ABSTRACT

My research focuses on issues related to housing returns in the U.S. housing market.

The first essay, “Explaining the Variance of the House Price Appreciation Rate” is motivated by an interesting geographic pattern in the variance of house price appreciation rate across the United States. Specifically, I find that the variance of the house price appreciation rate (or housing returns) is significantly higher along the coastal areas than inland areas of the United States. This paper investigates the geographic pattern both theoretically and empirically. The theoretical model of this paper is founded on the monocentric urban area model developed by Capozza and Helsley (1988 & 1990). Based on extensions to this model, the relationships between the variance of the house price appreciation rate and exogenous variables related to the urban economy are identified. My empirical results are largely consistent with the model’s predictions. In particular, I show that the variance of housing returns is significantly and positively related to the variance of household income, the degree of land regulation (or growth control), land leverage, and transportation costs in each MSA.

The second essay, “Skewness of the Housing Returns and Housing Tenure Choice” focuses on the third moment of the housing returns and investigates its influence on a household’s tenure choice (the choice between owning and renting). In the finance literature, it is generally accepted that, holding the mean and variance of the returns constant, risk-averse households have a preference for financial investments with positively skewed
returns. Homes are a large investment for most American families and thus I examine in this paper whether the preference for a positive skewness also exists in the housing market. Based on the model of housing tenure choice developed by Henderson & Ioannides (1983) and Fu (1995), I derive the relationship between housing tenure choice and the skewness of housing returns. The empirical investigation supports my theoretical conjecture that the skewness of housing returns is positively related to a household’s likelihood of being a home owner.

The third essay, “A Spatial Econometric Approach to MSA Level Housing Returns” employs newly developed spatial panel data models to investigate the spillover mechanism in the U.S. housing market, focusing on determining whether housing returns are more highly correlated among geographically nearby cities or among cities that are economically similar. Two different weight matrices are integrated into the spatial econometric model; the first weight matrix is constructed based on geographic distances among MSAs and the second matrix uses measures of economic similarities among cities. My empirical regression results show that both the inter-MSA spatial correlations in housing returns are significant and positive, but with the geographic correlation appearing to be stronger and more significant than the correlation based on economic similarities. The empirical investigation also provide evidences that housing prices changes in one MSA area are correlated not only to their own lagged changes, but also correlated to the present and lagged housing prices changes in their neighboring or similar MSAs.
Dedicated to my parents and my husband
ACKNOWLEDGMENTS

I am deeply grateful to my advisor Professor Donald Haurin, for his guidance, inspiration and encouragement throughout my dissertation research. His encyclopedic knowledge in economics, rigorous attitude towards research and incessant patience in mentoring me have greatly contributed to the completion of this thesis.

I thank Professor Lixin Ye, Professor Bruce Weinberg and Professor Lucia Dunn for stimulating discussions and valuable suggestions that led to great improvements of this work, Professor Lung-fei Lee, Professor David Blau, and Professor Huanxing Yang for helpful advice at various stages of my study at Ohio State. I also would like to extend my special thanks to Professor Hajime Miyazaki. His warmhearted help makes the challenging years of PhD study so much easier. I have been truly blessed with many amazing friends over the years at Ohio State. I thank them all for their friendship.

Finally, I would like to thank my parents, Guolin Liu and Xiaoping Wang, and my beloved husband, Xiaodong Liu, for their unconditional love, support and belief in me. They make all my effort worthwhile.
VITA

November 1, 1979 . . . . . . . . . . . . . . . . . . . . . . . . . . . . Born — Quzhou, China

2001 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . B.A. Economics, Fudan University

2005 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . M.A. Economics, The Ohio State University

2009 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . M.A.S. Statistics, The Ohio State University

FIELDS OF STUDY

Major Field: Economics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
</tr>
<tr>
<td>Dedication</td>
</tr>
<tr>
<td>Acknowledgments</td>
</tr>
<tr>
<td>Vita</td>
</tr>
<tr>
<td>List of Tables</td>
</tr>
<tr>
<td>List of Figures</td>
</tr>
</tbody>
</table>

## Chapters:

1. EXPLAINING THE VARIANCE OF THE HOUSE PRICE APPRECIATION RATE 1
   1.1 Introduction 1
   1.2 The Model 6
       1.2.1 Household’s Budget Constraint 6
       1.2.2 Stochastic Assumptions 7
       1.2.3 Price of Urban Land 8
       1.2.4 Determinants of the Variance of the House Price Appreciation Rate Implied by the Model 10
       1.2.5 Extensions of the Original Model 16
   1.3 Testable Hypotheses Derived from the Models 24
   1.4 Empirical Investigation 26
       1.4.1 Cross-Sectional Regression Model 27
       1.4.2 Data 28
       1.4.3 Small Sample Regression 31 |
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4.1 Econometric Theory and Methodology of Missing Data Imputation</td>
<td>84</td>
</tr>
<tr>
<td>for Spatial Panel Models</td>
<td></td>
</tr>
<tr>
<td>3.4.2 The Spatial Econometric Model with Two Weight Matrices and</td>
<td>88</td>
</tr>
<tr>
<td>Missing Data</td>
<td></td>
</tr>
<tr>
<td>3.4.3 Estimation Results</td>
<td>89</td>
</tr>
<tr>
<td>3.5 The Spatial Dynamic Panel Model with One Weight Matrix</td>
<td>90</td>
</tr>
<tr>
<td>3.6 Conclusion</td>
<td>94</td>
</tr>
<tr>
<td>3.7 Tables</td>
<td>96</td>
</tr>
<tr>
<td>Bibliography</td>
<td>99</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>43</td>
</tr>
<tr>
<td>1.2</td>
<td>44</td>
</tr>
<tr>
<td>1.3</td>
<td>45</td>
</tr>
<tr>
<td>1.4</td>
<td>46</td>
</tr>
<tr>
<td>1.5</td>
<td>47</td>
</tr>
<tr>
<td>1.6</td>
<td>47</td>
</tr>
<tr>
<td>1.7</td>
<td>48</td>
</tr>
<tr>
<td>2.1</td>
<td>70</td>
</tr>
<tr>
<td>2.2</td>
<td>71</td>
</tr>
<tr>
<td>3.1</td>
<td>97</td>
</tr>
<tr>
<td>3.2</td>
<td>97</td>
</tr>
<tr>
<td>3.3</td>
<td>97</td>
</tr>
<tr>
<td>3.4</td>
<td>98</td>
</tr>
<tr>
<td>3.5</td>
<td>98</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Average House Price Appreciation Rate and Variance of House Price Appreciation Rate by Census</td>
<td>49</td>
</tr>
</tbody>
</table>
CHAPTER 1

EXPLAINING THE VARIANCE OF THE HOUSE PRICE APPRECIATION RATE

1.1 Introduction

This paper is motivated by an interesting pattern I discovered from a geographic investigation of the MSA (metropolitan statistical area) level house price appreciation rates across the United States. More specifically, in the study, I obtained the MSA level house price index data (dated from 1975 to 2006) from Fannie Mae and Freddie Mac; I then grouped these 379 MSAs by 9 census regions and calculated their average real house price appreciation rate (or real housing returns\(^1\)) and average variance of real housing returns respectively.

Figure (1.6.2) is a map of the USA with average house price appreciation rate and average variance of appreciation rate labeled in each text box for each census region. Some interesting geographic patterns were detected. First, we can easily see that coastal regions usually have higher average house price appreciation rates than inland regions. For example, housing appreciation rates of the inland regions such as West North Central, East North Central and East South Central are all at the 1% level, and West South Central region’s appreciation rate is even lower, which is only 0.1%. On the other hand, the coastal

\(^1\)In this paper, we define housing returns as house price appreciation rate, so basically they have the same meaning.
regions have much higher average real appreciation rates. The Pacific region, for example, has the highest appreciation rate, which is 3.7%. And all the other coastal regions including mountain area have appreciation rates around 2% or higher.

Another notable feature we can find from the map is that the average variance of house price appreciation rate is also significantly higher along the coastal regions than it is among inland areas. For instance, the average variance of house price appreciation rate is as high as 0.006 for New England and Pacific Region, while it is only 0.001 for most inland areas like East North Central, West North Central, East South Central, etc. Table (1.6) in the appendix presents the top 10 MSAs with most volatile housing returns, and table (1.7) gives a list of 10 MSAs with least volatile housing returns. We can see that all the MSAs in table (1.6) are from coastal regions, for example, California. On the other hand, most of the MSAs listed in table (1.7) are from inland areas such as Illinois, Wisconsin, Tennessee, etc. This geographic pattern indicates that coastal regions, which enjoy higher average house price appreciation rates, usually face more volatile housing returns as well.

Are there any existing models or theories which can explain the spatial patterns as described above? The literature regarding the house price appreciation rates is somewhat more developed than the literature about the variance of house price appreciation rate. According to Malpezzi (1996), one of the best studies of cross-MSA house prices is that of Ozanne and Thibodeau (1983). In that paper, the authors set up a cross-section model to explain hedonic house price indices for 59 major metropolitan areas. Explanatory variables in their model include the median household income, the number of households, an MSA-specific nonhousing price index, the mortgage interest rate, etc. Abraham and Hendershott (1993) expressed real house price changes as a function of the rate of changes in
employment, real income growth, real construction cost inflation and changes in real after-tax interest rates. Case and Shiller (2003) described the house price bubble\(^2\) and argued that although income growth alone explains virtually all housing price increases for more than 40 states of the United States; there was still evidence for the existence of house price bubbles according to their recent survey conducted in 2003 in four metropolitan areas: Los Angeles, San Francisco, Boston and Milwaukee. Generally speaking, in the existing literature on house price changes, a large number of papers focus on demand-side factors such as income and demographic variables and consider them as major determinants of changes in house price.

During recent years, more and more studies have started to emphasize the effects of supply-side variables on house price dynamics. For example, Malpezzi (1996) studied supply side factors with a particular focus on determinants related to regulation. His empirical work suggested that higher regulation levels raise house prices and lower homeownership rates. Another innovative part of Malpezzi’s work is that he constructed indices which reflect regulatory regimes in different markets. Glaeser, Gyourko and Saks (2005) indicated that the elasticity of housing supply plays a very important role in determining the level of house price appreciation. To be more specific, in places with inelastic housing supply, a positive demand shock (e.g., increase in productivity) will result in a higher level of housing prices and an unchanged population. In contrast, in places where the housing supply is more elastic, a positive demand shock usually leads to bigger cities and higher population levels. Changes in house prices are relatively much smaller. And usually, the inelasticity of housing supply is closely related to restrictive zoning and other land use regulations. McMillen and McDonald (1993, 1998, 1999, 2002) did a series of studies on the effects of

\(^2\)The term "bubble", according to Case and Shiller, refers to a situation in which excessive public expectations of future price increases cause the price to be temporarily elevated.
zoning ordinances introduced to Chicago in 1923. After controlling for initial land use and the endogeneity of zoning decisions, they found that residential zoning led to higher land value growth rates than commercial zoning.

Compared with the relatively more established literature on house price appreciation, very limited amount of research is devoted to explaining the variance of house price appreciation, despite the importance of this topic. As claimed by Crone and Voith (1999), for most American families, housing investment is probably the largest single investment and also the major part of their wealth portfolio. Furthermore, the housing market is also an important part of the U.S. economy and quite comparable to other major economic components. For example, the total value of owner occupied housing in the U.S. has exceeded 20 trillion dollars according to the 2006 Census Bureau American Factfinder Community Survey. During the same time period, the total assets of commercial banks in the U.S. was about 9.7 trillion according to the Federal Reserve Statistical Release. Hence, understanding the variation of housing returns is important not only for an individual’s welfare consideration, but also crucial for the overall economy.

Among the very small amount of papers in the literature on this topic, Berkovec (1989) developed a theoretical model which demonstrated the positive relationship between risk and expected returns in the owner-occupied housing market. Gat (1994) used this model to examine risk and return in neighborhood housing markets in Tel Aviv and empirically found a significantly positive relation between risk and return in the housing market. Crone and Voith (1999) introduced a model based on homeowner’s utility maximization and their empirical results also supported the positive relationship between housing returns and the volatility of returns. All these works emphasized the existence of trade-off between risk and return in the housing market, which, to some extent, provided some insights into this
problem, both theoretically and empirically. However, their arguments are mostly focused on consumer’s utility maximization instead of a general equilibrium problem. Furthermore, their research did not answer the questions about why housing markets in some regions are more volatile than in others.

Miller and Peng (2006) used a six-equation VAR model to study the dynamic interaction between the volatility of house price appreciation and a few economic and demographic variables, including the rate of home value appreciation, the per capita personal income growth rate, the population growth rate, the change of unemployment rate, etc. They found evidence that volatility is Granger-caused by the house price appreciation rate and GMP growth rate. Cannon, Miller and Pandher (2006) used ZIP code level housing data to study the cross-sectional role of volatility, price level, stock market risk and idiosyncratic volatility in explaining housing returns. The empirical findings of these works conveyed some enlightening information about the cross sectional pattern of housing returns volatility but neither of them provided theoretical models for their empirical conjectures.

The main intention of this paper, therefore, is to set up an appropriate model and find out potential variables from the model that can explain the geographic pattern of significant differences in the average variance of house price appreciation rates across different regions of the United States. More specifically, the theoretical model in this paper is founded on the monocentric urban area model developed by Capozza and Helsley (1988 & 1990). Based on extensions to this model, the variance of house price appreciation rate is derived and variables which may be closely related to the variation of the housing returns are identified from the model. Directions of the relationships are discussed and empirical work also is provided which tests the conjectures generated by the theoretical model.
The paper proceeds as follows. Section 1.2 introduces the basic model and calculates the variance of house price appreciation rate followed by comparative static analysis. Extensive studies are then conducted to analyze the variance of the housing returns under some specific conditions. Section 1.3 presents a series of testable hypotheses implied by the theoretical models of section 1.2. Section 1.4 provides empirical regression results and the last section concludes.

1.2 The Model

The theoretical model in this paper is founded on the monocentric urban area model developed by Capozza and Helsley (1988 & 1990). More specifically, in their papers, Capozza and Helsley derived urban land prices in a fixed lot size model under both certainty and uncertainty conditions, and the model presented here extends their model to analyze the variance of the housing returns.

To begin, suppose for simplicity that there is a small urban area with all the employment concentrated at the central business district (CBD). Households commute to this CBD to work every weekday and the distance from home to CBD is denoted by \( z \). The distance from the boundary of the urban area to the CBD at time \( t \) is denoted by \( z(t) \). Furthermore, consumption of land is fixed at one unit per household, at any time and any location. Therefore, fundamentally, this is a monocentric urban area model with fixed lot size.

1.2.1 Household’s Budget Constraint

Aside from the consumption of fixed size of land, household also derives utility from a composite numeraire good \( x \). Therefore, for a typical household living at location \( z \), the budget constraint is

\[
y(t) = x + R(t, z) + kz, \tag{1.1}
\]
where \( y(t) \) is household income at time \( t \), \( R(t, z) \) is land rent and \( k \) is per-unit-of-distance transportation cost.

At equilibrium, as we assume costless migration, the utility level achieved by each household in the urban area has to be equalized, that is,

\[
V(R(t, z), y(t)) = \overline{V},
\]

(1.2)

where \( V(.) \) is the indirect utility function and \( \overline{V} \) is the utility level of every household. With fixed lot size, (1.2) implies that in order to keep the national utility level at \( \overline{V} \), all households should consume the same amount of consumption goods, \( \overline{x} \), so that

\[
U(\overline{x}, 1) = \overline{V},
\]

(1.3)

where \( U(.) \) is the direct utility function of households. Accordingly, the function of land rent is:

\[
R(t, z) = y(t) - (\overline{x} + kz).
\]

(1.4)

### 1.2.2 Stochastic Assumptions

In this model, the household income \( y(t) \) is assumed to follow a stochastic process. The story behind this assumption of stochastic household income, as told by Capozza and Helsley, is that the urban area produces products and exports them to an expanding world market. Shocks to the income \( y(t) \) thus come from the shocks in world demand, which is transmitted through the labor market.

More precisely, the sequence of random variable \( \{y(t), t \geq 0\} \) is a Brownian motion process with drift \( \mu > 0 \) and variance parameter \( \sigma^2 \). Therefore, "household income rises over time in an uncertain fashion". According to the basic properties of Brownian motion, we then get the following implications for the stochastic income sequence:
(i) $y(t+s) - y(t) \sim N(\mu_s, \sigma^2 s)$ for all $t$, which is known as the stationary increments property;

(ii) for all $t_0 < t_1 < \ldots < t_n$, $y(t_1) - y(t_0), y(t_2) - y(t_1), \ldots, y(t_n) - y(t_{n-1})$ are independent random variables. This is the independent increments property;

(iii) $y(0) = 0$, that is, the initial value of income is zero;

(iv) $y(t) = \mu t + \sigma W(t)$, where $\{W(t), t \geq 0\}$ is a standard Brownian motion process with drift 0 and variance parameter 1.

From equation (1.4), we can clearly see that, $R(t, z)$ should also follow a stochastic process, since it is a function of $y(t)$. To be more specific, the sequence $\{R(t, z), t \geq 0\}$ should also have properties (i) and (ii), with (iii) replaced by

(iii') $R(0, z) = -(\bar{x} + k z)$

and property (iv) replaced by

(iv') $R(t, z) = R(0, z) + \mu t + \sigma W(t)$

Property (i), (ii) and (iv') imply that we can represent $R(t + s, z)$ as:

$$R(t + s, z) = R(t, z) + \mu s + \sigma W(s).$$  \hspace{2cm} (1.5)

### 1.2.3 Price of Urban Land

Derivations of urban land price follow closely those of Capozza and Helsley (1988 & 1990)’s. Thus only the most important equations are presented here. Detailed procedures for deriving the model can be found in Capozza and Helsley (1988 & 1990).

