Chapter 3

Methodology

3.1 Research Design

There is much literature about the determinant factors of the real estate price using empirical methods. This paper set up a real estate price cointegration equation, establishing a long-term equilibrium price of real estate of various factors. A real estate bubble is a reflection of an economic bubble, and the real asset bubble is always the most important thing when discussing real assets. However, in the real estate price bubble debate, there is not a definitely correct answer. Therefore, this paper not only introduces the definition of the real estate bubble, but also explains how the bubble forms. Also, this paper finds a method to test the bubble's character using panel data of major cities to set up a model and use the empirical results to prove the existence of an asset price bubble.

In addition, this paper uses panel data from two-dimensional cross-section and time series to study the development of the urban real estate situation and test the bubbles character. In order to avoid a time series generating spurious regression phenomenon, this paper uses panel unit root test and panel cointegration test methods to eliminate spurious regression. In this way, panel data can be estimated effectively. The regression process reveals some variables of the real estate industry in long-term equilibrium.

All in all, this search is based on theoretical and empirical analysis. As an empirical study, this research will focus on the combination of qualitative analysis and quantitative analysis. Analytical methods include econometrics panel data analysis, panel unit root test, panel cointegration test and other methods.

3.2 Overview of Data Sources

The data sources of this paper are the yearly data from twenty one main cities in China from 2000-2010. Table 3.1 gives definitions of continuous variables, which have been compiled by the National Bureau of Statistics of China. The data is published in the China Statistical Yearbook, China City Statistical Yearbook, China Urban Life & Price Yearbook, and China Real Estate Statistics Yearbook respectively.

In order to eliminate the impact of the price level, these raw data were deflated. All economic factors impacting on the empirical analysis of real estate price in a log linear reduced form. All nominal variables are converted into real values by using the consumer price index (2000year=100%). The per capita disposable income of urban residents, land transaction price index, and the urban real GDP are divided by the consumption price index. About data in detect bubble, the urban housing sales price index, and urban housing rents price, the paper uses the same method to get the actual value. Real interest rate is obtained by a five-year bank mortgage loan interest rate minus consumer price index rate.

Table3.1: Variable Definitions and Data Sources

Notation	Variable name	Definition	Measuring unit	Data source
HP	House Price	Sale price index	%	China real estate statistical yearbook
PPI	Income	Per capital disposable income	RMB	National Bureau of statistics of China
R	Interest Rate	Nominal mortgage interest rate with above 5 year maturity	%	The people's Bank China
GDP	Gross Domestic Product	Real gross domestic product	100 million RMB	China City Statistical Yearbook
LP	Land Price	Land transaction index	%	China real estate statistical yearbook

Note: All nominal variables are transformed into real variables by using the consumer price Index base on 2000 years (2000year=100%)

3.3 Econometric Methods

3.3.1 Panel Unit Root Tests

(1) LLC Test

Levin, Lin and Chu (2002) suggested the use of panel unit root test by carring out separate ADF regressions. The generated ADF regression see equation (1)

$$\Delta y_{it} = \delta_i y_{it-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mt} + \varepsilon_{it} \quad \text{m=1, 2, 3} \tag{1}$$

where:

Δy_{it} = difference term of y_{it}

y_{it-1} = panel data

$\delta_i = \rho$ -1

P_i = the number of lag order for difference terms

$\alpha_{mi} d_{mt}$ = exogenous variable in model such as city fixed effects and individual time trend

ε_{it} = the error term of equation (1)

Having determined autoregression order P_i in equation (1) Regress Δy_{it} and y_{it-1} against Δy_{it-L} (L=1,…, P_i) and the appropriate deterministic variables, d_{mt}, then save the residuals $\hat{e}_{it}$ and $\hat{e}_{it-1}$ from these regressions. Specifically,

$$\hat{e}_{it} = \Delta y_{it} - \sum_{L=1}^{p_i} \hat{\pi}_{iL} \Delta y_{it-L} - \hat{\alpha}_{mi} d_{mt} \tag{2}$$

And

$$\hat{v}_{it-1} = y_{it-1} - \sum_{L=1}^{p_i} \hat{\pi}_{iL} \Delta y_{it-L} - \hat{\alpha}_{mi} d_{mt}$$

(3)

