and used to remove constant from *the* time series *y*

$$y_t^* = y_t - \widehat{\beta_0} \tag{28}$$

The ADF test is calculated based on the time series by

$$\Delta y_t^* = \emptyset_1 y_{t-1}^* + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i}^* + \varepsilon_t \tag{29}$$

GLS estimates the linear trend in the model**.** Transformation is extended by $z_1 = 1, z_t = t - \rho(t-1), where\ \rho = 1 + \frac{\bar{c}}{T}. where\ \bar{c} = 13.5$. Estimates of parameters are calculated as follow

$$y_t = \beta_0 x_t + \beta_1 z_t + \varepsilon_t \tag{30}$$

Estimated parameters and removed trend from the time series

$$y_t^* = y_t - (\widehat{\beta_O} + \widehat{\beta_1} t) \tag{31}$$

Finally, the ADF test is applied to the transformed time series, i.e. the test statistic is obtained from the following equation:

$$\Delta y_t^* = \beta_0 + \emptyset_1 y_{t-1}^* + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i}^* + \varepsilon_t \tag{32}$$

The critical values for the ADF-GLS test obtained by simulation in Elliot, Rothenberg and Stock (1996) show that for models without constant they are the same as in the case of the ADF test. For the remaining models, critical values of the ADF-GLS test are used as indicated in Elliot, Rothenberg and Stock (1996) as well.

3.5.2 Markov-Switching Engle-Granger (MS-EG) cointegration test

For the long-run study, our data show different patterns in the trend over different time periods. Therefore, Hamilton (1989)'s Markov Switching Model is adopted to replace the Augmented Dickey-Fuller test in the Engle-Granger 2-step Error Correction Model (ECM) to allow intermittent cointegrating relationships. The Markov Switching Model is a nonlinear regression that allows different model structures across different regimes. That is, the MS-EG model allows each pair of variables to be cointegrated over some periods of time and non-cointegration in other periods. Let $s_t = \{1,2\}$ unobserved state variable, equation (2) with regime switching can be written as

$$\hat{u}_t = \begin{cases} \rho_{10} + \rho_{11}\hat{u}_{t-1} + v_{1t} & if\ s_t = 1 \\ \rho_{20} + \rho_{21}\hat{u}_{t-1} + v_{2t} & if\ s_t = 2 \end{cases} \tag{33}$$

where v_{1t} and v_{2t} are i.i.d. and s_t follows a first-order Markov process. Similar to standard unit-root tests, $\rho_{11} \geq 1$ indicates that $\hat{u}_t$ has unit-root and, thus, y_t and x_t are not cointegrated in regime 1. If $\rho_{11} < 1$, then $\hat{u}_t$ is stationary and thus, y_t and x_t are cointegrated in regime 1. The stationary of $\hat{u}_t$ and the cointegration of y_t and $x_t$in regime 2 is determined by ρ_{21}. As results, different regimes may have different cointegrating parameters and allows intermittent cointegration between two variables. The estimates for the long-run relationship presenting equation 1 are only statistically meaningful in regimes that y_t and x_t are cointegrated.

3.5.3 Engle-Granger 2-step Error Correction Model (ECM)

Yule (1936) and Granger and Newbold (1974) were the first to draw attention to the problem of spurious correlation and find solutions on how to address it in time series analysis. Given two completely unrelated but integrated (non-stationary) time series, the regression analysis of one on the other will tend to produce an apparently statistically significant relationship, and thus a researcher might falsely believe to have found evidence of a true relationship between these variables.

To examine the long-run and short-run linkages between housing price indices, CPI and economic growth, Engle and Granger (1987) recommend a two-step procedure for cointegration analysis. For the long-run equation, we have

$$y_t = \beta_0 + \beta_1 x_t + u_t, \tag{34}$$

where y_t is the dependent variable and x_t is the independent variable at year t. β_0 is the intercept, β_1 is the cointegrating coefficient capturing the long-run relationship between x and y and u_t is the residuals. More specifically, for model (1), y_t is public HPI and x_t is CPI. For model (2), y_t is public HPI and x_t is ln(GPD per capita). For model (3), y_t is private HPI and x_t is CPI. Finally, for model (4), y_t is private HPI and x_t is ln(GPD per capita).

For non-stationary variables to establish a long-run relationship, the variables must be cointegrated. To examine the cointegration, the Augmented Dickey-Fuller unit-root test is applied on the residual $\hat{u}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t$ from above long-run regression equation. Specifically,

$$\Delta\hat{u}_t = \phi\hat{u}_{t-1} + \sum_{i=1}^{p-1}\phi_i\Delta\hat{u}_{t-1} + v_t, \tag{35}$$

Under the null hypothesis there exists a unit root in the potentially cointegrating regression residual, while under the alternative hypothesis, the residuals are stationary. Thus, if we reject the null hypothesis, it indicates that a stationary linear combination of the non-stationary variables is found, i.e. the non-stationary variables are cointegrated.

The cointegration regression only considers the long-run linkages between the level series of variables, while the Error Correction Model (ECM) is developed to measure any dynamic adjustments between the first differences of the variables. A simple error correction term is the residual from equation (1), and the Error Correction Model (ECM) is then defined as

$$\Delta y_t = \alpha\hat{u}_{t-1} + \gamma\Delta x_t + \epsilon_t, \tag{36}$$

where ϵ_t is i.i.d. and the first difference of y_t can be explained the lagged of $\hat{u}_{t-1}$ and the first difference of x_t. The $\hat{u}_{t-1}$ is the one period lagged value of the residuals from estimation of equilibrium error term, or in another word, a disequilibrium error term occurred in the previous period. For cointegrated series, the error correction term $\hat{u}_{t-1}$, which represents the speed of adjustment toward the long-run values, offers an added explanatory variable to explain Δy_t.

As a wrap of this Engle-granger two steps cointegration

(1) Estimate the long-run equation

The OLS residuals resulted from equation (35). A test of cointegration is a test if the residual is stationary or not. This is determined by ADF tests on the residuals, with the MacKinnon (1991) critical values adjusted for the number of variables, which denotes as n.

The traditional diagnostic test is unimportant as the only important question is the stationary or otherwise of the residuals.

(2) Estimate the Error Correction Model

ECM describes how y and x behave in the short run consistent with a long run cointegrating relationship.

$$\Delta y_t = \phi_0 + \sum_{j=1} \phi_j \Delta y_{t-j} + \sum_{h=0} \theta_h \Delta x_{t-h} + \alpha \widehat{u_{t-1}} + \varepsilon_t \quad (37)$$

By using t-ratios and diagnostic testing of the error, the term is appropriate. The adjustment coefficient α must be negative.