As defined by Capozza and Helsley (1990), "in a competitive market, the price of land equals the expected present value of future land rents." Let $r$ denote the common discount rate shared by risk neutral land owners, and we assume that the current rent level $R(t, z)$ is common knowledge. Therefore, $P^{u}(t, z)$, which is the price of one unit of urban land at
time $t$, location $z$, can be defined as:

$$P^u(t, z) = E \left\{ \int_t^\infty R(\tau, z)e^{-r(\tau-t)} d\tau \mid R(t, z) \right\}, \quad z \leq \bar{z}(t). \quad (1.6)$$

The equation for the price of one unit of agricultural land is different. Originally, land in agricultural use earns rent $A$. Let $C$ denotes the capital cost of converting a unit of agricultural land to urban land. For example, it can be the cost related to constructing roads, providing water, sewer and electricity, etc. So, as stated by Capozza and Helsley, "urban land is produced from agricultural land by adding a certain amount of capital cost per unit." Once the land is converted to urban use, it starts to receive urban rents from then on and we assume that the decision to convert agricultural land to urban use is irreversible. Therefore, the price of one unit of agricultural land at time $t$, location $z > \bar{z}(t)$ is

$$P^a(t, s, z) = E \left\{ \int_t^{t+s} Ae^{-r(\tau-t)} d\tau + \int_{t+s}^\infty R(\tau, z)e^{-r(\tau-t)} d\tau - Ce^{-rs} \mid R(t, z) \right\}, \quad (1.7)$$

where $t+s$ is the date the agricultural land is converted to urban land and $s$ is a "stopping time". Landowners then maximize expected value of land by choosing the optimal conversion time. That is, landowners solve the maximization problem

$$\max_s P^a(t, s, z)$$

When $\sigma \neq 0$, that is, when there is uncertainty involved in household’s income, the model can be solved as a hitting time problem thanks to the special properties possessed by Brownian motion process, which the sequence of land rents follows. The optimal conversion rule, turns out to be:

$$R^* = A + rC + \frac{r - \lambda \mu}{\lambda r}, \quad \quad (1.8)$$

3Technically, it is still possible to convert urban land back to agricultural use. But the cost incurred is usually huge considering the existing urban constructions, the already changed environment, etc., which makes the reversion unrealistic. So here we assume irreversibility for the sake of simplicity.
where $R^*$ stands for the level of urban land rent at which it is optimal to convert land from agricultural use to urban use. $\lambda$ here is a parameter and more specifically, $\lambda = \sqrt{\mu^2 + 2ae^2r - \mu}$ and we assume that $\lambda \leq r/\mu$.

Accordingly, after knowing $R^*$, we can now get the equilibrium urban land rent function:

$$R(t, z) = A + rC + \frac{r - \lambda \mu}{\lambda r} + k(\bar{z}(t) - z), \ z \leq \bar{z}(t). \quad (1.9)$$

The price of urban land with uncertainty is thus

$$P^u(t, z) = \frac{A}{r} + C + \frac{\mu}{r^2} + \frac{r - \lambda \mu}{\lambda r^2} + \frac{k}{r}[\bar{z}(t) - z], \ z \leq \bar{z}(t). \quad (1.10)$$

So, the price of urban land here consists of agricultural land rent, $A/r$, the conversion cost $C$, a growth premium $\mu/r^2$, the irreversibility premium $(r - \lambda \mu)/\lambda r^2$, and the value of accessibility, $k/r[\bar{z}(t) - z]$.

### 1.2.4 Determinants of the Variance of the House Price Appreciation Rate Implied by the Model

#### Urban Land Price and House Price

So far, what we have derived are equations of land prices instead of house prices. Because the dynamics of house price appreciation is the major concern of this paper, we need to relate land price to house price. Consider a standard Cobb-Douglas production function of housing with constant returns to scale$^4$:

$$H^* = B N^\alpha K^{1-\alpha} \quad (1.11)$$

$^4$I choose a Cobb-Douglas production function because it gives the cleanest relationship between changes of land price and changes of house price. We can also use other forms of production functions, for example, a general CES production function, which gives similar results.

Labor factor can also be put into the production function. However, although labor cost might differ spatially, the level of labor cost is usually stable and therefore does not contribute significantly to the volatility.
where $B$ is a conversion constant, $N$ is the input of land, $K$ is the input of capital, $\alpha$ is the share of land in producing housing and $1 - \alpha$ is the share of capital in the production of houses. We assume that the input market is competitive and factors are paid in compliance with their value of marginal product. Let $P^h(t, z)$ denote the price of house at time $t$ location $z$, $P_k$ represent the price of capital which is a constant,\(^5\) and $P^u(t, z)$ stand for the price of urban land at time $t$ location $z$. We then get equations:

$$P_k = P^h(t, z)(1 - \alpha)H/K$$  \hspace{1cm} (1.12)

$$P^u(t, z) = P^h(t, z)\alpha H/N.$$  \hspace{1cm} (1.13)

Solve (1.12) and (1.13) for $N$ and $K$ and substitute them into the housing supply function (1.11), yielding:

$$P^h(t, z) = \Phi P^u(t, z)\alpha$$  \hspace{1cm} (1.14)

where $\Phi = P_k^{1-\alpha}/B\alpha^\alpha(1 - \alpha)^{1-\alpha}$, which is a constant.

Equation (1.14) implies that house price is closely related to land price. And it also leads to the relation

$$\frac{\partial P^h(t, z)}{P^h(t, z)} = \alpha \frac{\partial P^u(t, z)}{P^u(t, z)}$$  \hspace{1cm} (1.15)

which indicates that the percentage change in house price is closely related to the percentage change in urban land price.

Therefore, it suffices to look at the determinants of urban land price and find out variables which may affect the volatility of the house price appreciation rate.

\(^5\)Obviously, $P_k$ could also vary overtime. However, letting $P_k$ vary is not so relevant in this paper because $P_k$ usually is spatially uniform and therefore does not play a role in explaining the spatial differences. So without loss of generality, here we assume it is a constant.
Variance of the House Price Appreciation Rate

We have showed that the price of urban land at location $z$ time $t$ when $\sigma \neq 0$ is

$$P^\mu(t, z) = \frac{A}{r} + C + \frac{\mu}{r^2} + r - \frac{\lambda \mu}{r^2} + \frac{k}{r}[\bar{z}(t) - z], \ z \leq \bar{z}(t).$$

(1.16)

And similarly, in this case, the endogenous $\bar{z}(t)$ is determined by

$$\bar{z}(t) = \frac{1}{k} \left( y(t) - \bar{x} - A - rC - \frac{r - \lambda \mu}{\lambda r} \right).$$

(1.17)

Substitute out $\bar{z}(t)$ in (1.16) using (1.17), the price of urban land becomes

$$P^\mu(t, z) = \frac{1}{r} \left( y(t) - \bar{x} - k z + \frac{\mu}{r} \right)$$

$$= \frac{1}{r} \left[ \mu t + \sigma W(t) - \bar{x} - k z + \frac{\mu}{r} \right].$$

(1.18)

The house price appreciation rate $\delta$ given location $z$ therefore equals

$$\delta_{t+1, t, z} = \frac{\mu + \sigma W(t + 1) - \sigma W(t)}{\mu t + \sigma W(t) - \bar{x} - k z + \frac{\mu}{r}}.$$

(1.19)

The variance of $\delta$, $\text{var}(\delta) = E\delta^2 - (E\delta)^2$. By the special properties of Brownian motion, $[W(t + 1) - W(t)]$ and $W(t)$ are independent and $E[W(t)] = 0$, we thus get

$$E\delta = \frac{E[\mu + \sigma W(t + 1) - \sigma W(t)]}{E[\mu t + \sigma W(t) - \bar{x} - k z + \frac{\mu}{r}]} = \frac{\mu}{\mu t - \bar{x} - k z + \frac{\mu}{r}}$$

$$E\delta^2 = \frac{E[\mu + \sigma W(t + 1) - \sigma W(t)]^2}{E[\mu t + \sigma W(t) - \bar{x} - k z + \frac{\mu}{r}]^2} = \frac{\mu^2 + \sigma^2}{(\mu t - \bar{x} - k z + \frac{\mu}{r})^2 + \sigma^2 t}.$$  

(1.20)

Therefore, the variance of the house price appreciation rate when the future is uncertain, will be

$$\text{var}(\delta) = \frac{\mu^2 + \sigma^2}{(\mu t - \bar{x} - k z + \frac{\mu}{r})^2 + \sigma^2 t} - \frac{\mu^2}{(\mu t - \bar{x} - k z + \frac{\mu}{r})^2}.$$  

(1.21)

From equation (1.21), we can see that the variance of the house price appreciation rate from this model, given location $z$ and when $\sigma \neq 0$, is related to $\mu$, the drift of the stochastic household income, $\sigma^2$, the variance of income, $\bar{x}$, the composite numeraire good, $k$, the per-unit-of-distance transportation cost and $r$, the discount rate or interest rate.
Comparative Static Analysis  Given (1.21), we can conduct a comparative static analysis to find out how these exogenous variables affect the variance of the housing returns.

First, consider the effect of income volatility $\sigma^2$ on the variance of the housing returns. For the sake of notation simplicity, let $\mu t - k z + \frac{\mu}{r} = G > 0$, (because $G = E [P^\mu(t, z)] * r$, and we assume the expected price of urban land is positive). Therefore,

$$\frac{\partial \text{var}(\delta)}{\partial \sigma^2} = \frac{G^2 + \sigma^2 t - (\mu^2 + \sigma^2)t}{[G^2 + \sigma^2 t]^2} = \frac{G^2 - \mu^2 t}{[G^2 + \sigma^2 t]^2} > 0. \quad (1.22)$$

This derivative is positive because $\text{var}(\delta) = \frac{\mu^2 + \sigma^2}{G^2 + \sigma^2 t} - \frac{\mu^2}{G^2} > 0$.

(1.22) tells us that the variance of the housing returns is increasing in the volatility of household income. So in explaining cross-sectional differences in housing returns variance, we hypothesize that housing returns are more volatile in areas where household income is more volatile.

Second, the derivative of $\text{var}(\delta)$ with respect to the drift $\mu$ of household income is:

$$\frac{\partial \text{var}(\delta)}{\partial \mu} = \frac{-2\mu \sigma^2 t G^4 - 2\mu \sigma^4 t^2 G^2 - 2\sigma^2 G(t + \frac{1}{r})(G^4 - 2\mu^2 t G^2 - \mu^2 \sigma^2 t^2)}{[G^2 + \sigma^2 t]^2 G^4} \quad (1.23)$$

The sign of the above derivative, however, is ambiguous given the current information available. Equation (1.22) tells us that $G^2 > \mu^2 t$. If we impose an additional constraint on $G$, such that

$$G^2 \geq \mu^2 t + \mu^2 t \sqrt{1 + \frac{\sigma^2}{\mu^2}}, \quad (1.24)$$

then $\frac{\partial \text{var}(\delta)}{\partial \mu} < 0$, which implies that the variance of the housing returns is decreasing in the drift level of household income. And because $E [y(t)] = \mu t$, $\mu$ thus also measures the expected level of household income at time $t$. Therefore, under the additional assumption (1.24), that is, when the expected price of urban land $E [P^\mu(t, z)]$, given location $z$ and time $t$, is greater than or equal to a particular level $\left(\mu^2 t + \mu^2 t \sqrt{1 + \sigma^2 / \mu^2}\right) / r^2$,
then housing returns are less volatile in areas where the expected level of household income is higher. On the other hand, if the additional assumption (1.24) is not imposed, the relationship between the variance of the housing returns and $\mu$ is ambiguous.

The relationship between $\text{var}(\delta)$ and $k$ is:

$$
\frac{\partial \text{var}(\delta)}{\partial k} = \frac{2\sigma^2 Gz(G^4 - 2\mu^2 t G^2 - \mu^2 \sigma^2 t^2)}{[G^2 + \sigma^2 t]^2 G^4}.
$$

(1.25)

Basically, the sign of the derivative $\frac{\partial \text{var}(\delta)}{\partial k}$ is also ambiguous without any additional assumptions imposed. It can be either positive or negative or zero under different circumstances:

$$
\begin{cases}
G^2 > \mu^2 t + \mu^2 t\sqrt{1 + \sigma^2/\mu^2} & \iff \frac{\partial \text{var}(\delta)}{\partial k} > 0 \\
\mu^2 t < G^2 < \mu^2 t + \mu^2 t\sqrt{1 + \sigma^2/\mu^2} & \iff \frac{\partial \text{var}(\delta)}{\partial k} < 0 \\
\text{if } G^2 = \mu^2 t + \mu^2 t\sqrt{1 + \sigma^2/\mu^2} & \iff \frac{\partial \text{var}(\delta)}{\partial k} = 0
\end{cases}
$$

(1.26)

So, (1.24) and (1.26) imply that if $G^2 > \mu^2 t + \mu^2 t\sqrt{1 + \sigma^2/\mu^2}$, then the variance of the housing returns will be decreasing in $\mu$ but increasing in the per-unit-of-distance transportation cost $k$. If $\mu^2 t < G^2 < \mu^2 t + \mu^2 t\sqrt{1 + \sigma^2/\mu^2}$, the variance of the housing returns is decreasing in $k$, but the effect of $\mu$ on the variance is ambiguous. If $G^2 = \mu^2 t + \mu^2 t\sqrt{1 + \sigma^2/\mu^2}$, the variance of the housing returns will be decreasing in $\mu$, but unaffected by the composite numeraire good $k$.

Similarly, the derivative of $\text{var}(\delta)$ with respect to the discount rate $r$ is,

$$
\frac{\partial \text{var}(\delta)}{\partial r} = \frac{1}{2r^2} \frac{\partial \text{var}(\delta)}{\partial k}
$$

(1.27)

Similar to the sign of the derivative $\frac{\partial \text{var}(\delta)}{\partial k}$, the sign of $\frac{\partial \text{var}(\delta)}{\partial r}$ also is ambiguous without any additional assumptions imposed. It can be either positive or negative or zero under different circumstances:

$$
\begin{cases}
G^2 > \mu^2 t + \mu^2 t\sqrt{1 + \sigma^2/\mu^2} & \iff \frac{\partial \text{var}(\delta)}{\partial r} > 0 \\
\mu^2 t < G^2 < \mu^2 t + \mu^2 t\sqrt{1 + \sigma^2/\mu^2} & \iff \frac{\partial \text{var}(\delta)}{\partial r} < 0 \\
\text{if } G^2 = \mu^2 t + \mu^2 t\sqrt{1 + \sigma^2/\mu^2} & \iff \frac{\partial \text{var}(\delta)}{\partial r} = 0
\end{cases}
$$

(1.28)
(1.28), (1.24) and (1.26) indicate that if $G^2 > \mu^2 t + \mu^2 t \sqrt{1 + \sigma^2 / \mu^2}$, the variance of the house price appreciation rate will be decreasing in $\mu$, but increasing in the transportation cost $k$ and discount rate $r$. If $\mu^2 t < G^2 < \mu^2 t + \mu^2 t \sqrt{1 + \sigma^2 / \mu^2}$, the variance of the housing returns is decreasing in $k$ and $r$, but the effect of $\mu$ on the variance is ambiguous, it could be either positive or negative or even zero. If $G^2 = \mu^2 t + \mu^2 t \sqrt{1 + \sigma^2 / \mu^2}$, the variance of the housing returns will be decreasing in $\mu$, but unaffected by the composite numeraire good $k$ and discount rate $r$.

We do not conduct comparative static analysis with regard to $\bar{x}$, since $\bar{x}$ is fixed at some level by the set up of the model.

We have seen from the above derivation that the comparative static analysis suggests a clear positive relationship between variance of the housing returns and volatility of household income. Without imposing additional constraints on parameters, the partial derivatives with respect to other variables all imply ambiguous relationships. However, we can still generate interesting and testable hypotheses from the analysis if we take a closer look at the constraints we imposed. So in general, the comparative static analysis is informative and provides us with testable hypotheses. More specifically, based on the above analysis, the following hypotheses can be tested in our empirical investigation:

i) The variance of the housing returns is positively related to the variance of household income $\sigma^2$.

ii) The variance of the housing returns is related to the drift of household income $\mu$, the per-unit-of-distance transportation cost $k$, and discount rate $r$, but the directions of relationship are ambiguous unless some specific constraints on parameters are imposed.

iii) The discount rate $r$ and the per-unit-of-distance transportation cost $k$ affect the variance of the housing returns in the same direction; i.e., if variance of the housing returns
is positively related to discount rate $r$, then it is also positively related to the per-unit-of
distance transportation cost $k$, and vice versa.

iv) If variance of the housing returns is positively related to transportation cost $k$ (or $r$), then it should be negatively related to the drift of household income $\mu$. Otherwise, the
relationship between variance of the housing returns and $\mu$ is ambiguous.

1.2.5 Extensions of the Original Model
The Variance of Housing Returns Under Growth Control

As we have mentioned in the introduction, more studies have started to emphasize the effects of supply-side variables on house price dynamics and the elasticity of housing supply is closely related to restrictive zoning and other land use regulations imposed by the government. However, in our previous discussion, households and land owners are the only two players in the model. Therefore, we now introduce government into this model as the third player and study the variance of the house price appreciation rate under the existence of a government’s growth control.

We assume that the objective function of the government depends on the welfare of both landowners and households. The welfare of landowners is reflected in total land rents earned while the welfare of households is reflected in the utility achieved. In our model of fixed lot size, the amount of land consumed by each household is fixed at one unit. In the previous section, households derive utility from consuming land and a composite numeraire good $x$. Here we let the utility level also depends negatively on the population $N(t)$ of the urban area with the utility function

$$U(x, N) = x - \beta N(t), \quad (1.29)$$

This is based on model in Brueckner (1998).
where $\beta \geq 0$ is the externality parameter. Here, the consumption of land is suppressed in writing the utility function because the amount of consumption is always fixed at one unit per household.

The objective function of the government, therefore, is to maximize the weighted summation of utility level achieved by the household and the total rent earned by the landowners by choosing an appropriate $\pi(t)$:

$$\max_{\pi(t)} \lambda_0 U + (1 - \lambda_0) \int_0^{\pi(t)} [y(t) - x - kz] dz,$$

where $\lambda_0$ is a welfare weight which satisfies $0 \leq \lambda_0 \leq 1$. And recall that $N(t) = \pi \pi(t)^2$.

Solving this maximization problem using Leibniz’s rule, we get the optimal size of urban area $\pi(t)^*$, which satisfies:

$$\pi(t)^* = \frac{1 - \lambda_0}{k(1 - \lambda_0) + 2\beta \pi \lambda_0} [y(t) - x].$$

Let $\psi^{-1} = \frac{1 - \lambda_0}{k(1 - \lambda_0) + 2\beta \pi \lambda_0}$, we then can denote $\pi(t)^*$ as:

$$\pi(t)^* = \psi^{-1} [y(t) - x].$$

Here, $0 \leq \psi^{-1} \leq 1$ is used to reflect the extent of growth control impose by the government and a larger $\psi$ indicates a more restrictive growth control measure. Therefore, the government conducts growth control by maintaining the ratio of city size and household income at a constant level. We should also note that the extent of growth control $\psi$ depends on $\beta$, $\pi$, $k$ and $\lambda_0$, while the only parameter that the government can directly control is the weight parameter $\lambda_0$. We can see that $\partial \psi / \partial \lambda_0 > 0$, which implies that the more weight that the government put on the welfare of household, the more restrictive growth control will be. For the sake of simplicity, we let $\pi(t)^* = \psi^{-1} y(t)$ in the following derivation.
Following the reasoning of the derivation in Capozza and Helsley’s, the price of urban land \( P^u(t, z) \) when the future is uncertain, now equals

\[
P^u(t, z) = \frac{1}{r} (A + r C + \frac{r - \lambda k y^{-1} \mu}{\lambda r} - k z) + \frac{E}{2} \left[ \int_t^\infty k y^{-1} y(\tau)e^{-r(\tau - t)} d\tau | y(t) \right].
\]  

(1.33)

Let \( \tau = t + s \) and with some derivation, equation (1.33) becomes

\[
P^u(t, z) = \frac{1}{r} (A + r C + \frac{r - \lambda k y^{-1} \mu}{\lambda r} - k z) + \frac{k}{\psi r^2} \mu + \frac{k}{\psi r} y(t)
\]  

(1.34)

Recall that \( \lambda = \frac{\sqrt{\mu^2 + 2\sigma^2 \gamma - \mu}}{\sigma^2} \).