After that take both $\hat{e}_{it}$ and $\hat{v}_{it-1}$ dividing by the regression standard error $\hat{\sigma}_{it}$ also can express more detail of these variable following that (see both equation (4) (5)

$$\tilde{e}_{it} = \frac{\hat{e}_{it}}{\hat{\sigma}_{\varepsilon i}}$$

(4)

$$\tilde{v}_{it-1} = \frac{\hat{v}_{it-1}}{\hat{\sigma}_{\varepsilon i}}$$

(5)

Where $\hat{\sigma}_{\varepsilon i}$ is estimated standard error from each ADF in equation (1) and lastly an estimate of the coefficient δ may be obtained from equation (6)

$$\tilde{e}_{it} = \delta \tilde{v}_{it-1} + \tilde{\varepsilon}_{it}$$

(6)

LLC (2002) show that under the null hypothesis, a modified t-statistics for the resulting $\hat{\delta}$ is asymptotical normally distributed as well as it has been presented give by equation (7).

$$t^*_{\delta}\hat{\delta} = \frac{t_{\delta} - N\tilde{T}\hat{S}_N\hat{\delta}_{\tilde{\varepsilon}}^{-2} STD(\hat{\delta})\mu^*_{m\tilde{T}}}{\sigma^*_{m\tilde{T}}}$$

(7)

Where:

t^*_{δ}=the standard t-statistic for $\hat{\delta}$ =0

$\hat{\delta}_{\tilde{\varepsilon}}^{-2}$ =the estimated variance of the error term ε

$STD(\hat{\delta})$=the standard error of $\hat{\delta}$

$\tilde{T}$ =T-($\sum_{i}^{P_i}$ / N) – 1

LLC (2002) panel unit root test has null hypothesis as panel data has unit root as well as can present below that:

- H_0: null hypothesis as panel data has unit root (assumes common unit root process)

- H_1: panel data has not unit root.

If t* is significant then conclusion is toreject null hypothesis or panel data does not have a unit root. Otherwise if t* is not significant then conclusion is to accept null hypothesis or panel data has unit root.

(2)Breitung Test

Breitung (2000) suggested the use of panel data following transformed data;

$$(\Delta y_{it})^{*} = s_t[\Delta y_{it} - \frac{1}{T-t}(\Delta y_{it+1} + ... + \Delta y_i T)] \quad \text{for t=1,….,T-1} \tag{8}$$

Where

$S_t^2 = (T-t)/(T-t+1)$

$s_{it-1}^{*} = y_{it-1} - y_{io} - \frac{t-1}{T}(y_{iT} - y_{io})$

Δy_{it} =panel data has been differenced

The panel unit root test for the null hypothesis proposed by Breitung (2000) is to reject the null for the small values of the following statistic:

$$\mathrm{B}_{nT} = (\frac{\hat{\sigma}^2}{nT^2}\sum_{i=1}^{n}\sum_{t=2}^{T-1}(y_{it-1}^{*})^2)^{-1/2}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=2}^{T-1}(\Delta y_{it})^{*} y_{it-1}^{*} \tag{9}$$

Where

$\hat{\sigma}^2$ is a consistent estimator of σ2

B_{nT} =t-statistic and it has been used to test panel unit test

The BnT (Breitung (2000) t-statistic) has non-stationary as null hypothesis as well as to show below that:

H_o : null hypothesis as panel data has unit root (assumes common unit root process)

H_1 : panel data has not unit root

If B_{nT} is significant then conclusion that rejects null hypothesis or panel data has not unit root. Otherwise if B_{nT} is not significant then conclusion is to accept null hypothesis or panel data has unit root.

(3) Im, Pesaran and Shin (IPS Test)

This test is generated following simple dynamic linear heterogeneous panel data model and can be written in equation (10)

$$y_{it} = (1-\phi_i)\mu_i + \phi_i y_{i,t-1} + \varepsilon_{it} \tag{10}$$

Where:

y_{it} =panel data

i =1,…., N are cross-section unit or series

t =1,…., T are observed over periods

ε_{it} =error term of equation (10)

Where initial values, y_{io} , to testing the null hypothesis of unit roots ϕ_i =1 for all i (10) can be expressed as

$$\Delta y_{it} = \alpha_i + \beta_i y_{i,t-1} + \varepsilon_{it} \tag{11}$$

Where:

$\Delta y_{it} = y_{i,t}$ - $y_{i,t-1}$ differential into y_i

α_i= (1- ϕ_i) μ_i

β_i=-(1- ϕ_i)

The null hypothesis of unit roots then becomes $H_0 : \beta_i$=0 for all i

Against the alternatives $H_1 : \beta_i$<0, i=1,2….. N_1. β_i=0, $i = N_1$+1, N_1+2, …..N

Im et al. (2003) first calculate the t-statistics for the α_i's in the individual ADF regressions (denoted as $t_i T_i(p_i)$) and then compute their average:

$$\bar{t}_{NT} = \frac{\sum_{i=1}^{N} t_i T_i(p_i)}{N} \tag{12}$$

For the general case with a non-zero *pi* for some cross-sections, the following statistic is asymptotically normally distributed:

$$W_{\bar{t}_{NT}} = \frac{\sqrt{(N)}(\bar{t}_{NT} - N^{-1}\sum_{i=1}^{N} E[t_i T\,(p_i)]}{\sqrt{N^{-1}\sum_{i=1}^{N} \text{var}[t_i T\,(p_i)]}} \tag{13}$$

Where $W_{\bar{t}_{NT}}$ is W-statistics has been used to test panel data based on Im, Pesaran and Shin (2003) techniques. Also this technique has non-stationary as null hypothesis as well as to show below that:

- H_0 null hypothesis as panel data has unit root (assumes individual unit root process)

- H_1 : panel data has not unit root

If $W_{\bar{t}_{NT}}$ is significant then conclusion that reject null hypothesis or panel data does not have unit root. Otherwise if $W_{\bar{t}_{NT}}$ is not significant then the conclusion is to can accept null hypothesis or panel data has unit root.

(4) Fisher-Type Test using ADF and PP-Test

Madala and Wu(1999) suggested all these procedures depend on different ways of combining the observed significance levels (p-values) from the different tests. If the test statistics are continuous, the significance levels π_i (i= 1, 2, . . . , N) are independent uniform (0, 1) variables, and -2$\log_e \pi_i$ has a χ^2 distribution with two degrees of freedom and can be written in equation(14)

$$P_\lambda = -2\sum_{i=1}^{N} \log_e \pi_i \quad (14)$$

Where:

P_λ=Fisher (Pλ) panel unit root test

$-2\sum_{i=1}^{N} \log_e \pi_i$=it has a χ2 distribution with 2N degree of freedom

N=all N cross-section;

In addition, Choi (2001) proposed use these *p*-values the test statistics and are defined as follows :

$$P = -2\sum_{i=1}^{N} Ln(p_i) \quad (15)$$

Where:

$-2Ln(p_i)$=a chi-square distribution with 2 degrees of freedom

$$Z = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \Phi^{-1}(p_i) \quad (16)$$

Where:

Z = Z-statistic panel data unit root test;

N= all N cross-section in panel data;

Φ (•) is the standard normal cumulative distribution function, and p_i is the P-value from the i^{th} test.

Both Fisher (Pλ) Chi-quare panel unit root test and Choi Z-statistics panel data unit root test have non-stationary as null hypothesis as well as to show below that:

- H_0 : null hypothesis as panel data has unit root (assumes individual unit root process)

- H_1 : panel data has not unit root

If both Fisher (Pλ) Chi-quare panel unit root test and Choi Z-statistics panel unit root test are significant then the conclusion is to reject null hypothesis or panel data does not have not unit root. Otherwise both If Fisher (Pλ) Chi-quare panel unit root test and Choi Z-statistics panel unit root test are not significant then the conclusion is to accept null hypothesis or panel data has unit root.

3.3.2 Panel Cointegration Tests

(1) Kao ADF Test

Kao(1999) uses both Dickey-Fuller (DF) tests and an augmented Dickey-Fuller (ADF) test to test the null of no cointegration. Also this test start with the panel residual-based tests for cointegration regression in panel data. The panel regression model is as follow (see equation 17)

$$y_{it} = \alpha_i + \beta x_{it} + e_{it} \qquad i=1,\ldots,N \quad t=1,\ldots T \tag{17}$$

Where y and x are presumed to be non-stationary and: (see equ (18))

$$\hat{e}_{it} = \rho \hat{e}_{it-1} + \upsilon_{it}$$

(18)

where $\hat{e}_{it}$ is the estimate of e_{it} from Eq. (17).