The appreciation rate turns out to be

\[
\delta = \frac{k y(t + 1) - k y(t)}{\frac{1}{\psi}(A + r C + \frac{1}{\lambda} - k z) + \frac{k}{\psi r} y(t)}
\]  

(1.35)

For notational simplicity, we let \( H = \psi (A + r C + \frac{1}{\lambda} - k z) + k \mu t \). Then,

\[
E \delta = \frac{k \mu}{H}, \quad E \delta^2 = \frac{k^2 (\mu^2 + \sigma^2)}{H^2 + k^2 \sigma^2 t}.
\]  

(1.36)

Therefore, the variance of the house price appreciation rate equals

\[
var(\delta) = \frac{k^2 (\mu^2 + \sigma^2)}{H^2 + k^2 \sigma^2 t} - \frac{k^2 \mu^2}{H^2}.
\]  

(1.37)

Regarding comparative static analysis, first, we check if variance of the housing returns is still increasing in the variance of income \( \sigma^2 \). By taking a partial derivative, we show

\[
\frac{\partial var(\delta)}{\partial \sigma^2} = \frac{k^2 H^4 (H^2 - k^2 \mu^2 t) + 2\sigma^2 k^2 H \psi \lambda^{-2} \frac{\partial \lambda}{\partial \sigma^2} (H^4 - 2k^2 \mu^2 t H^2 - k^4 \mu^2 \sigma^2 t^2)}{\left[H^2 + k^2 \sigma^2 t^2\right]^2 H^4}.
\]  

(1.38)
We know that $\partial \lambda / \partial \sigma^2 < 0$ and $H^2 > k^2 \mu^2 t$ given $\text{var}(\delta) > 0$. However, the sign of the partial derivative is ambiguous based on the current information we have. If we impose an additional constraint on $H$, such that

$$H^2 \leq k^2 \mu^2 t + k^2 \mu^2 t \sqrt{1 + \sigma^2 / \mu^2}$$

(1.39)

then $\partial \text{var}(\delta) / \partial \sigma^2 > 0$. That is, the variance of the housing returns is increasing in the variance of household income. Otherwise, without additional constraints being imposed on $H$, the direction of the effect is ambiguous.

Similarly, the partial derivative with respect to $\psi$, which is the extent of growth control, is

$$\frac{\partial \text{var}(\delta)}{\partial \psi} = \frac{-2k^4 \sigma^2 H(A + rC + \frac{1}{\chi} - kz)(H^4 - 2k^2 \mu^2 t H^2 - k^4 \mu^2 \sigma^2 t^2)}{[H^2 + k^2 \sigma^2 t]^2 H^4}.$$ 

(1.40)

Again, the sign of this partial derivative can be positive, negative, or even zero, depending on the parameters. More specifically,

$$\begin{cases} 
H^2 > k^2 \mu^2 t + k^2 \mu^2 t \sqrt{1 + \sigma^2 / \mu^2} \iff \partial \text{var}(\delta) / \partial \psi < 0 \\
k^2 \mu^2 t < H^2 < k^2 \mu^2 t + k^2 \mu^2 t \sqrt{1 + \sigma^2 / \mu^2} \iff \partial \text{var}(\delta) / \partial \psi > 0 \\
H^2 = k^2 \mu^2 t + k^2 \mu^2 t \sqrt{1 + \sigma^2 / \mu^2} \iff \partial \text{var}(\delta) / \partial \psi = 0
\end{cases}$$

(1.41)

The marginal effect of $A$ and $C$ on the variance of the housing returns, behave exactly like the growth control parameter $\psi$, that is

$$\begin{cases} 
H^2 > k^2 \mu^2 t + k^2 \mu^2 t \sqrt{1 + \sigma^2 / \mu^2} \iff \partial \text{var}(\delta) / \partial A < 0, \partial \text{var}(\delta) / \partial C < 0 \\
k^2 \mu^2 t < H^2 < k^2 \mu^2 t + k^2 \mu^2 t \sqrt{1 + \sigma^2 / \mu^2} \iff \partial \text{var}(\delta) / \partial A > 0, \partial \text{var}(\delta) / \partial C > 0 \\
H^2 = k^2 \mu^2 t + k^2 \mu^2 t \sqrt{1 + \sigma^2 / \mu^2} \iff \partial \text{var}(\delta) / \partial A = 0, \partial \text{var}(\delta) / \partial C = 0
\end{cases}$$

(1.42)

The sign of the partial derivative with respect to $k$, the per-unit-of-distance transportation cost, is ambiguous. So in the case of growth control, the theoretical model only implies that the variance of the housing returns is closely related to the transportation cost and we will find out the direction of their relationship through empirical investigation later.
Finally, the partial derivative with respect to the growth rate $\mu$ is,

$$\frac{\partial \text{var}(\delta)}{\partial \mu} = \frac{-2k^4\sigma^2\mu t H^4 - 2k^6\sigma^4\mu t^2 H^2}{[H^2 + k^2\sigma^2 t]^2 H^4}$$

$$-\frac{2k^2\sigma^2 H(-\psi^2 \lambda - 2\frac{\partial \lambda}{\partial \mu} + kt)(H^4 - 2k^2\mu^2 t H^2 - k^4\mu^2\sigma^2 t^2)}{[H^2 + k^2\sigma^2 t]^2 H^4}$$

(1.43)

$\partial \lambda/\partial \mu < 0$, therefore, if $H^2 \geq k^2\mu^2 t + k^2\mu^2 t\sqrt{1 + \sigma^2/\mu^2}$, then $\partial \text{var}(\delta)/\partial \mu < 0$.

Otherwise, the relationship is ambiguous. This result is quite similar to that of the previous case, where no growth control is present.

Based on the above comparative static analysis, the following hypothesis can be tested empirically:

i) With growth controls, the variance of the housing returns is related to the variance of household income $\sigma^2$, drift of household income $\mu$, the extent of growth control $\psi$, per-unit-of-distance transportation cost $k$, agricultural land rent $A$ and conversion cost $C$. However, the directions of those relationships are ambiguous unless some specific constraints on parameters are imposed.

ii) The growth control parameter $\psi$, conversion cost $C$ and agricultural land rent $A$ affect the variance of the housing returns in the same direction: i.e., if variance of the housing returns is positively related to the growth control parameter $\psi$, then it is also positively related to $A$ and $C$.

iii) If the variance of the housing returns is positively related to the extent of growth control $\psi$, the agricultural land rent $A$, and the conversion cost $C$, then it must also be positively related to the variance of household income $\sigma^2$.

iv) If the variance of the housing returns is negatively related to the extent of growth control $\psi$, the agricultural land rent $A$, and the conversion cost $C$, then it must also be negatively related to the drift of household income $\mu$. 

20
Land Leverage

There is still one variable which is worth taking into consideration. In previous section, we’ve shown that the percentage change in house price is closely related to the percentage change in urban land price via the equation

\[
\frac{\partial P^h(t, z)}{P^h(t, z)} = \alpha \frac{\partial P^u(t, z)}{P^u(t, z)},
\]

where \(\alpha\) represents the share of land in producing housing under the standard Cobb-Douglas production function of housing \(H^r = B N^\alpha K^{1-\alpha}\). The parameter \(\alpha\) here can have another interpretation. In particular, it also stands for "land leverage", which is analogous to finance leverage and defined as the land-to-total value ratio. Following section 1.2.4, the land-to-total value ratio is calculated as:

\[
\frac{N}{H} \cdot \frac{P^u(t, z)}{P^h(t, z)} = \frac{N}{H}, B \alpha \left(\frac{K}{N}\right)^{1-\alpha} = \alpha.
\]

Therefore, equation (1.44) not only tells us that the percentage change in house price is closely related to the percentage change in urban land price, it also implies that the house price dynamics are positively related to the extent of land leverage. To get a better understanding of the variance of the housing returns, we can further derive (1.44) by calculating the variance of both sides of the equation and get:

\[
\text{var} \left( \frac{\partial P^h(t, z)}{P^h(t, z)} \right) = \alpha^2 \text{var} \left( \frac{\partial P^u(t, z)}{P^u(t, z)} \right).
\]

From equation (1.46), we notice that it is \(\alpha^2\) that links the variance of the housing returns to the variance of the land price appreciation rate. This amplifies the effect of land leverage on explaining the cross-sectional differences of housing returns volatility. We can see this through a very simple numeric example. Suppose we have two metropolitan areas with almost the same economic fundamentals (income dynamics, population level,
etc.) so that they have very similar variances of land price appreciation. However, the land leverage of one area (say, area A) is four times larger than the land leverage of the other area (say, area B), that is \( \alpha_A = 4\alpha_B \). Therefore, with land leverage \( \alpha \) being squared in (1.46), the variance of the housing returns in area A will be 16 times larger than the variance of the housing returns of area B, although they have a very close volatility of the land price appreciation rate.

Therefore, we can claim that metropolitan areas with a higher extent of land leverage usually will have higher levels of housing returns volatility. This conjecture adds to the hypotheses we test empirically.

**Variance of the Housing Returns when the Population Also Evolves Stochastically**

Instead of letting population vary according to exogenous household income, we now consider the case when population also evolves stochastically as household income does, such that

\[
y(t) = \mu t + \sigma W(t)
\]

\[N(t) = N(0)e^{\delta t + \sigma_n W(t)}
\]

where \( N(t) \) is the population level of the urban area at time \( t \). (1.48) means that an urban area’s population grows exponentially at a stochastic Brownian motion growth rate with drift \( \delta \) and variance parameter \( \sigma_n \). In our monocentric fixed lot size model, (1.48) also implies that:

\[
\bar{z}(t) = \bar{z}(0)e^{\delta t/2 + \sigma_n W(t)/2}
\]
In this special case, instead of varying according to \(y(t)\), the boundary of urban area \(z(t)\) now grows at some stochastic rate. Therefore, the utility level achieved by each household is no longer fixed, which means that \(R(t, z)\) no longer follows a Brownian motion process, but instead follows some unknown stochastic process.

Assume that for this special case, the reservation rent level \(R^*\), which is the level of urban land rent when it is optimal to convert, equals \(A + rC + \eta(\mu, \sigma^2, r, g, \sigma_n^2)\). Here, \(\eta\) is the irreversibility premium and it is a function of \(\mu, \sigma^2, r, g\) and \(\sigma_n^2\).

By a derivation similar to the previous case, we get that

\[
P^\mu(t, z) = 1/r(A + rC + \eta - kz) + \mathbb{E} \left[ \int_t^\infty k\bar{z}(\tau)e^{-r(\tau-t)}d\tau \mid \bar{z}(t) \right]
\]

\[
= 1/r(A + rC + \eta - kz) + k\bar{z}(t)\mathbb{E} \left[ \int_t^\infty e^{g(t-\tau)/2}e^{\sigma_n[W(\tau)-W(t)]/2}e^{-r(\tau-t)}d\tau \mid \bar{z}(t) \right]
\]

\[
= 1/r(A + rC + \eta - kz) + k\bar{z}(t)\mathbb{E} \left[ \int_0^\infty e^{gs/2}e^{\sigma_n[W(t+s)-W(t)]/2}e^{-rs}ds \mid \bar{z}(t) \right]
\]

\[\text{(1.50)}\]

According to the moment generating function of the standard normal distribution, (1.50) can be simplified to:

\[
P^\mu(t, z) = 1/r(A + rC + \eta - kz) + k\bar{z}(t)\int_0^\infty e^{gs/2}e^{\sigma_n^2/8}e^{-rs}ds
\]

\[
= 1/r(A + rC + \eta - kz) + \frac{8}{8r - 4g - \sigma_n^2}k\bar{z}(0)e^{g/2 + \sigma_n W(t)/2}
\]

\[\text{(1.51)}\]

Therefore, the house price appreciation rate is:

\[
\delta(t, t+1, z) = \frac{8}{8r - 4g - \sigma_n^2}k\bar{z}(0)\left[e^{g(t+1)+\sigma_n W(t+1)/2} - e^{g(t)+\sigma_n W(t)/2}\right] - 1/r(A + rC + \eta - kz) + \frac{8}{8r - 4g - \sigma_n^2}k\bar{z}(0)e^{g/2 + \sigma_n W(t)/2}/2
\]

\[\text{(1.52)}\]

\[7\text{Recall that in previous case when only }y(t)\text{ follows a stochastic process, we can get explicit expression of the irreversibility premium, which is }\frac{r^2 \mu}{\sigma^2}.\]
Given the complexity of the variance of the appreciation rate, comparative static analysis is difficult. The only conclusion we can draw is that the variance of the housing returns when we have both stochastic income and stochastic population growth will be related to drift of income growth $\mu$, variance of income growth $\sigma^2$, transportation cost $k$, discount rate $r$, drift of population growth rate $g$, variance of population growth rate $\sigma_n^2$, agricultural land rent $A$ and conversion cost $C$.

The utility achieved by each household in this case is $U(x, 1)$, where $x$ here is no longer fixed, but instead:

$$x = y(t) - R(t, z) - kz$$
$$= \mu t + \sigma W(t) - A - rC - \eta - k\bar{z}(0)e^{gt/2+\sigma_n^2 W(t)/2}.$$  \hspace{1cm} (1.53)

Therefore, household’s utility now varies stochastically.

1.3 Testable Hypotheses Derived from the Models

According to the comparative static analysis conducted, following hypotheses are empirically testable:

**Under the assumption of an open city or free migration** The following hypotheses can be tested:

1) The variance of the housing returns is positively related to the variance of household income $\sigma^2$.

2) The variance of the housing returns is related to the drift of household income $\mu$, the per-unit-of-distance transportation cost $k$, and discount rate $r$, but the directions of the relationships are ambiguous unless some specific constraints on parameters are imposed.
3) The discount rate $r$ and the per-unit-of-distance transportation cost $k$ affect the variance of the housing returns in the same direction; i.e., if the variance of the housing returns is positively related to the discount rate $r$, then it is also positively related to the per-unit-of-distance transportation cost $k$, and vice versa.

4) If the variance of the housing returns is positively related to the transportation cost $k$ (or $r$), then it should be negatively related to the drift of household income $\mu$. Otherwise, the relationship between variance of the housing returns and $\mu$ is ambiguous.

5) The variance of the housing returns is positively related to the land leverage $a$.

**Under the assumption of existence of growth controls**  The following hypotheses can be tested:

1) With growth control, the variance of the housing returns is related to the variance of household income $\sigma^2$, the drift of household income $\mu$, the extent of growth control $\psi$, the per-unit-of-distance transportation cost $k$, the agricultural land rent $A$ and the conversion cost $C$. However, the directions of these relationships are ambiguous unless some specific constraints on parameters are imposed.

2) The growth control parameter $\psi$, the conversion cost $C$ and the agricultural land rent $A$ affect the variance of the housing returns in the same direction: i.e., if variance of the housing returns is positively related to the growth control parameter $\psi$, then it is also positively related to $A$ and $C$.

3) If the variance of the housing returns is positively related to the extent of growth controls $\psi$, (or the agricultural land rent $A$, the conversion cost $C$) then it must also be positively related to the variance of household income $\sigma^2$. 
4) If the variance of the housing returns is negatively related to the extent of growth control \( \psi \), (or the agricultural land rent \( A \), the conversion cost \( C \)) then it must also be negatively related to the drift of household income \( \mu \).

5) The variance of the housing returns is positively related to land leverage \( a \).

**Under the assumption of stochastic population growth**  
Testable hypotheses are:

1) The variance of the housing returns when we have both stochastic income and stochastic population growth will be related to the drift of income growth \( \mu \), the variance of income growth \( \sigma^2 \), the per-unit-of-distance transportation cost \( k \), the discount rate \( r \), the drift of population growth rate \( g \), the variance of population growth rate \( \sigma_n^2 \), the agricultural land rent \( A \) and the conversion cost \( C \).

2) The variance of the housing returns is positively related to land leverage \( a \)

**1.4 Empirical Investigation**

In this section, we use MSA level data to test the hypotheses derived from our theoretical model. In section 1.3, we derived different sets of testable hypotheses under different assumptions, so we will conduct empirical investigations of those hypotheses separately. That is, we will employ different regression models under different circumstances. Furthermore, considering the feasibility of conducting empirical investigation under the case of stochastic population growth\(^8\), we focus our empirical work on testing the hypotheses under the cases of open city and growth control.

\(^8\)For the case of stochastic population growth, we only have one single specific hypothesis, which makes the empirical work not as meaningful.
1.4.1 Cross-Sectional Regression Model

We use cross-sectional linear regression models to test the hypotheses implied by the theoretical model. Regressors differ according to the different assumptions.

**Under the Assumption of Open City or Free Migration**

According to the hypotheses listed in section 1.3, the linear regression model under the assumption of free migration would be:

\[
\text{var}(\delta)_i = \alpha_0 + \alpha_1 \mu_i + \alpha_2 \sigma_i^2 + \alpha_3 t \mu_i + \alpha_4 \text{leverage}_i + \alpha_5 \delta_i + \epsilon_i. \tag{1.54}
\]

Here \(\alpha_0\) is a constant, \(\text{var}(\delta)_i\) represents the variance of the housing returns for MSA \(i (i = 1, 2, ..., n)\). \(\mu_i\) is the drift or growth rate of per capita income for each MSA, \(\sigma_i^2\) is the variance of per capita income at MSA level, \(t \mu_i\) represents the transportation cost of MSA \(i\), which is approximated by the inverse of average commuting speed of each MSA, \(\text{leverage}_i\) denotes the share of land in the total housing value for each MSA, \(\delta_i\) is the MSA level average housing returns and finally, \(\epsilon_i\) is the standard Gaussian error.

We can see that the average housing returns of each MSA \(\delta_i\) is also involved as a regressor. It is common knowledge that in financial market there is a positive relation between risk and return of investment assets. In other words, stocks with higher returns usually are more volatile. Such a positive relation was shown to also exist in the owner-occupied housing market through a theoretical model by Berkovec(1989). The empirical work conducted by Cannon et.al.(2006) also demonstrated a significantly positive relation between risk and return in the U.S. housing market. That is, riskier and more volatile housing markets usually provide higher average housing returns. Therefore, here we would like to include \(\delta\) as another regressor in the empirical model so that we can also check the existence of this relationship with MSA level data. Furthermore, the discount rate \(r\) is
not included in the regression, given the fact that discount rates do not differ much across MSAs; therefore it does not contribute much to the spatial difference in housing return volatility.