To test the null hypothesis of no cointegration amounts to test: H_0:ρ = 1 in equation(18) and against the alternative that y and x are contitegrated (i,e, H_1 : ρ < 1) Kao (1999) developed both DF-Type test statistics and ADF test statistics were used to test cointegration in panel also both DF-Type (4 Type) test statistics and ADF test statistics (see more detail of 4 type test statistics in Kao (1999).

(2) Pedroni Test

Pedroni (1999) describes the framework for testing for cointegration in panel datasets with *m* = 2*, ...,M* explanatory variables and Pedroni (2004) covers the case for just one regressor. The hypothesized cointegrating regression is

$$y_{it} = \mu_i + \omega_{it} + \varphi_i x_{i,t} + \zeta_{it}$$

(19)

Where:

T= the time dimension (*t* = 1*, ..., T,)*

N= the cross-sectional dimension (*i* = 1*, ..., N.)*

φ_i = slope coefficient in across individual panel member

μ_i = fixed-effects parameter in across individual panel member

ω_{it} = a coefficient with an individual time trend

There are seven residual-based statistics proposed by Pedroni. (see more detail of 7 type test statistics in Pedroni (1999))

The first four are based on pooling along the within-dimension and test the null hypothesis of no cointegration:

H_o: $\gamma_i = 1$

H_1: $\gamma_i = \gamma < 1$

Where *γi* is the autoregressive coefficient of the residual $\hat{\zeta}_{it}$ extracted from estimating the regression equation Eq.19.

The remaining three statistics are based on pooling along the between dimension and again test the null hypothesis of no cointegration:

H_o: $\gamma_i = 1$

H_1: $\gamma_i < 1$

And this research focus on ADF test statistic based on residual-based test follow concept of Kao (1999)to test cointegration in panel and also this research focus on PP-test statistic based on concept of Pedroni to test cointegration in panel. Both ADF–statistics and PP-statistic have same null hypothesis of no cointegration in panel.

3.3.3 Hausman Test

A Hausman test of whether the fixed effects or random effects model is appropriate. Consider the linear model

$$y = bX + e, \tag{20}$$

where

y is univariate

X is vector of regressors,

b is a vector of coefficients

e is the error term.

There are two estimators for *b*: b_0 and b_1. Under the null hypothesis, both of these estimators are consistent, but b_1 is efficient (has the smallest asymptotic variance), at least in the class of estimators containing b_0. In a fixed-effects kind of case. H_o: the random effects would be consistent and efficient.

Under the alternative hypothesis, b_0 is consistent, whereas b_1 isn't.H_1: that random effects would be inconsisitent.

Then the Hausman statistic is:

$$H = (b_1 - b_0)'(\text{var}(b_0) - Var(b_1)^{*}(b_1 - b_0)$$

(21)

Where * denotes the Moore–Penrose pseudoinverse. This statistic has asymptotically the chi-squared distribution with the number of degrees of freedom equal to the rank of matrix Var(b_0) − Var(b_1).

If the Hausman test statistic is large, we reject the null hypothesis, one or both of the estimators is inconsistent. One must use a fixed effects model. If the statistic test is small we accept the null hypothesis, one may get random effects.

3.3.4 Estimating Panel Cointegration Model

There are various approachs for estimating a cointegration test using panel data such as Pedroni (2000, 2001) approach, Chiang and Kao (2000, 2002) approach and Breitung (2002) approach. The various estimators available include with and between groups. But this paper focuses on the OLS estimator and dynamic OLS estimator .OLS and dynamic are both a parametric approaches, While DOLS estimators include lagged first-differenced term are explicitly estimated as well as consider a simple two variable panel regression model:(see detail calculated of OLS,DOLS in equation(23) and (25)

$$y_{it} = \alpha_i + \beta x_{it} + \mu_{it}$$

(22)

Consider a standard panel OLS estimator for the coefficient β of regression model (34) given as

$$\hat{\beta}_{OLS} = (\sum_{i=1}^{N}\sum_{t=1}^{N}(x_{it} - \overline{x}_i)^2)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{N}(x_{it} - \overline{x}_i)(y_{it} - \overline{y}_i)$$