**Under the Assumption of Growth Control**

We have known above that when growth control exists (the city is no longer completely open to free migration), a different set of testable hypotheses come from the theoretical model. Accordingly, our regression model changes to:

\[
\text{var}(\delta)_i = \beta_0 + \beta_1 \mu_i + \beta_2 \sigma_i^2 + \beta_3 t_i + \beta_4 \text{reg}_i + \beta_5 \text{aggr}_i + \beta_6 \text{leverage}_i + \beta_7 \delta_i + \eta_i. \tag{1.55}
\]

\(\beta_0\) here is a constant, \(\text{reg}_i\) measures the regulation or growth control level of each city, which corresponds to variable \(\psi\) in the theoretical model. \(\text{aggr}_i\) represents the agricultural land rent level and \(\eta_i\) is the standard Gaussian error. Conversion cost \(C\) is not included as a regressor in the model, due to unavailability of the data. We have shown that conversion cost refers to the capital cost of converting a unit of agricultural land to urban land. Data related to this type of cost is difficult to obtain at the MSA level. It is reasonable to assume that conversion costs do not vary greatly cross-sectionally, therefore we omit it in the regression.

**1.4.2 Data**

The data sets used in our empirical investigation come from various sources.

The variance of the house price appreciation rate and the average housing returns are calculated using the CMHPI (Conventional Mortgage House Price Index) data provided by Freddie Mac. The CMHPI is a weighted, repeat-sales index, meaning that it measures
average price changes in repeat sales or refinancings on the same properties. This information is obtained by reviewing repeat mortgage transactions on single-family properties whose mortgages have been purchased or securitized by Fannie Mae or Freddie Mac since January 1975. The CMHPI is updated each quarter as additional mortgages are purchased or securitized by Fannie Mae and Freddie Mac. The new mortgage acquisitions are used to identify repeat transactions for the most recent quarter and for each quarter since the first quarter of 1975. The CMHPI data provides indexes for all nine Census Divisions, the 50 states and the District of Columbia, and every MSA in the U.S. Based on this HPI, we can compute MSA specific annual house price appreciation rate on a quarterly basis from the first quarter of 1975 to the second quarter of 2006. Given the time series of housing returns of each MSA, we can then calculate the average housing returns of each city and also the variance of the housing returns accordingly. Furthermore, the original CMHPI is not adjusted for inflation, so we used the CPI obtained from Federal Reserve Bank of St. Louis to adjust the house price index for inflation in order to get the mean and variance of real housing returns.

The drift and variance of household income are calculated using MSA level per capita personal income data provided by the Bureau of Economic Analysis (BEA). This is a yearly data set dated from 1969 to 2005. The CPI also is used here to adjust per capita income for inflation so that we obtain the drift and volatility of a household’s real income.

The regulation index, which measures the extent of regulation or growth control of each MSA, comes from Malpezzi’s 1996 paper "Housing Price, Externalities, and Regulation in U.S. Metropolitan Areas". In this paper, Malpezzi constructed a city-specific regulatory index by adding the unweighted values of seven variables collected by the Warton research project, documented by Linneman et. al. (1990) and Buist (1991). The higher the score is,
the more regulated the city is. More specifically, 56 major MSAs of the U.S. are covered in this index with San Francisco scored as the highest (29) and Chicago as the lowest (13).

The per-unit-of-distance transportation cost, known as \( k \) in the theoretical model and \( tc_i \) in our regression model, is approximated by the inverse of average daily commuting speed (miles per minutes) of each MSA. This MSA level travel speed data set is obtained from the National Household Travel Survey (NHTS) 2001, which provides authoritative data on travel by all modes of transportation, for all travel purposes, and all travel distances.

Agricultural land rent data is approximated by the average returns to agricultural land (dollars per acre). The data can be calculated using the Census of Agriculture data from the United States Department of Agriculture (USDA), and is computed as \( \frac{\text{value of all agricultural products sold} + \text{government payment} - \text{total farm production expenses}}{\text{total farmland area}} \). The data is at the state level, not MSA specific, so we used state level data to approximate MSA level data in this case.

Finally, the land leverage data, which measures land’s share of home value, is obtained from Davis and Palumbo’s 2006 paper "The Price of Residential Land in Large U.S. Cities". Unfortunately, their leverage data only covers about 50 major cities of the U.S., which means that the sample size of our empirical regression, due to the constraints on leverage data availability, might be limited to around 50. A feasible solution to this data issue is to enlarge the sample size using techniques of data imputation, the details of which will be elaborated later in this paper.
1.4.3 Small Sample Regression

Due to constraints on the availability of land leverage data, as we mentioned above, we first conduct an empirical investigation using cross-sectional data on 52 major MSAs. Later, we expand the sample size and do the empirical investigation using a larger sample.

Regression Results

Table (1.2) displays the results of the small sample linear regression. The second column of the table shows the empirical regression results under the case of free migration or an open city (no growth control), the third column gives the regression results under the case of growth control. Because the value of variance of per capita income $\sigma^2$ is very large and will lead to really small coefficient estimator, we used the standard deviation $\sigma$ instead, in our regression.

First, let’s check the conjectures derived from our theoretical model under the assumption of an open city, by looking at the second column of table (1.2). We can clearly see that the first hypothesis is consistent with the empirical result of the regression. That is, the variance of the housing returns is positively and significantly related to the variance of household income. The second hypothesis is also verified in the empirical investigation, because all the variables are statistically significant. Furthermore, we can see that the variance of the housing returns is significantly and negatively related to the drift of household income $\mu$, and it is significantly and positively related to the variable $tc$, which is corresponding to the per-unit-of-distance transportation cost $k$, therefore, hypothesis 4 is consistent with the empirical outcome. The last hypothesis is also verified. The variance of housing returns is positively and significantly related to leverage, the share of land in
total housing value. We also noticed that $\delta$, the average housing returns, turns out to be significantly and positively related to the variance of housing returns, therefore, the positive relation between risk and return is shown to also exist in the housing market by our empirical investigation. Hypothesis 3 can not be tested since we did not include the discount rate $r$ in the regression due to the reason we mentioned above. So generally speaking, under the case of open city or free migration, hypotheses derived from our theoretical model are very consistent with the empirical outcomes obtained from the linear regression.

We now look at the last column of table (1.2) and test the hypotheses derived from the theoretical model under the assumption of the existence of growth controls. The regression results indicate that almost all regressors are statistically significant, except for agricultural land rent. Therefore, the first hypothesis is consistent with our empirical outcome. Hypothesis 3 also is verified because the regression tells us that the variance of the housing returns is positively and significantly related to $reg$, the regulation or growth control level of each MSA, which corresponds to variable $\psi$ in the theoretical model; at the same time, it is also positively and significantly related to the variance of household income. Hypothesis 5 is also consistent with the empirical results. The variance of the housing returns is positively and significantly related to land leverage. Additionally, the empirics support the positive relationship between risk and return again-the t-statistics for $\delta$ is positive and large. The second hypothesis can not be tested given the fact that agricultural land rent is not significant in the regression, and hypothesis 4 is also irrelevant since we have found that the variance of the housing returns is positively, not negatively related to $\psi$, which corresponds to the variable $reg$ in our regression.

In summary, for both cases, almost all the hypotheses derived from our theoretical model are consistent with the empirical outcomes from the small sample size cross-sectional
regression. The variance of the housing returns is shown to be positively related to the variance of household income, land leverage, the average housing returns, the level of growth controls and the per-unit-of-distance transportation cost. Drift of household income is shown to be negatively related to the variance of the housing returns.

1.4.4 Larger Sample Regression

We have seen that the small sample regression has empirically verified almost all the hypotheses implied by our theoretical model under both cases. However, it is a regression with only 52 observations, that is, only 52 major MSAs of the U.S. are involved in the regression. So, will the cross-sectional regression produce similar outcomes if more MSAs are involved? How robust is our empirical investigation? To check this, we need to expand the sample size and run the regression with more observations. As we mentioned above, availability of land leverage data is the major constraint on the sample size of the regression, therefore, the first thing is to obtain more observations of land leverage data and we accomplish this by using the techniques of missing data imputation.

Leverage Data Imputation

Generally speaking, the basic idea of data imputation is to use the data we already have to predict the missing observations. In practice, we can have missing observations on dependent variables and/or explanatory variables. The specific case we have for land leverage data has a missing structure as

\[
\begin{bmatrix}
    \text{leverage}_1 \\
    \text{leverage}_{\text{miss}}
\end{bmatrix} = \begin{bmatrix}
    X_1 \\
    X_2
\end{bmatrix} \beta + \begin{bmatrix}
    u_1 \\
    u_2
\end{bmatrix},
\]

(1.56)

where \(\text{leverage}_1\) represents the part of leverage data we already have, which is limited to around 50 cities, \(\text{leverage}_{\text{miss}}\) represents the part of unobserved land leverage data,
which belongs to the remaining MSAs. $X_1$ and $X_2$ are both observed explanatory variables for leverage data, and $u_1, u_2$ are Gaussian errors. Furthermore, $\text{leverage}_1$ is $N_1$ by 1, $\text{leverage}_{\text{miss}}$ is $N_2$ by 1, $X_1$ is $N_1$ by $k$, $X_2$ is $N_2$ by $k$, $\beta$ is $k$ by 1, $u_1$ is $N_1$ by 1, and $u_2$ is $N_2$ by 1.

Obviously, from (1.56), to predict the missing leverage data, we first need to find out the components of $X$, that is, we need to know the variables that can explain leverage. To do this, the housing production function may help.

**Explanatory Variables for Land Leverage Implied by the Model**

First, we return to function (1.11), which is a standard Cobb-Douglas housing production function with constant returns to scale. In section 1.2.5, we showed that $\alpha$ is the share of land in producing housing and it can also be interpreted as "land leverage", which is analogous to financial leverage and defined as the share of land in total housing value.

Considering a more general case, that is, instead of being specifically Cobb-Douglas, let the production function take a general CES (constant elasticity of substitution) form:

$$H = B[\theta(a_K K)^\gamma + (1 - \theta)(a_N N)^\gamma]^{1/\gamma}$$

(1.57)

where $0 < \theta < 1$ is the share parameter and $\gamma$ determines the degree of substitutability of the inputs and $\gamma \leq 1$. The parameters $B, a_K, a_N$ depend upon the units in which the output and inputs are measured and play no important role.

Therefore, from (1.57), we can see that the marginal product of capital $K$ and marginal product of land $N$ are:

$$MP_K = B \left[ \theta(a_K K)^\gamma + (1 - \theta)(a_N N)^\gamma \right]^{1/\gamma - 1} \theta(a_K K)^{\gamma - 1} a_K$$

(1.58)

$$MP_N = B \left[ \theta(a_K K)^\gamma + (1 - \theta)(a_N N)^\gamma \right]^{1/\gamma - 1} (1 - \theta)(a_N N)^{\gamma - 1} a_N.$$  

(1.59)
We assume that the input market is competitive and factors are paid in compliance with their value of marginal product. Let $P_N$ denote the price of land, $P_K$ represent the price of capital and $P_H$ stands for the price of housing. We then have following equations:

$$P_K = P_H \cdot MP_K$$  \hfill (1.60)

$$P_N = P_H \cdot MP_N$$  \hfill (1.61)

The land leverage, which is defined as the land-to-total value, therefore is equal to

$$\text{Land Leverage} = \frac{P_N \cdot N}{P_H \cdot H} = \frac{MP_N \cdot N}{H}. \hfill (1.62)$$

Substituting out $H$ using equation (1.57) and $MP_N$ using equation (1.59), the land leverage is

$$\text{Land Leverage} = \frac{(1 - \theta)(a_N N)^\gamma}{\theta(a_K K)^\gamma + (1 - \theta)(a_N N)^\gamma}. \hfill (1.63)$$

From (1.63), we can see that when $\gamma = 0$, which is the case of the Cobb-Douglas production function, the land leverage is equal to $(1 - \theta)$, which is exactly the share of land in producing housing.

Furthermore, using equation (1.61), we can express land leverage in an alternative way such that

$$\text{Land Leverage} = \left(\frac{P_H}{P_N}\right)^{\frac{\gamma}{1 - \gamma}} (Ba_N)^{\frac{\gamma}{1 - \gamma}} (1 - \theta)^{\frac{1}{1 - \gamma}}. \hfill (1.64)$$

As mentioned above, the parameters $B$, $a_K$, and $a_N$ are just conversion constants and play no important roles, thus land leverage is basically related to the price of housing $P_H$, the price of land $P_N$, the degree of substitutability $\gamma$ and the share of land in producing housing $(1 - \theta)$. More specifically, the land leverage is positive related to the share of land in producing housing, but the direction of the relationship with $P_H$ or $P_N$ depends on the value of $\gamma$, and the direction of the effect of $\gamma$ on land leverage also is ambiguous.
Therefore, possible candidates for explanatory variables can be the price of housing $P_H$, the price of land $P_N$, land’s share in producing housing $(1 - \theta)$, and some variable which can approximate the degree of substitutability $\gamma$. Empirically, we will use MSA level median house price to represent $P_H$; data for price of land $P_N$ is difficult to obtain, so we will instead regard the land price as a function of economic fundamentals such as income, population and agricultural land rent. $\gamma$ is the degree of substitutability, and more specifically, for the CES production function with constant return to scale as shown in equation (1.57), the elasticity of substitution is equal to $\frac{1}{1 - \gamma}$, and recall that $\gamma \leq 1$. Therefore, when $\gamma$ goes to negative infinity, the elasticity of substitution goes to zero, which means no substitution between land and capital is possible. On the other hand, when $\gamma$ approximates 1, then the elasticity of substitution approaches infinity, which implies perfect substitution. In cities with more restrictive regulations on land development (for example, San Francisco), it is more difficult to substitute between land and capital, which means that the degree of substitution is low. Thus it is quite sensible to use regulation or the growth control level of each MSA to approximate $\gamma$ empirically.

In a preliminary regression, with regressors being median house price, square of median house price, population, income, agricultural land rent and regulation level, the outcome indicated that only median house price, square of median house price and agricultural land rent are significant, all others are not. Compared with the regression which only includes those significant explanatory variables, the complete one shows no improvement on the goodness of fit of the model. So to be more efficient, we choose only median house price, square of median house price and agricultural and rent as components of $X$ in model (1.56).
Naive Imputation Method  The naive imputation method is simple and involves only 2 steps. We first run OLS using the first $N_1$\(^9\) observations and find $\hat{\beta} = [X'_1 X_1]^{-1} X'_1 leverage_1$. Table (1.3) in the appendix shows the results of the regression, and we can see that the land leverage is positively and significantly related to median house price and agricultural land rent, but negatively and significantly related to square of median house price. The goodness of fit of the regression is high with $R^2$ equaling 0.76. In the second step, conditional on $X_2$, the predicted values of $leverage_{miss}$, denoted by $leverage_{miss}$, are given by $X_2\hat{\beta}$. After imputation, including the already known and augmented observations, we will finally get 271 observations of land leverage data, therefore, the sample size of our empirical investigation can be a much larger 271 compared to the previous 52.

MCMC Imputation Method  MCMC (Markov Chain Monte Carlo) is a widely used method for missing data imputation. It is used to generate pseudo-random draws from multinomial and otherwise intractable probability distributions via Markov chains. In MCMC simulation, one constructs a Markov chain long enough for the distribution of the elements to stabilize to a stationary distribution, which is the distribution of interest. This is called a single imputation. Usually, a single imputation can not adequately handle the missing-data uncertainty. This is the essential rationale for using a multiple imputation procedure, which is to repeatingly simulate steps of the chain and generate draws from the distribution of interest.

The Markov Chain calculations are implemented using the SAS Proc MI algorithm, with number of imputations equaling 20 and length of the Markov Chain being 100. After

\(^9\) $N_1$ equals 40 here, $N_2$ equals 231, and median house price data comes from Census 2000 by U.S. Census Bureau.
MCMC imputation, we again get 271 observations of land leverage data, which enlarges the sample size of our empirical regression of housing returns volatility.

**Regulation Data**

The regulation index, which measures the approximate level of regulation or growth control of each MSA, comes from Malpezzi’s 1996 paper; 56 major cities of the U.S. are covered in this index. Since our larger sample regression will cover 271 MSAs of U.S., we need regulation index data for the remaining 215 MSAs. The data is constructed based on the 56 indices and more specifically, the already-known 56 indices are used to construct state level regulation indices, and the rest of the 215 MSAs will be assigned indices according to the states they belong to. For example, Denver is among the already-known 56 MSAs and has a regulation score 17, therefore, Colorado Springs, which also belongs to the state of Colorado, but is among the unknown 215 MSAs, will accordingly be assigned a regulation score equaling 17.

**Regression Results**

After imputing land leverage data, we now have a larger sample size of 271, regressions are run using the augmented data set and outcomes are given in table (1.4) and table (1.5) respectively, according to the different imputation methods.

**Results with Naively Imputed Leverage Data** Table (1.4) provides us the results of the larger sample regression of housing returns variance, using naively imputed land leverage data. Compared with table (1.2), which is the small sample (observations=52) regression of housing returns variance, results are quite similar.
Under the assumption of an open city, the variance of the housing returns is positively and significantly related to the volatility of household income, land leverage, transportation cost, and the average housing returns; on the other hand, it is negatively and significantly related to the drift of household’s income. So the regression outcomes are very much similar to those of the small sample.

Under the assumption of growth control, the regression outcomes are also consistent. The variance of the housing returns is again shown to be positively and significantly related to the volatility of household income, land leverage, transportation cost, the average housing returns, and the regulation level of the city; the drift of household income is negatively and significantly related to the variance of house price appreciation rate. Agricultural land rent is not statistically significant, as it is in the small sample regression.

In summary, the large sample regression with naively imputed land leverage data gives similar outcomes to the small sample regression. Therefore, the hypotheses implied by our theoretical model, under both cases, are also verified by empirical investigation with the larger sample size, which proves the robustness of our empirical work.

Results with MCMC Imputed Leverage Data  We now turn to the regression with MCMC imputed land leverage data. Table (1.5) shows the regression results. The empirical outcomes of table (1.5) are highly consistent with those of table (1.4).

Under the case of an open city or free migration, the variance of the housing returns is positively and significantly related to the volatility of household income, land leverage, transportation cost, and average housing returns; on the other hand, it is negatively and significantly related to the drift of household’s income.
Under the assumption of growth control, the variance of the housing returns is positively and significantly related to the volatility of household income, land leverage, transportation cost, average housing returns, and regulation level of the city; drift or growth rate of household income is negatively and significantly related to the variance of the house price appreciation rate. Agricultural land rent again is not statistically significant.