(23)

Where:

i = cross-section data and N is the number of cross-section;

t = time series data and T is the number of time series data;

$\hat{\beta}_{OLS}$ = the coefficient of a standard panel OLS estimator;

x_{it} = all exogenous variable in model;

$\overline{x}_i$ = the average of x_{it} ;

y_{it} = all endogenous variable in model;

$\overline{y}_i$ = the average of y_{it}

Pedroni (2001) has also constructed a between-dimension, group-means panel DOLS estimator that incorporate corrections for endogenous and serial correlation parametrically. This is done by modifying equation (22) to include lead and lag dynamics. Consequently, the DOLS regression becomes

$$y_{it} = \alpha_i + \beta_i x_{it} + \sum_{k=-K_t}^{K_t} \gamma_{it} \Delta x_{it-k} + \mu_{it}^{*} \tag{24}$$

Form this regression, Pedroni(2001) construct the group-mean panel DOLS estimator as

$$\overline{\beta}_{DOLS} = [N^{-1} \sum_{i=1}^{N} (\sum_{t=1}^{T} z_{it} z_{it}^{'})^{-1} (\sum_{t=1}^{T} z_{it} \overline{s_{it}}] \tag{25}$$

Where:

i = cross-section data and N is number of cross-section data;

t = time series data and T is number of time series data;

$\overline{\beta}_{DOLS}$ = Full modified OLS estimator;

z_{it} = the 2(K+1)*1 vector of regressors $z_{it} = (x_{it} - \overline{x}_i, \Delta x_{it-K,}, ..., \Delta P_{it+k})$

$x_{it} = x_{it} - \overline{x}_i$

$\overline{x_i}$ = average of x_{it} ;

Δx_{it-k} =differential term of X

3.4 Characterizing House Price Dynamics

Jud and Winkler (2002) employed a fixed-effects model to examine the factors that influence real housing price changes in a sample of 130 metropolitan areas during the 1984 to 1998 period. The fixed-effects model is traditional demand-side and supply-side factors to analysis housing price influencing factors. This paper follows the framework developed by Jud and Winkler (2002) to investigate the long-term determinants of house price movements.

It is assumed that in each period, in each a city, there is a fundamental value of housing that is largely determined by economic conditions:

$$P_{it}^{*} = f(X_{it}) \tag{26}$$

Where P_{it}^{*} is the real fundamental value of house prices in city i at time t, $f(.)$ is a function and X_{it} is a vector of macroeconomic variables that determine house price fundamentals. We choose two blocks of explanatory variables based on theoretic reasoning or previous empirical work.

The first block of explanatory variables are demand-side factors, including real GDP, the real mortgage rate, due to the high correlation between income and population (0.97), (Yan, Feng, Bao 2010). We use per capita disposable income multiplied by population. We posit that higher income and higher population tend to encourage greater demand for new housing and housing improvements. In addition, the mortgage rate is expected to be negatively related to housing prices. A higher mortgage rate entails higher amortization, which, in turn, impinges on the cash flow of households. This reduces the affordability of new housing, dampens housing demand and pushes down house prices.

While the second block of supply-side factor is captured by land prices. The land supply, which refers to the Land Transaction Price in cities, and land prices can help capture an important driving force of China's property prices—that of local governments' fiscal financing through land sales. In the long run, an increase in land supply tends to bring up house prices. We expect a positive relationship between land transaction price and equilibrium house prices.

The simple model is specified as follows:

$$HP_{it} = f(PPI_{it}, R_{it}, GDP_{it}, LP_{it}) \tag{27}$$

And equation (2) can be expressed in logarithmic form equation number (27).

$$\ln HP_{it} = \alpha + \beta \ln PPI_{it} + \gamma \ln R_{it} + \delta \ln GDP_{it} + \theta \ln LP_{it} + \mu_{it} \tag{28}$$

Where:

$\ln HP_{it}$ =logarithm of selling price index of real estate in city i at time t

$\ln PPI_{it}$ = logarithm of real per capita disposable income in city i at time t

$\ln R_{it}$ = logarithm of real interest rate in city i at time t

$\ln GDP_{it}$ = logarithm of real Gross Domestic Product in city i at time t

$\ln LP_{it}$ = logarithm of real land transaction price index in city i at time t

μ_{it} = independently distributed random error term, with zero mean and constant

α β γ δ θ=parameters of be estimated .assume that β>0, γ<0, δ>0, θ>0

Several things are worth mentioning. First, this research adopted a general-to-specific approach in assessing the determinants of house price fundamentals. That is, we started by including the whole list of possible explanatory factors to investigate their long-term relationship with house prices, using panel data techniques. Only regressors found to be significant at the five percent level are retained.