Given the high similarity between the regression outcomes of naively imputed data and MCMC imputed data, the large sample regression again provides very similar outcomes to the small sample regression. Our empirical investigation gives us consistent outcomes independent of it being a small sample size regression or a larger sample size one, no matter whether the augmented data is naively imputed or MCMC imputed. Therefore, the empirical verification of the hypotheses derived from our theoretical model is ensured to be robust.

1.5 Conclusion

This paper derives a theoretical model to explain the geographic pattern of significant differences in the average variance of house price appreciation rates across different regions of the United States. The theoretical model of this paper is founded on the monocentric urban area model developed by Capozza and Helsley (1988 & 1990). Based on extensions to this model, the relationships between the variance of the house price appreciation rate and exogenous variables related to the urban economy are identified. The nature of these relationships is discussed and testable hypotheses implied by the theoretical models are derived. Results of empirical investigation are consistent with the model’s prediction. Specifically, the variance of housing returns is found to be significantly and positively related
to the variance of household income, the land regulation level (or growth control level) of each MSA, land leverage, and transportation costs.

Returning to the question we raised in the introduction part, which motivates the whole paper, that is, why does there exist such a vivid geographic pattern of the variance of the housing returns as shown in Figure 1? In other words, why are the housing returns much more volatile in coastal areas of the U.S than in the inland areas?

As we can see from tables (1.4) and (1.5), our large sample regression empirically explained about 60% of the spatial pattern of the variance of the housing returns. More specifically, as we have found both theoretically and empirically that variance of the housing returns is positively related to the variance of household income, therefore, the more volatile housing returns which coastal areas are facing could partially be explained by the more volatile household income that coastal areas are confronted with. Furthermore, we also showed that variance of the housing returns is positively related to land leverage, transportation cost and regulation level of a city, therefore, the usually higher land leverage level, more expensive transportation cost and higher extent of regulation or growth control level also contribute to the higher volatility of housing returns in coastal area MSAs. Additionally, in this paper, we empirically reconfmed the existence of positive relationship between risk and return in the housing market, therefore, we can also claim that the relatively higher volatility that coastal areas are facing is partially compensated by the higher average housing returns they are enjoy, which is indeed accurate looking at figure 1.

Research on this topic can be further extended by looking deeper into the special case when the population also evolves stochastically like income does. Although this paper deals with this issue to a certain extent, the topic will still benefit from more extensive research. The empirical investigation can also be carried out more thoroughly in order
to gain deeper insight. For example, the imputation methods might still have room to be improved; or even better, if we could obtain the complete original data set on land leverage and regulation of all the 379 MSAs, we can then run the entire empirical regressions again and test our theoretical hypotheses using more complete data.
1.6 Tables and Figures

1.6.1 Tables

Table 1.1: Descriptive Statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Obs.</th>
<th>Mean</th>
<th>Median</th>
<th>Std.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var of Housing Return</td>
<td>271</td>
<td>0.003</td>
<td>0.002</td>
<td>0.002</td>
<td>0.000</td>
<td>0.016</td>
</tr>
<tr>
<td>Average Housing Return $\delta$</td>
<td>271</td>
<td>0.015</td>
<td>0.012</td>
<td>0.014</td>
<td>-0.022</td>
<td>0.056</td>
</tr>
<tr>
<td>Income Drift $\mu$</td>
<td>271</td>
<td>174.2</td>
<td>168.4</td>
<td>57.0</td>
<td>14.8</td>
<td>375.1</td>
</tr>
<tr>
<td>Income Volatility $\sigma$</td>
<td>271</td>
<td>1971.3</td>
<td>1911.7</td>
<td>579.1</td>
<td>660.2</td>
<td>4159.4</td>
</tr>
<tr>
<td>Land Leverage(naive)</td>
<td>271</td>
<td>0.379</td>
<td>0.356</td>
<td>0.128</td>
<td>0.128</td>
<td>0.885</td>
</tr>
<tr>
<td>Land Leverage(MCMC)</td>
<td>271</td>
<td>0.389</td>
<td>0.382</td>
<td>0.154</td>
<td>0.081</td>
<td>0.885</td>
</tr>
<tr>
<td>Trans.Cost $tc$</td>
<td>271</td>
<td>1.926</td>
<td>1.923</td>
<td>0.138</td>
<td>1.661</td>
<td>2.759</td>
</tr>
<tr>
<td>Regulation Level</td>
<td>271</td>
<td>19</td>
<td>18</td>
<td>2.954</td>
<td>13</td>
<td>29</td>
</tr>
<tr>
<td>Aggr.Land Rent</td>
<td>271</td>
<td>69.61</td>
<td>56.28</td>
<td>55.99</td>
<td>2.15</td>
<td>288.52</td>
</tr>
</tbody>
</table>

Transportation cost is calculated by taking inverse of average speed of commuting (average miles per minutes). Agricultural land rent is approximated by the average return to agri.land (dollars per acre).
Table 1.2: Cross-Sectional Regression With Sample Size=52

<table>
<thead>
<tr>
<th>Variables</th>
<th>Open City</th>
<th>Growth Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>-0.00743</td>
<td>-0.00791</td>
</tr>
<tr>
<td></td>
<td>(-2.81)**</td>
<td>(-2.87)***</td>
</tr>
<tr>
<td>( \mu )</td>
<td>-0.00012</td>
<td>-0.00012</td>
</tr>
<tr>
<td></td>
<td>(-2.58)**</td>
<td>(-2.48)**</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
<tr>
<td></td>
<td>(2.39)**</td>
<td>(2.26)**</td>
</tr>
<tr>
<td>leverage</td>
<td>0.00465</td>
<td>0.00414</td>
</tr>
<tr>
<td></td>
<td>(2.05)**</td>
<td>(1.75)*</td>
</tr>
<tr>
<td>( \delta )</td>
<td>0.07801</td>
<td>0.08072</td>
</tr>
<tr>
<td></td>
<td>(3.11)**</td>
<td>(3.01)***</td>
</tr>
<tr>
<td>tc</td>
<td>0.00443</td>
<td>0.00384</td>
</tr>
<tr>
<td></td>
<td>(3.48)***</td>
<td>(2.92)***</td>
</tr>
<tr>
<td>reg</td>
<td>0.00011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.73)*</td>
<td></td>
</tr>
<tr>
<td>aggr</td>
<td>-0.00000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.91)</td>
<td></td>
</tr>
<tr>
<td>observations</td>
<td>52</td>
<td>52</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.775</td>
<td>0.788</td>
</tr>
</tbody>
</table>

dependent variable: variance of housing return

t statistics in parentheses

* significant at 10%, ** significant at 5%

*** significant at 1%
Table 1.3: Regression of Land Leverage

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>-0.09951 (-1.13)</td>
</tr>
<tr>
<td>median house price</td>
<td>0.00521 (5.01)**</td>
</tr>
<tr>
<td>median house price squared</td>
<td>-0.000008 (-3.17)***</td>
</tr>
<tr>
<td>agricultural land rent</td>
<td>0.00065 (2.41)**</td>
</tr>
<tr>
<td>observations</td>
<td>40</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.76</td>
</tr>
</tbody>
</table>

dependent variable: land leverage

* t statistics in parentheses
** significant at 10%, *** significant at 5%
**** significant at 1%
Table 1.4: Cross-Sectional Regression With Naively Imputed Land Leverage Data

<table>
<thead>
<tr>
<th>Variables</th>
<th>Open City</th>
<th>Growth Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>-0.01063</td>
<td>-0.01126</td>
</tr>
<tr>
<td></td>
<td>(-8.10)**</td>
<td>(-8.27)**</td>
</tr>
<tr>
<td>$\mu$</td>
<td>-0.00005</td>
<td>-0.00005</td>
</tr>
<tr>
<td></td>
<td>(-6.95)**</td>
<td>(-6.51)**</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
<tr>
<td></td>
<td>(5.90)**</td>
<td>(5.51)**</td>
</tr>
<tr>
<td>leverage</td>
<td>0.00593</td>
<td>0.00528</td>
</tr>
<tr>
<td></td>
<td>(5.46)**</td>
<td>(4.22)**</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.05128</td>
<td>0.04880</td>
</tr>
<tr>
<td></td>
<td>(5.66)**</td>
<td>(5.25)**</td>
</tr>
<tr>
<td>$tc$</td>
<td>0.00548</td>
<td>0.00519</td>
</tr>
<tr>
<td></td>
<td>(7.99)**</td>
<td>(7.26)**</td>
</tr>
<tr>
<td>reg</td>
<td>0.00009</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.69)*</td>
<td></td>
</tr>
<tr>
<td>aggr</td>
<td>-0.00000</td>
<td>(-0.96)</td>
</tr>
<tr>
<td>observations</td>
<td>271</td>
<td>271</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.606</td>
<td>0.611</td>
</tr>
</tbody>
</table>

Dependent variable: variance of housing return

T statistics in parentheses

* significant at 10%, ** significant at 5%

*** significant at 1%
Table 1.5: Cross-Sectional Regression With MCMC Imputed Land Leverage Data

<table>
<thead>
<tr>
<th>Variables</th>
<th>Open City</th>
<th>Growth Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>constant</td>
<td>-0.01072</td>
<td>-0.01158</td>
</tr>
<tr>
<td></td>
<td>(-7.92)***</td>
<td>(-8.36)***</td>
</tr>
<tr>
<td>$\mu$</td>
<td>-0.00006</td>
<td>-0.00005</td>
</tr>
<tr>
<td></td>
<td>(-7.49)***</td>
<td>(-6.81)***</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.00001</td>
<td>0.00001</td>
</tr>
<tr>
<td></td>
<td>(6.91)***</td>
<td>(6.22)***</td>
</tr>
<tr>
<td>leverage</td>
<td>0.00267</td>
<td>0.00204</td>
</tr>
<tr>
<td></td>
<td>(3.52)***</td>
<td>(2.59)***</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.06910</td>
<td>0.05888</td>
</tr>
<tr>
<td></td>
<td>(8.44)***</td>
<td>(6.58)***</td>
</tr>
<tr>
<td>$tc$</td>
<td>0.00571</td>
<td>0.00508</td>
</tr>
<tr>
<td></td>
<td>(8.08)***</td>
<td>(6.97)***</td>
</tr>
<tr>
<td>reg</td>
<td>0.00014</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.76)***</td>
<td></td>
</tr>
<tr>
<td>aggr</td>
<td>-0.00000</td>
<td>(-0.46)</td>
</tr>
<tr>
<td>observations</td>
<td>271</td>
<td>271</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.581</td>
<td>0.595</td>
</tr>
</tbody>
</table>

Dependent variable: variance of housing return

* Significant at 10%, ** Significant at 5%
*** Significant at 1%

Table 1.6: 10 Most Volatile Housing Returns MSAs

<table>
<thead>
<tr>
<th>Rank</th>
<th>MSA</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Honolulu,HI</td>
<td>0.0156</td>
</tr>
<tr>
<td>2</td>
<td>Fairbanks,AK</td>
<td>0.0120</td>
</tr>
<tr>
<td>3</td>
<td>Barnstable Town,MA</td>
<td>0.0092</td>
</tr>
<tr>
<td>4</td>
<td>Oxnard- Thousand Oaks-Ventura, CA</td>
<td>0.0086</td>
</tr>
<tr>
<td>5</td>
<td>New Haven-Milford, CT</td>
<td>0.0084</td>
</tr>
<tr>
<td>6</td>
<td>Riverside-San Bernardino-Ontario, CA</td>
<td>0.0083</td>
</tr>
<tr>
<td>7</td>
<td>Los Angeles-Long Beach-Glandale, CA</td>
<td>0.0083</td>
</tr>
<tr>
<td>8</td>
<td>Modesto, CA</td>
<td>0.0080</td>
</tr>
<tr>
<td>9</td>
<td>Santa Ana-Anaheim-Irvine, CA</td>
<td>0.0080</td>
</tr>
<tr>
<td>10</td>
<td>San Jose-Sunnyvale-Santa Clara, CA</td>
<td>0.0078</td>
</tr>
</tbody>
</table>
Table 1.7: 10 Least Volatile Housing Returns MSAs

<table>
<thead>
<tr>
<th>Rank</th>
<th>MSA</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Goldsboro,NC</td>
<td>0.0003</td>
</tr>
<tr>
<td>2</td>
<td>Appleton, WI</td>
<td>0.0003</td>
</tr>
<tr>
<td>3</td>
<td>Clarksville, TN-KY</td>
<td>0.0004</td>
</tr>
<tr>
<td>4</td>
<td>Florence, SC</td>
<td>0.0004</td>
</tr>
<tr>
<td>5</td>
<td>Decatur, IL</td>
<td>0.0004</td>
</tr>
<tr>
<td>6</td>
<td>Fayetteville, NC</td>
<td>0.0004</td>
</tr>
<tr>
<td>7</td>
<td>Warner Robins, GA</td>
<td>0.0004</td>
</tr>
<tr>
<td>8</td>
<td>Tuscaloosa, AL</td>
<td>0.0004</td>
</tr>
<tr>
<td>9</td>
<td>Oshkosh-Neenah, WI</td>
<td>0.0004</td>
</tr>
<tr>
<td>10</td>
<td>Erie, PA</td>
<td>0.0005</td>
</tr>
</tbody>
</table>
Figure 1.1: Average House Price Appreciation Rate and Variance of House Price Appreciation Rate by Census
CHAPTER 2

SKEWNESS OF THE HOUSING RETURNS AND HOUSING TENURE CHOICE

2.1 Introduction

It is widely recognized in the housing literature that owning a house involves both a consumption choice and an investment portfolio decision. Owner-occupied housing is a major investment for most households in the U.S. and other countries. Also, it has been shown both theoretically and empirically in the finance literature that risk-averse individuals who exhibit decreasing absolute risk aversion prefer investments with positively skewed returns. As housing is usually the largest part of people’s investment portfolio, then do risk-averse individuals also show a preference for positive skewness in the housing market? That is, does the skewness of the housing returns affect households’ housing tenure choice? Does positive skewness of the housing returns increase the likelihood of owner-occupation? Answers to these questions are not straightforward because housing is different from stocks and bonds in that it is not only an investment decision but also a consumption choice. Therefore, what is observed in the finance literature does not necessarily also hold in the housing market, which motivates this paper.

10 A positively skewed distribution has a fat right tail and a negatively skewed distribution has a fat left tail.
Literature related to this topic includes studies of housing tenure choice and papers on preference for positive skewness. The literature on skewness preference has been well established. Arditti (1967) showed both theoretically and empirically that risk averse investors prefer positively skewed returns. The author claimed that this is a common-sense result because a risk averter is reluctant to undertake any investment that presents him with the possibility of a large loss and only a limited gain. Tsiang (1972) theoretically proved individual’s preference for skewness in any economic decision involving an uncertain outcome. More specifically, he showed that skewness preference is a general pattern of behavior towards uncertainty for all risk-averse individuals with decreasing or constant absolute risk-aversion with respect to increases in wealth. Kraus and Litzenberger (1976) developed a market equilibrium model and showed that skewness is preferred and is relevant to the market valuation of risky assets. Scott and Horvath (1980) theoretically proved that for the usual risk-averse investor, the preference direction is positive for every odd central moment of returns and negative for every even central moment.

Topics on housing tenure choice also are well explored. Many demographic and social economic variables are found to be closely related to people’s housing tenure choice in the literature, among which income uncertainty is considered as an important factor. Haurin and Gill’s (1987) empirical findings showed that the consumption of housing falls when the uncertainty of receipt of income increases. Dynarski and Sheffrin (1985) studied the transitory income shocks and found that positive transitory shocks increase the likelihood of home purchases. Haurin (1991) found a negative relationship between income risk and the likelihood of home ownership. A 10% increase in income variability reduced homeownership by the same amount as a 5% decrease in income. Recent work of Diaz-Serrano (2005) considered the skewness of household income and housing tenure choice. Using data of
Germany and Spain, he observed that households located in more positively skewed income distributions show a greater propensity for home ownership. In terms of household’s wealth portfolio management, a recent paper by Sinai and Souleles (2005) claimed that rental market could also be risky, and therefore owning a house can be used as a hedge against rent risk. Other variables that may affect housing tenure choice include transaction costs of owning a house, expected length of stay, mortgage interest rate, depreciation rate, household’s tax rate, property tax rate, commuting costs, relative price of owner-occupied housing to rental housing of the same quality, etc. Demographic variables including race, ethnicity, marital status, and gender have also been shown to have effects on people’s housing tenure choice.

The main purpose of this paper is to set up an appropriate theoretical model to derive the relationship between the skewness of the housing returns and housing tenure choice and test the theoretical predictions empirically. More specifically, based on the models of housing tenure choice developed by Henderson & Ioannides (1983) and Fu (1995), this paper derives the direction of the relationship between housing tenure choice and skewness of the housing returns from the model. Theoretical conjectures derived from the model are then tested empirically using the MSA level data sets.

The paper proceeds as follows. Section 2.2 introduces the basic model and derives the relationship between the skewness of the housing returns and housing tenure choice through comparative static analysis. Impacts of other variables of interest on housing tenure choice also are discussed and testable predictions are yielded. Section 2.3 provides empirical investigation and section 2.4 concludes. The last section points out policy implications of our findings and some potential related topics for future research.
2.2 The Model

The theoretical model is based on the models of housing tenure choice developed by Henderson & Ioannides (1983) and Fu (1995).

2.2.1 Basic Assumptions

Owning a house involves both a consumption choice and a portfolio decision. We denote housing consumption with $h_c$ and housing investment with $h_i$. Household derives utility from current housing consumption $h_c$ and a numeraire non-housing good $x$ in period 1, and from future consumption in period 2. Household’s future consumption depends on savings $s$ and an uncertain equity in housing investment $h_i$. The uncertainty comes from the house price appreciate rate (or housing return) $\theta$, which is a random variable with $E(\theta) = \overline{\theta}$, $var(\theta) = \sigma^2$, $E(\theta - \overline{\theta})^3 = \mu_3$, and zero-valued moments above the third order.

Due to institutional constraints, homeowners must own at least as much housing as they consume. So, if $h_i \geq h_c$, households will owner-occupy the house. If $h_i < h_c$, households either choose to rent, or distort their investment and consumption choices, raising $h_i$ to $h_c$ and owner-occupy. Therefore, as $h_i$ increases and $h_c$ decreases (or $h_c$ fixed), the likelihood of owner-occupying increases.

Additionally, as supported by empirical and experimental evidence in the literature, we assume that households are risk averse and exhibit decreasing absolute risk aversion.

2.2.2 A Household’s Budget Constraints

The household receives incomes of $y_1$ and $y_2$ in periods 1 and 2 respectively. Let $P$ be the price per unit of housing, $L$ be an interest-accrual mortgage loan per unit of housing, $R$
be the imputed rental income for period 1 per unit of housing, $s$ be savings in period 1, and $r$ be a riskless interest rate. The budget constraints are:

$$W_1 = y_1 - s - (P - L - R)h_i$$

(2.1)

$$W_2 = y_2 + s(1 + r) + [P(1 + \theta) - L(1 + r)]h_i,$$

(2.2)

where $(P - L - R) > 0$ is the net of rental income down-payment on housing investment in period 1, $[P(1 + \theta) - L(1 + r)]h_i$ is the equity in period 2 and $W_1$ and $W_2$ denote the consumption budget for period 1 and 2, respectively. To avoid dealing with the costs of defaults, we assume $[P(1 + \theta) - L(1 + r)] > 0$. Furthermore, we also have a liquidity constraint which assumes that $s \geq 0$, so that people cannot borrow against future. So in period 1, household has housing investment demand $h_i$, which has a rental value of $R$. At the same time, the household rents $h_c$ units of housing for consumption purpose at a price of $R$. In period 2, the household can sell the asset for an unknown return of $\theta$.

### 2.2.3 Household’s Utility Maximization

Let $V_1(W_1)$ and $V_2(W_2)$ denote the indirect utility functions for period 1 and period 2, respectively. And because we assume that households are risk averse and exhibit decreasing absolute risk aversion, we then have $V_i' = dV_i/dW_i > 0$ for $i = 1, 2$; $V_i'' = d^2V_i/dW_i^2 < 0$ for $i = 1, 2$; $V_i''' = d^3V_i/dW_i^3 > 0$ for $i = 1, 2$.

---

11 We assume that households leave no bequests and therefore saving in period 2 is zero.

12 As we’ve mentioned, $W_1$ is the budget for current consumption of $h_c$ and $x$. So given that the price of numeraire good $x$ is assumed to be 1, $W_1 = x + Rh_c$.

13 Utility functions that have these properties include the negative exponential function $V(W) = B(1 - e^{-\alpha W})$, and the family of constant elasticity utility functions: $V(W) = \left[1/(1 - a)\right]W^{1-a}$ $(a > 0)$, and $V(W) = \log W$, etc.
Risk-averse household selects \( h_i, s, h_c \), and \( x \), subject to the budget constraints, so as to maximize expected utility:

\[
V = V_1(W_1) + E[V_2(W_2)].
\] (2.3)

From (2.2), we get

\[
E(W_2) = \bar{W}_2 = y_2 + s(1 + r) + [P(1 + \bar{\theta}) - L(1 + r)] h_i,
\] (2.4)

Therefore, it follows that

\[
W_2 - \bar{W}_2 = P h_i(\theta - \bar{\theta}).
\] (2.5)

Using Taylor expansion around \( \bar{W}_2 \) and using equation (2.5), we can get:

\[
E[V_2(W_2)] = V_2(\bar{W}_2) + \frac{1}{2!} V''_2(\bar{W}_2) P^2 h_i^2 \sigma^2 + \frac{1}{3!} V'''_2(\bar{W}_2) P^3 h_i^3 \mu_3.
\] (2.6)

Because the focus of this paper is on the impact of the skewness of the housing returns on housing tenure choice, we only look at the first order condition with respect to \( h_i \); that is, \( \partial V / \partial h_i = 0 \).

From (2.3) and (2.6), the first order condition with respect to \( h_i \) is

\[
\frac{\partial V}{\partial h_i} = 0
\]

\[
= V'_1(W_1) \alpha + V'_2(\bar{W}_2) \beta + \left( V''_2(\bar{W}_2) P^2 h_i \sigma^2 + \frac{1}{2} \beta V'''_2(\bar{W}_2) P^2 h_i^2 \sigma^2ight)
\]

\[
+ \frac{1}{2} V''_2(\bar{W}_2) P^3 h_i^2 \mu_3 + \frac{1}{6} V'''_2(\bar{W}_2) \beta P^3 h_i^3 \mu_3,
\] (2.7)

where \( \alpha = [- (P - L - R)] < 0 \), \( \beta = [P(1 + \bar{\theta}) - L(1 + r)] > 0 \).

\(^{14}\)Discounting factor could be added in front of \( E[V_2(W_2)] \) to discount the utility derived from the future period (period 2), but the results are not affected no matter the discounting factor is present or not.
2.2.4 Comparative Static Analysis

To find out the relationship between skewness of the housing returns and household’s housing tenure choice, we can do the comparative static analysis using (2.7) and get:

\[
\frac{dh_i^*}{d\mu_3} = \left[ \frac{1}{2} V_2''(W_2) + \frac{1}{6} V_2'''(W_2) \beta h_i^* \right] P^3 h_i^* \left( -\frac{\partial^2 V}{\partial h_i^2} \right)^{-1}. \tag{2.8}
\]

We know \( V_2''(W_2) > 0 \), and we can show \( \partial^2 V / \partial h_i^2 < 0 \), so the sign of \( dh_i^* / d\mu_3 \) depends on \( \frac{1}{6} V_2'''(W_2) \beta h_i^* \). By Scott and Horvath (1980), \( V_2'''(W_2) < 0 \), but if the fourth derivative of the utility given budget level \( W_2 \) is extremely small which makes \( \frac{1}{6} V_2'''(W_2) \beta h_i^* \) ignorable, then \( dh_i^* / d\mu_3 > 0 \). This is very likely to happen as we’ve mentioned that utility functions which have properties we assumed in this paper include the negative exponential function \( V(W) = B(1 - e^{-\gamma W}) \), the family of constant elasticity utility functions such as \( V(W) = \left[ 1/(1 - a) \right] W^{1-a} \) \((a > 0)\), and \( V(W) = \log W \), etc. The fourth derivatives of all these utility functions involve the fourth power of a positive number which is smaller than one. For example, when \( V(W) = \log W \), we have \( V'''(W) = -6(1/W)^4 \), which could be extremely small. And in fact we can analytically show that when \( V(W) = \log W \), \( dh_i^* / d\mu_3 \) is positive.\(^{16}\)

Given fixed period 1 income \( y_1 \) and a binding liquidity constraint \((s \equiv 0)\), we have \( dW_1^* / d\mu_3 = a(dh_i^* / d\mu_3) < 0 \), that is, \( dh_c^* / d\mu_3 < 0 \). This implies that ceteris paribus, a larger value of skewness leads \( h_i^* \) to increase and \( h_c^* \) to decrease, therefore, skewness of the housing returns has a positive impact on likelihood of owning.

\(^{15}\)\( \frac{\partial^2 V}{\partial h_i^2} = a^2 V''_1 + E \{ V''_2 \beta^2 \} < 0 \)

\(^{16}\)Please refer to the appendix.

\(^{17}\)Following Fu (1995), we assume that the liquidity constraint is binding in our model. That is, households neither borrow nor save.

\(^{18}\)We assume that the signs of the comparative statics of \( W_1^* \) apply to those of \( h_c^* \), as housing is a normal good and \( h_c^* \) increases with \( W_1^* \).
Other interesting comparative statics may include the effects of expected housing returns and variance of the housing returns on a household’s housing tenure decision. The comparative static analysis with respect to $d\bar{\theta}$ is:

$$
\frac{dh_i^*}{d\bar{\theta}} = \left[ V''_2(W_2) P + \frac{1}{2} V'''_2(W_2) \beta P h_i^* + \frac{1}{6} V'''_2(W_2) P^3 h_i^* \sigma^2 \right] \left( -\frac{\partial^2 V}{\partial h_i^*} \right)^{-1}
$$

(2.9)

The sign of $dh_i^*/d\bar{\theta}$ is ambiguous given the properties of utility function $V(W)$. Therefore, the net impact of a change in $\bar{\theta}$ (expected level of housing returns) on housing tenure choice can not be determined. This result is consistent with Fu (1995)’s findings, and the ambiguous sign implies that a higher $\bar{\theta}$, "on the one hand, makes housing investment less attractive by reducing the expected future marginal utility and on the other hand, it makes housing investment more attractive by raising the expected return of housing investment and reducing the risk premium."

The comparative static analysis with respect to $\sigma^2$ is:

$$
\frac{dh_i^*}{d\sigma^2} = \left[ V''_2(W_2) P^2 h_i^* + \frac{1}{2} \beta V'''_2(W_2) P^2 h_i^* \right] \left( -\frac{\partial^2 V}{\partial h_i^*} \right)^{-1}
$$

(2.10)

The sign of $dh_i^*/d\sigma^2$ is also ambiguous since $V''_2(W_2) < 0$ and $V'''_2(W_2) > 0$. Therefore, the net impact of a change in $\sigma^2$ (variance of the housing returns) on housing tenure choice can not be determined. This is also consistent with Fu (1995)’s conclusion, the economic implication of this ambiguous sign of $dh_i^*/d\sigma^2$ is that a larger $\sigma^2$, "on the one hand, makes housing investment more attractive by raising the expected marginal utility and on the other hand, it makes housing investment less attractive by increasing the risk premium."

However, similar as the comparative statics of $h_i^*$ in terms of $\mu_3$, we can again claim that $dh_i^*/d\sigma^2$ is very likely to be negative, given the forms of the utility functions that have
properties we assumed in this paper. And again, we can show that if \( V(W) = \log W \), then \( dh^*_{i}/d\sigma^2 \) is negative.\(^{19}\) Consequently, with fixed period 1 income \( y_1 \) and binding liquidity constraint \((s \equiv 0)\), we have \( dW_1^*/d\sigma^2 = \alpha (dh^*_i/d\sigma^2) > 0 \), that is, \( dh^*_c/d\sigma^2 > 0 \). So as \( \sigma^2 \) increases, demand for housing investment \( h^*_i \) decreases while demand for housing consumption \( h^*_c \) increases. Therefore, we claim that the variance of the housing returns \( \sigma^2 \) is very likely to be negatively related to people’s likelihood of owning.

Furthermore, the partial derivative with respect to the rent level \( R \), turns out to be:

\[
\frac{dh^*_i}{dR} = \left[ V''_1(W_1)h^*_i \alpha + V'_1(W_1) \right] \left( -\frac{\partial^2 V}{\partial h^*_i} \right)^{-1} > 0,
\]

(2.11)
given that \( V''_1(W_1) < 0, V'_1(W_1) > 0 \) and \( \alpha < 0 \). On the other hand, we have \( dh^*_c/dR < 0 \).\(^{20}\) This implies that as the rental value of a property increases, the investment demand for housing increases while the consumption demand for housing decreases, and therefore, the likelihood of owning a house increases.

Following Fu (1995), let \( W = y_1 + \frac{y_2}{1+r} \) denote the household’s wealth (or the present value of household’s total lifetime income), and we use \( W/P \), which is the ratio of household’s wealth to house price, to measure the affordability of housing, we then can check the comparative statics of \( h^*_i \) and \( h^*_c \) in terms of \( W/P \).

\[
\frac{dh^*_i}{d(W/P)} = \left[ \frac{1}{P} \left( \frac{dh^*_i}{dW} \right)^{-1} - \frac{W}{P^2} \left( \frac{dh^*_i}{dP} \right) \right]^{-1}
\]

(2.12)

\( dh^*_i/dW \) can be proved to be positive according to Fu (1995)\(^{21}\). However, \( dh^*_i/dP \), which is the comparative static analysis with regard to house price \( P \) is very complicated.

\(^{19}\)Please refer to the appendix.

\(^{20}\)\( dW_1/dR \) could be negative, zero or positive, but we assume that \( dh^*_c/dR < 0 \), given the fact that \( R \) is the price of housing consumption.

\(^{21}\)Please refer to proposition (4) of Fu (1995) and the corresponding proof in that paper’s appendix.
and does not produce a clear direction of relationship with $h^*_t$. Therefore, the net impact of the affordability of housing $W/P$ can not be determined from the model. We will find it out by our empirical investigation.

### 2.3 Empirical Investigation

#### 2.3.1 Hypotheses To Be Tested

We’ve found from the theoretical model that people’s housing tenure choice should be closely related to the first three moments of housing returns, and also to a household’s income, house price, and rental price, among which the skewness of the housing returns is of our major interest. More specifically, from the theoretical analysis, we proposed that it is very likely that the skewness of the housing returns is positively related with the likelihood of owning a house, that is, the more right skewed the housing return is, the more likely that people will choose to be a homeowner instead of a renter, ceteris paribus. We also found that it is very likely that the variance of the housing returns is negatively related to the likelihood of owning. That is, the more volatile the housing market is, the less likely that people will choose to become owner-occupiers. We also obtained a prediction about the relationship between rent level and likelihood of owning, and it is positive. Our comparative static analysis gave us ambiguous signs with respect to all the remaining variables, and we will find out the directions of those relationships through our empirical work.
2.3.2 Data

Our data sets are at the MSA\textsuperscript{22} level and 273 MSAs of the United States are included. Data for different variables comes from different sources.

Housing tenure data is obtained from Census 2000 conducted by U.S. Census Bureau. For each MSA, we obtain the number of households who choose to own and number of households who choose to rent respectively, from which we can calculate the homeowner-ship rate for each MSA.

Moments of housing returns are calculated using CMHPI (Conventional Mortgage House Price Index) data provided by Freddie Mac and Fannie Mae. The CMHPI data is published quarterly from 1975 to the present for each MSA, from which we can compute the average housing return\textsuperscript{23}, variance (or standard deviation) of housing return, and skewness of housing return at MSA level\textsuperscript{24}. The CMHPI data is not adjusted for inflation, so in order to obtain the moments of real housing returns, we used CPI (Consumer Price Index) from the Federal Reserve Bank to obtain an inflation adjusted house price index.

Median house price data for each MSA is obtained from Census 2000 and so is median household income data.

Rental price for each MSA is provided by HUD (U.S. Department of Housing and Urban Development), and more specifically, we used the MSA level monthly rental price of a typical 2-bedroom apartment in year 2000.

Descriptive statistics of the data sets are provided in table (2.1) in the appendix.

\textsuperscript{22}MSA: Metropolitan Statistical Area. Usually refers to an urban area with at least 50,000 population plus the surrounding counties.

\textsuperscript{23}Housing return, refers to the year over year quarterly increase rate of house price. For this paper, we only used data up to the last quarter of 2000, because the data for all the other variables are from Census 2000.

\textsuperscript{24}Skewness is calculated using formula $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$, where $s$ is the estimated standard deviation and calculated as $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$.
2.3.3 Weighted Least Square Regression Model

A linear cross-sectional model as in equation (2.13) is used in our empirical regression.

\[
\ln \frac{\bar{P}_i}{1 - \bar{P}_i} = \alpha + \beta_1 \text{mean}_i + \beta_2 \text{var}_i + \beta_3 \text{skewness}_i + \beta_4 \text{rent}_i + \beta_5 \text{incomehp}_i \\
+ \beta_6 \text{propertytax}_i + \beta_7 \text{education}_i + \beta_8 \text{white}_i + \beta_9 \text{asian}_i \\
+ \beta_{10} \text{black}_i + \varepsilon_i, \tag{2.13}
\]

where \(\alpha\) is a constant, \(\beta_1, ... \beta_6\) are parameters to be estimated, \(\varepsilon_i\) is the error term.

\(\bar{P}_i\) is the observed homeownership rate for MSA \(i\), we make logit transformation here to avoid the problem of getting some estimated \(\hat{P}_i\) that is either larger than 1 or negative. \(\text{mean}_i\) is the average real house price appreciation rate for MSA \(i\), \(\text{var}_i\) is the sample variance of the real house price appreciation rate, \(\text{skewness}_i\) is the skewness of the real house price appreciation rate, \(\text{rent}_i\) is the average rental price for a typical 2-bedroom apartment in MSA \(i\), \(\text{incomehp}_i\) is the household income over house price ratio, which measures the housing affordability. \(\text{propertytax}_i\) is the MSA level property tax rate, \(\text{education}_i\) is the percentage of population with Bachelor’s degree or higher for MSA \(i\), the last three regressors \(\text{white}_i, \text{asian}_i\) and \(\text{black}_i\) represent percentage of white, Asian and African American population respectively.

The first 5 explanatory variables are from our theoretical model, and the remaining 5 are variables that are widely used as explanatory variables for people’s housing tenure choice in the literature. We will run 2 regressions separately, with the first one only including the first 5 regressors derived from the theoretical model, and the second one involving all the ten variables.
According to our theoretical predictions, we would expect: 1) $\beta_2$, $\beta_3$, and $\beta_4$ should all be significant; 2) $\beta_3$, the coefficient for skewness of the housing returns, should be significant and positive; 3) $\beta_2$, the coefficient for variance of the housing returns, should be significant and negative. 4) $\beta_4$, the coefficient for rental price, should also be significant and positive.

In terms of the estimation method, weighted least square estimation is used because we assume heteroskedasticity in the error term $\varepsilon_i$. We believe that the variance of the error term decreases as population increases, since then $\overline{P_i}$ is a better estimate of the true $P_i$. Therefore, population is used as the weight for the weighted least square regression.

2.3.4 Results

WLS Regression Results

Table (2.2) in the appendix presents the main results of the WLS regression. Model I of the table presents results for the regression with the 5 variables derived from the theoretical model.

We can see that all the 5 explanatory variables in model I are significant. This is consistent with our theoretical model’s implication. Therefore, mean, variance and skewness of real housing returns are shown to be significantly related to people’s housing tenure choice, so are rental price and the income over house price ratio.

The skewness of the housing returns is found to be positively related to the likelihood of owning, which is consistent with our theoretical conjecture. So, controlling for other factors, right (or positively) skewed housing returns have a positive impact on people’s housing tenure choice, that is, there is also a preference for positive skewness in the housing market. More specifically, according to our coefficient estimation, each one standard deviation increase in the skewness of the housing returns will approximately double the
odds of owning (instead of renting). The intuition behind this preference for skewness is that, compared with a symmetric or negatively skewed distribution, a large deviation below the mean is less likely to occur for positively skewed distributions. Therefore, holding mean and variance constant, risk averse people tend to prefer investments returns with positively skewed distributions.

We also find that likelihood of being an owner-occupier is positively related to the mean (or average) housing returns, that is, ceteris paribus, households in areas with higher average housing returns are more likely to buy houses instead of renting. This is quite reasonable and easy to interpret. Considering buying a house as an investment behavior, higher average housing returns make owning a house more attractive, and this increases people’s likelihood of owning, ceteris paribus. Or, using Fu (1995)’s reasoning, our empirical results indicate that people tend to put more value on higher expected returns and lower risk premium rather than on higher expected future marginal utility of housing investment.