Second, this paper used four standard panel unit root tests such as LLC (2002) panel unit root test, Breitung (2000) panel unit root test, IPS (2003) panel unit root test, Maddala and Wu (1999) and Choi (2001) panel unit root test(see Table 3.2). Moreover, the panel cointegration test based on Pedroni residual cointegration tests and Kao residual cointegration tests will use to test in panel among the variables.(see table 3.3). Before the estimator, the choice between fixed effects and random effects estimators was base on Hausman test for panel data model. The OLS estimator and DOLS estimator were uses to find the long-run relationship of house price and economic fundamentals (see Figure 3.1).

Table3.2: 4 Standard Method Tests of Panel Data of Unit Root Test

	assumes common unit root process		assumes individual unit root process		
	LLC test	Breitung test	IPS test	Fisher type test	
				ADF test	PP test
H_o	Variable has unit root				
H_1	Variable has not unit root				
Variables PP PPI R GDP LP	If test statistics is significant ↓ Reject H_o ↓ Panel data has not unit root ↓ Variable is stationary		If test statistic is not significant ↓ Accept H_o ↓ Panel data has unit root ↓ Variable is not stationary		

Table 3.3: Panel Cointegration Test Based on Kao Test and Pedronic Test

Type test	Homogeneous coefficient		A common coefficient	Hetergenous coefficient
	Kao test		Pedronic test	
	DF test	ADF test	Panel cointegration statistic	Group mean cointegration statistic
H_0	ρ =1 there is no panel cointegration		$\hat{\rho}_i$=1 there is no panel cointegration	
H_1	ρ <1 among of each data has panel cointegration		$\hat{\rho}_i$ <1 among of each data has panel cointegration	
result	DF test and ADF test statistics are significant ↓ Reject H_0 ↓ among of each data has panel cointegration		PP statistics are significant ↓ Reject H_0 ↓ among of each data has panel cointegration	