The volatility of the housing returns, as measured by the sample variance of the housing returns, is found to be negatively related to the likelihood of owning, which is consistent with our theoretical prediction. This is also quite intuitive given that we assume most people are risk-averse. A highly volatile housing market makes people feel anxious and leads to increasing concerns about the appreciation potential of their housing investment. This will discourage people from making housing investments, and therefore reduce the probability of home ownership. Again, this negative relationship reconfirms people’s preference for higher expected returns and a lower risk premium over higher expected future marginal utility of housing investment, which also gives empirical evidence that most people are risk averse and exhibit decreasing absolute risk aversion.
The rental price is shown to be positively related to the probability of owning. This complies with our theoretical prediction and is also consistent with empirical findings presented by many others in the literature. Renting and owning can be considered as substitutes to each other, therefore, as rental price increases, the demand for owner-occupied housing increases accordingly, leading to a positive relationship between rental price and people’s likelihood of owning, ceteris paribus. Or, deriving intuition from our theoretical model, rental price $R$ is the price of housing consumption, as rental price increases, people will substitute housing investment for housing consumption, leading to increased likelihood of becoming homeowners.

Income over house price ratio, which measures the housing affordability for a particular MSA, is positively related to the likelihood of owning. This is straightforward in that people usually would like to own instead of to rent if housing is affordable, considering the many advantages of becoming homeowners, especially the advantage of tax deduction attached to it.

Model 2 is the regression model which involves all of the explanatory variables as listed in equation (2.2). The estimation results are also presented in table (2.2). Compared with model 1, we can see that we get very similar estimates of the coefficients for the first 5 variables. Again, skewness is found to be positively and significantly related to people’s likelihood of owning, and so are the mean (or average) housing returns, rental price and income over house price ratio. The variance of housing returns is still negatively related to the likelihood of becoming homeowners. Furthermore, property tax is significantly and negatively related to the likelihood of owning, which is quite intuitive. Other things equal, higher property tax increases the user cost of owning a house, and therefore discourages
people from buying a house. Education is also found to have a negative impact on the likelihood of owning. Recall that the regressor education measures percentage of population with Bachelor’s degree or higher for each MSA, so this regression result implies that people in areas with higher overall education level tends to be less likely to own. A possible story behind this negative relationship is that people with higher educational achievement usually relocate more frequently and therefore have a shorter expected length of stay on average, which makes them less willing to buy a house and get completely settled down. Finally, white is shown to be significantly and positively related to homeownership rate, which implies that areas with higher percentage of white people in the population tend to have higher homeownership rates. The percentages of Asian and African American people in the population are not significant in the regression.

To conclude, by using MSA level data, our project shows that the 5 variables from the theoretical model are significantly related to household’s housing tenure choice. More specifically, the skewness of housing returns, average housing returns, rental price, and income over house price ratio are positively related to people’s likelihood of being home owners, while volatility of the housing returns is negatively related to the probability of owning. Additionally, we also found that property tax and education are negatively related to people’s likelihood of being home owners, while percentage of white people in the population is positively related to it. Overall our empirical model explains about 75% of the cross-MSA variation in the homeownership rate.

2.4 Conclusion

House is the largest investment for most American families and preference for investment with positively skewed returns is well recognized in finance literature. Buying a house
is a investment decision as well as a consumption choice, which makes housing investment different from any other financial investments such as stocks and bonds. Therefore, what works for the finance market does not automatically also work for the housing market. This paper studies whether this preference for skewness also exists in the housing market and whether the skewness of the housing returns affects households’ housing tenure choice. Based on the theoretical model of housing tenure choice developed by Henderson & Ioannides (1983) and Fu (1995), this paper derives the relationship between housing tenure choice and the skewness of the housing returns from the model and predicts that it is very likely that the skewness of housing returns is positively related to the likelihood of owning a house, ceteris paribus.

Our empirical investigation supports our conjectures and shows that skewness of the housing returns is positively related to people’s likelihood of being a owner-occupier, and therefore, preference for positive skewness also exists in the housing market.

We also find that likelihood of being an owner-occupier is positively related to the mean (or average) housing returns and negatively related to volatility of the housing returns. This gives empirical evidence for people’s preference for higher expected returns and lower risk premium over higher expected future marginal utility of housing investment.

Additionally, the income over house price ratio, rental price, and percentage of white people in the population are found to be positively related to the probability of owning, as indicated by our regression results. Property tax and the overall education level, on the other hand, has negative relationships with the likelihood of being a owner-occupier.
2.5 Future Research

2.5.1 Policy Implications

We’ve shown both theoretically and empirically in this paper that preference for positively skewed investment returns also exists in the housing market. Therefore, other things equal, we can argue that risk-averse people in areas with more right skewed housing returns will be more willing or more likely to become a homeowner instead of a renter.

We know that homeownership is now considered as the "American dream" in the United States and has been continuously supported by the U.S. government. This paper has presented theoretical and empirical evidence that skewness of the housing returns does play a role in people’s housing tenure decision. Therefore, if promoting homeownership is still one of the major tasks in a policy maker’s mind, then policies that could have an effect on the skewness of the housing returns should be appropriate. One example is the house price insurance program proposed by Case and Shiller, the so called "Chicago Plan". In the plan you can sell contracts that will reap you a profit if local prices fall, allowing you to lock in the current value of your home. If this plan is adopted, it should push the distribution of the housing returns rightward, shortening the left tail of the distribution and resulting in a more positively skewed housing returns. The predicted outcome is that households become more likely to own a house because with house price insurance, they no longer need to worry too much about suffering a loss in housing investments.

2.5.2 Determinants of the Skewness of the Housing Returns

Policy makers would be more informed if they know more about potential determinants of the skewness of the housing returns. If we look at the summary statistics in table (2.1), we can see that there is a huge variation in the skewness of the real housing returns across
MSAs. For examples, the highest skewness equals 2.168 (Madison, WI), while the lowest is equal to -2.053 (Albany, GA), and the standard deviation is 0.728. So why does the skewness of the housing returns differ so dramatically across different places? Which factors could be potential determinants of or be closely related to the skewness of the housing returns? At MSA level, could the level of growth controls or land use regulation be a potential determinant? Is there something else involved? Further research is needed. At a more micro level, could the quality of school districts be a possible candidate? Evidence has been provided that house prices in good school districts have downward rigidity during housing market recessions. Therefore, it is quite likely that housing returns could be more positively skewed for houses in good school districts. This could be an interesting topic for further research, also.
2.6 Tables and Proofs

2.6.1 Proof of statement related to footnote 17

If \( V(W) = \log W \), (2.8) becomes:

\[
\frac{dh_i^*}{d\mu_3} = \left[ \frac{1}{2} V''(W_2) + \frac{1}{6} V'''(W_2) \beta h_i^* \right] P^3 h_i^{*2} \left( -\frac{\partial^2 V}{\partial h_i^2} \right)^{-1}
\]
\[
= \left[ \frac{1}{2} \times 2 \frac{1}{W_2} - \frac{1}{6} \times 6 \frac{1}{W_2} \beta h_i^* \right] P^3 h_i^{*2} \left( -\frac{\partial^2 V}{\partial h_i^2} \right)^{-1}
\]
\[
= \frac{1}{W_2} (W_2 - \beta h_i^*) P^3 h_i^{*2} \left( -\frac{\partial^2 V}{\partial h_i^2} \right)^{-1}
\]
\[
= \frac{1}{W_2} (y_2 + s(1 + r) + \beta h_i^* - \beta h_i^*) P^3 h_i^{*2} \left( -\frac{\partial^2 V}{\partial h_i^2} \right)^{-1}
\]
\[
= \frac{1}{W_2} (y_2 + s(1 + r)) P^3 h_i^{*2} \left( -\frac{\partial^2 V}{\partial h_i^2} \right)^{-1} > 0
\]

2.6.2 Proof of statement related to footnote 20

If \( V(W) = \log W \), (2.10) becomes:

\[
\frac{dh_i^*}{d\sigma^2} = \left[ V''(W_2) P^2 h_i^* + \frac{1}{2} \beta V'''(W_2) P^2 h_i^{*2} \right] \left( -\frac{\partial^2 V}{\partial h_i^2} \right)^{-1}
\]
\[
= \left[ -\frac{1}{W_2} P^2 h_i^* + \frac{1}{2} \times 2 \frac{1}{W_2} \beta P^2 h_i^{*2} \right] \left( -\frac{\partial^2 V}{\partial h_i^2} \right)^{-1}
\]
\[
= \left[ -\frac{1}{W_2} (W_2 - \beta h_i^*) \right] P^2 h_i^{*} \left( -\frac{\partial^2 V}{\partial h_i^2} \right)^{-1}
\]
\[
= \left[ -\frac{1}{W_2} (y_2 + s(1 + r)) \right] P^2 h_i^{*} \left( -\frac{\partial^2 V}{\partial h_i^2} \right)^{-1} < 0
\]
2.6.3 Tables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Median</th>
<th>Stdev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>homeownership rate</td>
<td>0.671</td>
<td>0.677</td>
<td>0.061</td>
<td>0.456</td>
<td>0.929</td>
</tr>
<tr>
<td>mean housing returns</td>
<td>0.006</td>
<td>0.005</td>
<td>0.013</td>
<td>-0.026</td>
<td>0.046</td>
</tr>
<tr>
<td>variance of housing returns</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
<td>0.000</td>
<td>0.009</td>
</tr>
<tr>
<td>skewness of housing returns</td>
<td>-0.133</td>
<td>-0.144</td>
<td>0.728</td>
<td>-2.053</td>
<td>2.168</td>
</tr>
<tr>
<td>rental price (thousand $)</td>
<td>0.550</td>
<td>0.528</td>
<td>0.113</td>
<td>0.358</td>
<td>1.332</td>
</tr>
<tr>
<td>income over hp ratio</td>
<td>0.413</td>
<td>0.423</td>
<td>0.078</td>
<td>0.177</td>
<td>0.650</td>
</tr>
<tr>
<td>property tax rate (%)</td>
<td>1.091</td>
<td>1.021</td>
<td>0.048</td>
<td>0.248</td>
<td>2.603</td>
</tr>
<tr>
<td>Bachelor’s degree rate (%)</td>
<td>23.079</td>
<td>22.500</td>
<td>7.019</td>
<td>11.000</td>
<td>47.600</td>
</tr>
<tr>
<td>% of asian population</td>
<td>2.059</td>
<td>1.300</td>
<td>3.333</td>
<td>0.100</td>
<td>46.000</td>
</tr>
<tr>
<td>% of black population</td>
<td>11.036</td>
<td>6.900</td>
<td>11.162</td>
<td>0.200</td>
<td>51.000</td>
</tr>
<tr>
<td>Variables</td>
<td>Model I</td>
<td>Model II</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>-0.103 (-0.79)</td>
<td>-1.022 (-6.17)**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>9.448 (9.40)***</td>
<td>6.913 (7.59)***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>var</td>
<td>-22.332 (-3.47)**</td>
<td>-14.836 (-2.33)**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>skew</td>
<td>0.030 (1.76)*</td>
<td>0.070 (4.27)***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rent</td>
<td>0.242 (2.62)***</td>
<td>0.225 (2.14)**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>income/hp</td>
<td>2.227 (9.45)***</td>
<td>2.736 (11.42)***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>propertytax</td>
<td></td>
<td>-0.021 (-7.91)***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>education</td>
<td>-0.006 (-2.56)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>white</td>
<td>0.010 (5.63)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>black</td>
<td>0.003 (1.31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>asian</td>
<td>0.002 (0.48)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>273</td>
<td>273</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²-adjusted</td>
<td>0.63</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

dependent variable: homeownership rate, t statistics in parentheses
* significant at 10%, ** significant at 5%, *** significant at 1%
3.1 Introduction

Research on spatial correlation of housing price and housing price changes has been done at different aggregation levels.

Basu and Thibodeau (1998) examined spatial autocorrelation in transaction prices of single-family properties in Dallas, Texas. Using a semilog hedonic house price equation and a spherical autocorrelation function with data for over 5000 transactions, they found strong empirical evidence of spatial autocorrelation in transaction prices within submarkets.

Tirtiroglu (1992) and Clapp and Tirtiroglu (1994), using data of Hartford, CT, found that housing price changes in a submarket not only depend on its own lagged values, but also depend on the lagged values of price changes in neighboring submarkets.

Holly, Pesaran and Yamagata (2006) modeled the dynamic adjustment of real house prices using state level data. They examined the "role of spatial factors, in particular the effect of contiguous states by use of a weight matrix". Their empirical results gave them evidence of "significant spatial effect, even after controlling for State specific real incomes, and allowing for a number of unobserved common factors."
A direct motivation for this paper is the work by Pollakowski and Ray (1997). In that paper, the authors examined the interrelationship among housing price changes in (1) different U.S. census divisions and (2) different primary MSAs within a consolidated MSA. They found evidence of diffusion between contiguous areas at metropolitan area level, but when the same VAR analysis is conducted at census region level, different diffusion patterns are found. Particularly, census region housing price changes do not seem to follow a spatial diffusion process, in which a shock in one division first hits neighboring divisions and then diffuses from those divisions to their neighboring divisions, and so on. Instead, correlations between neighboring regions are not so significant. For example, Pacific region is more related with West South Central and East South Central, instead of related to the neighboring Mountain region. The authors argued that the spillover effects of an economic shock on housing market "would not necessarily occur between neighboring areas, but would be based on economic interrelationships".

The purpose of this paper is to apply recent advances in spatial econometric models of panel data, to examine the spillover mechanism in the US housing market at Metropolitan Statistical Area (MSA) level. In particular, we are interested in finding out whether housing returns are more highly correlated among economically similar cities (in various geographic locations) or among nearby cities. Methodologically, we use two different weight matrices\(^\text{25}\) in the spatial econometric model simultaneously. The first matrix is based on geographic dependence among MSAs and the second matrix is based on the economic similarities between MSAs. We control for fundamental changes such as real income and population dynamics as well as MSA specific individual effects. Our empirical regression results shows that both the inter-MSA spatial correlations in housing returns are significant

\(^{25}\) In spatial econometric literature, weight matrix is used to calibrate the degree of cross section dependence.
and positive, but with the geographic correlation appearing to be stronger and more significant than the correlation based on economic similarities. Furthermore, we also employ a spatial dynamic panel model to check for the spillover/diffusion mechanism with lagged housing returns. We found that housing prices changes in one MSA area are correlated not only with their own lagged changes, but also correlated with the present and lagged housing prices changes in their neighboring or similar MSAs.

This paper extends the existing literature in two ways:

1. Advanced econometric methodologies are applied in this paper to conduct the empirical investigation. A brand new method of data imputation specifically designed for spatial panel model is applied in this paper to tackle the unbalanced panel data problem we encountered in our empirical investigation. And, a newly developed spatial dynamic panel model with both time and individual effects is also employed in our empirical analysis.

2. In addition to the conventional distance-based weight matrix, this paper innovatively constructs a weight matrix based on the economic similarities between MSAs.

The paper proceeds as follows. Section 3.2 introduces in detail the MSA level panel data set and the construction of two weight matrices. Section 3.3 presents the spatial econometric models with two weight matrices, and the corresponding empirical results with chopped panel data. Section 3.4 develops econometric theory and methods for missing data imputation to deal with the unbalanced panel problem, and provides the estimation results. Section 3.5 uses a spatial dynamic panel model to test the hypothesis of informational inefficiency in the housing market. The last section concludes.
3.2 Data

3.2.1 Data for Real Housing Returns

The dependent variable, which is the real house price appreciation rate (or real housing returns) is calculated using Freddie Mac’s Conventional Mortgage House Price Index (CMHPI) data. The CMHPI is a weighted, repeat-sales index, meaning that it measures average price changes in repeat sales or refinancings on the same properties. This information is obtained by reviewing repeat mortgage transactions on single-family properties whose mortgages have been purchased or securitized by Fannie Mae or Freddie Mac since January 1975. The CMHPI is updated each quarter as additional mortgages are purchased or securitized by Fannie Mae and Freddie Mac. The new mortgage acquisitions are used to identify repeat transactions for the most recent quarter and for each quarter since the first quarter of 1975. The CMHPI data provides indexes for all nine Census Divisions, the 50 states and the District of Columbia, and all the 379 MSAs in the U.S.. Based on CMHPI, we can compute MSA specific annual housing returns on a quarterly basis\textsuperscript{26}, from the first quarter of 1976 to the present. Furthermore, the original CMHPI is not adjusted for inflation, so we used the CPI obtained from Federal Reserve Bank of St. Louis to adjust the house price index for inflation.

However, one major limitation of the CMHPI data set is that it is an unbalanced panel. Among all the 379 MSAs, only 55 of them (mostly big major MSAs) have complete observations from the first quarter of 1975 to present. All the other MSAs start to have observations at some date later than 1975, with the first available date varying largely among

\textsuperscript{26}Housing return, refers to the year over year quarterly increase rate of house price. More specifically, to adjust for seasonality in housing prices, it is calculated using the formula: \((P_t / P_{t-4}) - 1\), where \(P_t\) is the real housing price index for current quarter, while \(P_{t-4}\) is the real housing price index for the same quarter of one year ago.
different MSAs. For example, El Paso, TX MSA has data available since the first quarter of 1979, while Yuma, AZ MSA doesn’t have any observations available until the first quarter of 1990. To deal with this missing data problem, we will first adopt the most conventional solution. That is, we will do the empirical investigation using a balanced panel data set, which is a subset of the original data set. The chopped data set will cover 158 MSAs from the first quarter of 1984 to the last quarter of 2006. Later, we will use the newly developed data imputation techniques for spatial econometric models to tackle the unbalanced panel problem caused by missing data issues. We then will have a much larger panel which covers almost all the MSAs of the U.S. (354 MSAs out of 379) with the time horizon from the first quarter of 1976 to the last quarter of 2006.

3.2.2 Data for Explanatory Variables

In our empirical regressions, we also control for fundamental changes such as real income and population dynamics.

As has been widely recognized in the literature, changes in income and population are the two major fundamental demand side factors for housing market dynamics. For example, supporting evidences can be found in recent works of Capozza, Hendershott and Mack (2004), Gallin (2006), and Wheaton and Nechayev (2007). Following the literature, here we include in our regression models the year-over-year quarterly growth rates of real disposable income\(^{27}\) and population for each MSA.

The quarterly data used to calculate changes in MSA total disposable income is obtained from Bureau of Economic Analysis (BEA) via Moody’s Economy.com, as is the data for population changes. Again, the income data is not adjusted for inflation, so we use CPI to deflate it in order to obtain the growth rates of real income.

\(^{27}\)Disposable income is gross income minus income tax on that income.
3.2.3 Weight Matrices

We use the first weight matrix to calibrate the geographic cross-sectional dependence among MSAs. In particular, the weight matrix is constructed so that the \((i, j)\) element \(w_{ij}\) of the matrix takes a value of 1 if the \(i\)th and \(j\)th MSAs are contiguous (share borders) and zero otherwise. The diagonal elements of the weight matrix are set equal to zero in order that no observation of real housing returns predicts itself. In the literature, there are many choices for specifying the weight matrix based on geographic information. For example, we can let \(w_{ij} = \frac{d_{\text{min}}}{d_{i,j}} \forall i \neq j\), where \(d_{\text{min}}\) is the distance between two closest MSAs, and \(d_{i,j}\) denotes the distance between MSA \(i\) and MSA \(j\). Or, we can let \(w_{ij} = e^{-d_{i,j}/1000} \forall i \neq j\), which is also a decreasing function in \(d_{i,j}\).