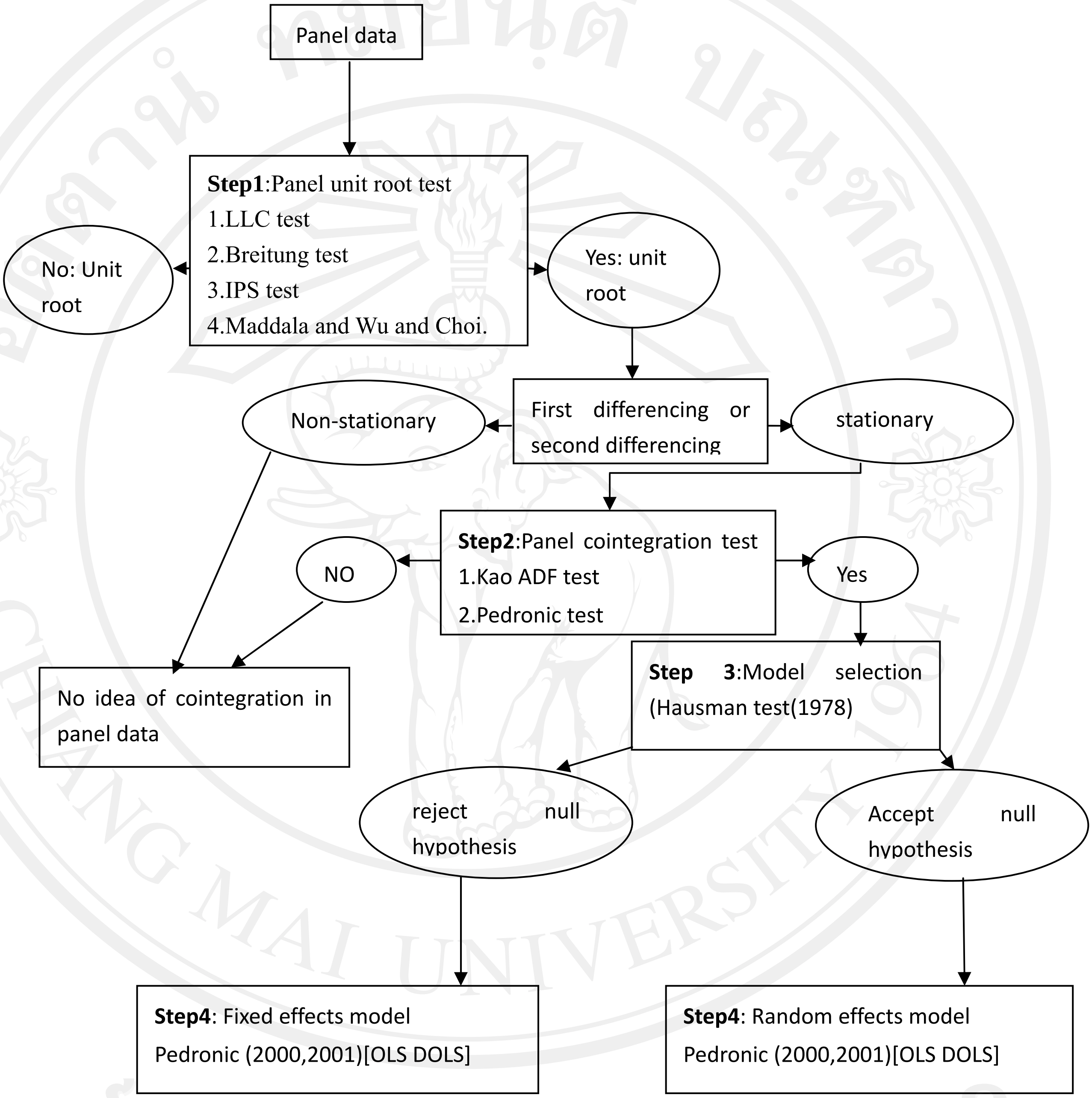

Figure 3.1: Panel Cointegration Framework

Source: modified from Harris and Sollis(2006)

3.5 The Real Estate Bubble Theory

The specific characteristics of housing have made it of important interest for researchers interested in asset price bubbles. Housing is not only a durable good with a long life, but also an immovable estate.

All characteristics make the real estate price which is determined by the rational expectation price. Thus, the deviation between the expectations of future and the actual situation will determine the an onset of real estate price bubble. This paper took the real estate price bubble into consideration when studying the economic factors influencing the urban real estate prices price.

3.5.1 The Definition of a Real Estate Bubble

The term ‘‘bubble’’ is widely used but rarely clearly defined. In The New Palgrave Dictionary (2008), Kindleberger defines a bubble as: a sharp rise in price of an asset or a range of assets in a continuous process, with the initial rise generating expectations of further rises and attracting new buyers – generally speculators interested in profits from trading rather than in its use or earning capacity. The rise is then followed by a reversal of expectations and a sharp decline in price, often resulting in severe financial or economic crises. Stiglitz (1990) often quoted definition as follows: ‘‘the basic intuition is straightforward: if the reason that the price is high today is only because investors believe that the selling price will be high tomorrow – when ‘fundamental’ factors do not seem to justify such a price – then a bubble exists.’’ Flood and Hodrick (1990) define a bubble as a deviation of the current market price of the asset (such as stocks or real estate) from the value implied by market fundamentals. Later, Smith and Smith (2006) present a more extended definition. They define a bubble as a situation in which the market prices of the assets rise far above the present value of the anticipated cash flow from the asset (what Kindleberger called the asset’s use or earning capacity).

3.6 Detecting Housing Bubble

A number of theories have been suggested to analyze housing bubble. One way to assess whether a housing bubble exists is to calculate the deviation between the observed housing price and the rational expectation price. In fact, the theoretic modeling of housing bubble was not developed until the school of rational expectations was formed in the later 20th century. Muth (1961) paved the way by first putting forward the hypothesis of rational expectations. The theory of rational expectations then became mature during the 1980s when the school of rational expectations formulated the mechanisms of bubble formation. The rational expectation model is built on the theory of asset pricing by incorporating the present value model. The theory suggests that housing price depends upon the return and utility of a property, i.e. the price is not only influenced by the use-return of housing during its tenure but also by the capital-return produced due to selling the property when the tenure expires. The rational expectation model has been widely used in identifying bubbles in earlier studies (Flood and Hodrick, 1990; Kim and Suh, 1993; Chan et al., 2001; Xiao and Tan, 2007)

The basic framework for analyzing bubbles in this approach is as follows. According to the theory of rational expectation, a bubble must grow fast enough to earn the expected return, that is, investors have to sell an asset for more in the future than it costs today, as the bubble pays no flow distribution. If a property's price does not reflect all the information about its intrinsic value, there exist ‘‘unexploited profit opportunities.’’ Then, someone knowing this can buy or sell the property to make a profit, thus driving the price toward equilibrium. All prices in markets will not reflect market fundamentals like future streams of profits and dividends until all profit opportunities have been exploited.

Mikhed and Zemˇcík (2009) employed panel data tests for unit roots and cointegration to determine whether house price reflect house-related earnings. The theoretical model described by Mikhed and Zemˇcík(2009) is briefly illustrated in the following.

Mikhed and Zemˇcík (2009) regarded a house as an investment asset and used a standard present-value formula to derive implications for the relationship between house prices and cash flows. The standard present-value formula is shown in the following:

$$P_{i,t} = E_t[\frac{R_{i,t+1} + P_{i.t+1}}{1+D}], \quad \text{i=1,….N,} \tag{29}$$

Where $P_{i,t}$ is a regional house price index, E_t is a mathematical expectation conditional on information at time t, $R_{i,t}$ is a cash flow associated with owning a house (i.e., rent), and r denotes a constant discount rate. This formula holds for all periods, so the regional house price index in periods t+1 can be described as follows

$$P_{i,t+1} = E_{t+1}[\frac{R_{i,t+2} + P_{i,t+2}}{1+r}] \tag{30}$$

Thus, equation (29) can be rewritten as:

$$P_{i,t} = E_{i,t}[\frac{R_{i,t+1}}{1+r} + \frac{R_{i,t+2}}{(1+r)^2} + \dots + \frac{R_{i,t+k}}{(1+r)^k} + \frac{P_{i,t+k}}{(1+r)^k}] \tag{31}$$

Mikhed and Zemˇcík (2009) proposed that a moment the no-bubble condition is:

$$\lim_{k\to\infty} E_t[\frac{P_{i,t+k}}{(1+D)^k}] = 0 \tag{32}$$

Which yield

$$P_{i,t}^F = \sum_{j=1}^{\infty} \frac{1}{(1+D)^j} E_t[C_{i,t+j}] \tag{33}$$

which is often referred to as price reflecting fundamentals.

Mikhed, Petr Zemˇcík(2009) defined the spread between the house price and cash flows as

$$S_{i,t} \equiv P_{it} - \frac{1}{D} C_{i,t}$$

(34)

Based on the no-bubbles condition, $S_{i,t}$ can be rewritten as

$$S_{i,t} = \frac{1}{D} E_t \sum_{j=1}^{\infty} \frac{\Delta C_{i,t+j+1}}{(1+D)^j} = \frac{1}{D} E_t [\Delta P_{i,t+1}]$$

(35)

Since $P_{i,t} / C_{i,t} = 1/D$ if $S_{i,t}$ =0, the stationary of $S_{i,t}$ implies that the house price-to-rent ratio will also be stationary. Alternately, based on Equation (35), if the series of rent is I(1) or I(2) and the no-bubbles condition holds, $S_{i,t}$ must be stationary. Therefore, there are two ways to ascertain whether or not bubbles exist. One is testing the stationary of the sale price house and rental price house, if housing price is not-stationary but rental price is stationary and the other is employing the cointegration test for the two variables, house sale and rent. To do so, we conduct the IPS test for unit roots and the Pedroni test for cointegration.

This research will select China's sales price indices of real estate and renting price indices of real estate represent respectively the urban real estate price level and rent level at the same period time. To investigate whether there is a long-run equilibrium relationship between house price and rents corresponding to the present value formula Eq.(29). To do so, we only conduct the IPS test for unit roots and Pedroni test for cointegration. The panel data tests used in this paper is briefly introduced before in econometric theory.