However, according to Militino et al. (2004), different specifications of weight matrices only show small practical differences, so we decide to use the most widely-used zero-one weight matrix to capture the geographic cross-sectional dependence among MSAs.

One of the major tasks of this paper is to construct a second weight matrix to capture the economic similarities between MSAs, given the fact that barely no preceding work on this has been done in the literature. We construct this matrix by using the "GDP by MSA" data from the BEA. BEA provides MSA level gross domestic product data decomposed into 20 different industries such as agriculture, mining, construction, manufacturing, education, information, etc. from 2001 to 2006.\(^{28}\) Using this information, we can calculate the average share of each industrial component in local GDP for each MSA. We then compute the correlations between any two MSAs to measure the economic similarities of these paired MSAs. Accordingly, the \((i, j)\) element of this weight matrix is the Pearson correlation based on the MSA level GDP decompositions between the \(i\)th and \(j\)th MSAs.

\(^{28}\)Please refer to the appendix for detailed decomposition.
Take New York MSA as an example, San Francisco, CA is found to be the MSA which is most similar to the New York MSA economically (correlation=0.926), followed by Boston, MA (correlation=0.908), Salt Lake City, UT (correlation=0.905), Bridgeport-Stamford, CT (correlation=0.897) and Los Angeles, CA (correlation=0.860). Among these five MSAs, we can see that only Bridgeport-Stamford, CT is neighboring with New York, the others are not. Instead, three of them are from the far west. Furthermore, the diagonal elements of the weight matrix are again set equal to zero so that no real housing returns predicts itself.

3.3 Spatial Econometric Models and Empirical Results with Chopped Panel Data

As we’ve mentioned, the real housing returns panel has an unbalanced pattern. So we first conduct the empirical investigation using a balanced panel data set, which is a subset of the original data set. The chopped data set consists of 158 MSAs from the first quarter of 1984 to the last quarter of 2006 (92 quarters in total). In the next section, we will expand the data set to include virtually all the MSAs, using the newly developed data imputation method by Wang (2009).

3.3.1 The Fixed Effects Model with Two Weight Matrices

The model

We put two weight matrices into the spatial autoregression model at the same time, in order to find out whether housing returns are more highly correlated among economically similar cities (in various geographic locations) or among nearby cities. In particular, at a given time $t$, the model specification is as follows:

$$
Y_{nt} = \lambda_1 W_{1n} Y_{nt} + \lambda_2 W_{2n} Y_{nt} + X_{nt} \beta + c_n + V_{nt},
$$

(3.1)
where $t = 1, 2, ..., 92$. The dependent variable $Y_{nt} = (y_{1t}, y_{2t}, ..., y_{it}, ..., y_{nt})'$ is a $n$ by 1 column vector, with $n = 158$. So it is the real housing returns for 158 MSAs at quarter $t$. $W_{1n}$ and $W_{2n}$ are 158 by 158 spatial weight matrices which are nonstochastic and generate the spatial dependence between cross sectional units $y_{it}$. More specifically, $W_{1n}$ is the weight matrix constructed based on geographic interdependence, and $W_{2n}$ is the weight matrix which measures the economic similarities between MSAs. $\lambda_1$ and $\lambda_2$ are the spatial autoregressive coefficients to be estimated which capture the spatial correlation of housing returns among MSAs, geographically and economically, respectively. Values of $\lambda_1$ and $\lambda_2$ should lie between $-1$ and $1$. $X_{nt}$ is an 158 by $k$ matrix of nonstochastic regressors, which, in our regression, $k$ is equal to 2, with the regressors being the quarterly income growth rate and population growth rate. $\beta$ is a $k$ by 1 column vector of parameters to be estimated. $c_n$ here is 158 by 1 column vector of individual effects, which does not vary over time. For example, it could probably capture the effect of city specific land use regulations on the housing returns dynamics.\(^{29}\) Other factors which are MSA specific and do not change over time may also be accommodated in $c_n$. As for now, we consider $c_n$ as fixed effects. Finally, $V_{nt} = (v_{1t}, v_{2t}, ..., v_{it}, ..., v_{nt})'$ is a 158 by 1 column vector of i.i.d. error terms. To give a more straightforward picture of our model specification, a typical observation of real housing returns $y_{it}$ for MSA $i$ and time $t$ is modeled as follows:

$$y_{it} = \lambda_1 \sum_{j=1}^{158} w_{1,ij} y_{jt} + \lambda_2 \sum_{j=1}^{158} w_{2,ij} y_{jt} + \beta_1 x_{1it} + \beta_2 x_{2it} + c_i + v_{it},$$ (3.2)

\(^{29}\)It has been recognized in the recent literature that in addition to the demand side factors such as income and population movements, the elasticity of housing supply also plays a very important role in determining the dynamics of housing return. And the land use regulation level is found to be very closely related to the elasticity of housing supply. For example, both theoretical and empirical evidences can be found in works of Malpezzi (1996, 1998), Glaeser, Gyourko and Saks (2005), McMillen and McDonald (1993,1998,1999,2002).
where $w_{1,ij}$ is the $(i, j)$ element of $W_1$, and $w_{2,ij}$ is the $(i, j)$ element of $W_2$. $x_{1it}$ is the growth rate of real disposable income, and $x_{2it}$ is the growth rate of population. $\beta_1$ and $\beta_2$ are parameters to be estimated, $c_i$ represents the MSA specific individual fixed effect, and $v_{it}$ is an i.i.d. error term.

The spatial autoregressive coefficients $\lambda_1$ and $\lambda_2$ are of our main interests. $\lambda_1$ measures the effects of all the other MSAs (weighted by the geographic distance) on the housing price dynamics of MSA $i$. And $\lambda_2$ captures the effects of all the other MSAs (weighted by the economic similarities) on the housing price dynamics of MSA $i$. A positive and significant $\lambda_1$ would imply that there is strong evidence of spatial dependence between contiguous MSAs in the U.S. housing market. That is, the housing returns of neighboring MSAs at the same time period have a positive effect on the housing returns of MSA $i$. On the other hand, a positive and significant $\lambda_2$ shows empirical evidence that spillover effects on housing market would probably be based on economic interrelationships. MSAs that are economically similar to MSA $i$ tend to have similar movements of housing prices as MSA $i$ has. By putting these two weight matrices together into the same regression model, we can then find out which interrelationship is more significant, by comparing the estimated $\lambda_1$ and $\lambda_2$.

**Estimation Methods**

To eliminate the individual fixed effects, we apply the within transformations on this model. Let $\tilde{Y}_{nt} = Y_{nt} - \bar{Y}_{nT}$, $\tilde{X}_{nt} = X_{nt} - \bar{X}_{nT}$, and $\tilde{\varepsilon}_{nt} = \varepsilon_{nt} - \bar{\varepsilon}_{nT}$, where $\bar{Y}_{nT} = \frac{1}{T} \sum_t Y_{nt}$, $\bar{X}_{nT} = \frac{1}{T} \sum_t X_{nt}$, and $\bar{\varepsilon}_{nT} = \frac{1}{T} \sum_t \varepsilon_{nt}$. The within transformation ignoring the individual fixed effects gives

$$
\tilde{Y}_{nt} = \lambda_1 W_{1n} \tilde{Y}_{nt} + \lambda_2 W_{2n} \tilde{Y}_{nt} + \tilde{X}_{nt} \beta + \tilde{\varepsilon}_{nt}.
$$

(3.3)
As $W_{1n} \tilde{Y}_{nt}$ and $W_{2n} \tilde{Y}_{nt}$ are endogenous, we estimate the transformed model by the 2SLS method with the IV matrix $Q_{nt} = [W_{1n} \tilde{X}_{nt}, W_{2n} \tilde{X}_{nt}, \tilde{X}_{nt}]$. Let $\delta = (\lambda_1', \lambda_2', \beta')'$, $\tilde{\epsilon}_{nt}(\delta) = \tilde{Y}_{nt} - \lambda_1 W_{1n} \tilde{Y}_{nt} - \lambda_2 W_{2n} \tilde{Y}_{nt} - \tilde{X}_{nt} \beta$, $Z_{nt} = [W_{1n} \tilde{Y}_{nt}, W_{2n} \tilde{Y}_{nt}, \tilde{X}_{nt}]$, and $P_{nt} = Q_{nt} (Q_{nt}' Q_{nt})^{-1} Q_{nt}'$. The 2SLS estimator is $\hat{\delta}_{2sls,n} = (\sum_t Z_{nt}' P_{nt} Z_{nt})^{-1} \sum_t Z_{nt}' P_{nt} \tilde{y}_{nt}$.

**Estimation Results**

The regression results are presented in table (3.2) in the appendix.

$\lambda_1$, the spatial autoregressive coefficient measuring the spatial correlation of housing returns among MSAs geographically, is estimated to be highly significant and equal to 0.8567, which is positive and close to 1. This result implies that controlling for other factors such as income growth, population growth, individual effects, as well as inter-MSA dependence based on economic similarities, there is still quite strong evidence that housing returns are closely and positively related among neighboring MSAs. So, neighboring cities tend to move in the same direction in terms of housing prices changes.

$\lambda_2$, the spatial autoregressive coefficient measuring the spatial correlation of housing returns among MSAs based on economic similarities, is also very significant, positive and equal to 0.3527. This gives evidences of positive correlation in housing price dynamics among economically similar MSAs, with other factors being controlled.

To make a comparison about which effect is more significant, we look at the estimates of $\lambda_1$ and $\lambda_2$. In terms of the values of estimates, $\lambda_1$ is more than double $\lambda_2$, which might imply that the geographic interrelationship appear to be stronger than the economic interrelationship. In terms of the t-test statistics, the t-ratio corresponding to estimated $\lambda_1$ is also much larger (40.7 vs. 5.1), which suggests that the spatial autoregressive coefficient based on geographic distances is more significant than the coefficient based on economic similarities.
Furthermore, the coefficients for both income growth and population growth are positive and significant. Therefore, controlling for MSA individual effects together with spatial autocorrelations, income growth rate and population growth rate are found to be significantly and positively related with the housing price changes. This is intuitive given the fact that income and population are considered to be the major demand side factors for housing prices, and this result is also consistent with the findings of previous researches.

To summarize, using the spatial econometric model with two weight matrices, we found that both the inter-MSA spatial correlations are significant and positive, but with the geographic correlation appearing to be stronger and more significant than the correlation based on economic similarities. R-square is 0.72, which implies that about 72% of the variation in housing returns across different MSAs and different time periods are explained by our regression.

3.3.2 The Random Effect Model with Two Weight Matrices

For robustness purpose, we now estimate model (3.1) as a random effect model, that is, we now consider the individual effect $c_n$ in model (3.1) as a random effect variable and re-estimate the model.

**Estimation Methods**

Again, we estimate the model by the 2SLS method. However, here we use different IV matrix for the random effect model. Following Kelejian and Prucha (1998), the IV matrix we employ here is:

$$Q_{nt} = [X_{1,nt}, W_{1n}X_{1,nt}, W_{2n}X_{1,nt}, W_{1n}^2X_{1,nt}, W_{2n}^2X_{1,nt}].$$
Note that for the fixed effect model, we use both income growth and population growth as our instrumental variables, but for the random effect model, we only used $X_{1,nt}$, the income growth variable. We believe that income growth rates rather than population growth rates are more likely to be uncorrelated with the individual effects which might capture the effects of city specific land use regulations. Intuitively, population changes and the level of land use regulation/growth controls should be correlated. As showed by Glaeser, Gyourko and Saks (2005), in cities with more restrictive regulation on new land development, a positive demand side shock usually leads to a higher housing price, rather than a bigger population or a larger city. On the other hand, in cities where the land use regulation level is low, it usually results in more population instead of higher housing prices. So regulation level, which could be captured by the individual effect $c_n$, is very likely to be correlated with the regressor of population growth rate. We thus use $X_{1,nt}$, the income growth rate as our instrumental variable, because there is no evidence of significant correlation between land use regulation and income dynamics.

Furthermore, we let $\delta = (\lambda_1', \lambda_2', \beta')', Z_{nt} = [W_{1n}Y_{nt}, W_{2n}Y_{nt}, X_{nt}]$, and $P_{nt} = Q_{nt}(Q_{nt}'Q_{nt})^{-1}Q_{nt}'$. The 2SLS estimator is given by

$$\hat{\delta}_{2sls,n} = (\sum_t Z_{nt}'P_{nt}Z_{nt}^*)^{-1} \sum_t Z_{nt}'P_{nt}Y_{nt}.$$

**Estimation Results**

Table (3.3) presents the estimation results for the random effect model. Compared with the results of the fixed effect model, no significance differences are noticed. Estimated $\lambda_1$ is still larger than $\lambda_2$ in magnitude, and also in terms of t-ratios. Both income growth rate and populations rate are significant and positively related to the housing price changes. R-squares are also very close.
So, again, we found that both the inter-MSA spatial correlations are significant and positive, but the geographic correlation appears to be stronger and more significant than the correlation based on economic similarities.

3.4 Spatial Econometric Models and Empirical Results with Larger but Unbalanced Panel Data

3.4.1 Econometric Theory and Methodology of Missing Data Imputation for Spatial Panel Models

In this section, we use new developed econometric theory of missing data imputation for spatial panel model, in order to tackle the problem of unbalanced housing returns data. More specifically, one limitation of the CMHPI data from Freddie Mac is that the pattern of the panel is unbalanced. That is, among all the 379 MSAs, only 55 of them (mostly big major MSAs) have complete observations from the first quarter of 1975 to present. All the other MSAs start to have observations at some date later than 1975, with the first available date varying largely among different MSAs.

Instead of chopping the data into a smaller but balanced subset, we have an alternative option, which is to use some econometric treatment methods to deal with the missing data problem so that we can be more efficient by making use of all the observations we have. Treatment methods for unbalanced/incomplete panel data have been well explored in the literature. Methods such as maximum likelihood (ML), restricted maximum likelihood (REML), analysis of variance (ANOVA), multiple imputations (MI), minimum norm quadratic unbiased estimation (MINQUE), and minimum variance quadratic unbiased estimation (MIVQUE)\textsuperscript{30} are introduced to solve the missing data problems. However, literature on dealing with unbalanced/incomplete panel data problems for spatial models

\textsuperscript{30}Please refer to Baltagi and Chang (1994) and Allison (2001) for details on these methods.
is virtually blank given the fact that the commonly used methods in the literature tend to be either over complicated or inapplicable for spatial models. For example, the often used method of MLE becomes quite challenging under the framework of spatial models, according to Lee (2004). Another widely adopted method of MI appears to be virtually inapplicable for this case.

Note that we only have missing data problem with the data set for real housing returns, which are the dependent variables in our econometric model setting. We have complete observations on all the other variables. An imputation method that uses the expected values of those missing data could be a feasible solution. Wang (2009) developed theory and methodology to deal with the missing data problems for spatial econometric models using GMM methods, specifically for data with missing observations only in dependent variables. We can apply this method here to tackle the unbalanced panel problem.

One issue that remains is on the individual effects. As we can see in section 3.3, we treat the individual effects as fixed effects and then also as random effects for robustness purpose. However, here we will only use random effects model. The reason, as stated by Wang (2009), is that when we treat the individual effects as fixed effects, the estimation procedure requires a data transformation to eliminate the fixed effects. But with missing observations in the data set, the data transformation procedure will mix up the observed and unobserved data which will cause problems in deriving the explicit form of the expected values of the unobserved data. Therefore, Wang (2009) propose a setting with the individual effects treated as random effects. Although random effects models are not frequently used in practice due to the possible correlations between the random effects and the explanatory variables, it works as long as we can find IVs.

A brief introduction of the GMM estimation methods is provided below.
Consider the following spatial panel model with $p$-order spatial lags ($T$ periods, $t = 1, ..., T$ and each period $n$ individuals $i = 1, ..., n$):

$$Y_{nt} = \sum_{j=1}^{p} \lambda_{0j} W_{jn} Y_{nt} + X_{nt} \beta_0 + \tilde{\xi}_n + V_{nt},$$

where $Y_{nt}$ is a $n \times 1$ vector of outcomes of the $n$ individuals at time $t$. $X_{nt}$ is an $n \times k$ matrix of exogenous variables representing the $n$ individuals’ exogenous characters at time $t$. $\tilde{\xi}_n$ is the $n \times 1$ vector of random individual effects, $\tilde{\xi}_i \sim i.i.d. (0, \sigma_{\tilde{\xi}}^2)$. $V_{nt} = (v_{1t}, ..., v_{nt})'$ with $v_{it} \sim i.i.d. (0, \sigma_v^2)$. And, $W_{jn}$’s are $p$ distinct $n \times n$ weight matrices which are assumed to be time invariant.

By stacking the model into a big vector, we have

$$y_{nT} = \sum_{j=1}^{p} \lambda_{0j} w_{nT} y_{nT} + x_{nT} \beta_0 + \epsilon_{nT},$$

where $y_{nT} = (Y'_{n1}, ..., Y'_{nT})'$, $w_{nT} = I_T \otimes W_{jn}$, $x_{nT} = (X'_{n1}, ..., X'_{nT})'$, $\epsilon_{nT} = [(\tilde{\xi}_n + V_{n1})', ..., (\tilde{\xi}_n + V_{nT})]'$.

So we might just regard that as a SAR (Spatial Autoregression model but with random component structure in the overall disturbances $\epsilon$.

We consider the unbalanced panel case where some of the observations in the outcome vectors are unavailable. Suppose $n'$ is the number of unobserved elements in $y_{nT}$. To simplify notation, we introduce a selection matrix $P_N$ which picks out the unobserved elements of $y_{nT}$ such that $Y_N = P_N y_{nT}$, where $N = nT$ and $Y_N = \left( \begin{array}{c} Y_{N}^{(o)} \\ Y_N^* \end{array} \right)$, where $Y_{N}^{(o)}$ is the $(N - n')$-dimensional observed subvector, but $Y_N^*$ is the remaining $n'$-dimensional unobserved subvector.

Therefore,

$$P_N y_{nT} = \sum_{j=1}^{p} \lambda_{0j} P_N w_{nT} y_{nT} + P_N x_{nT} \beta_0 + P_N \epsilon_{nT},$$

86
So we have $X_N = P_N x_n T$, $W_{j:N} = P_N w_{jnT} = P_N [I_T \otimes W_{jn}]$ and $\epsilon_N = P_N \epsilon_{nT}$

The model is then simplified to,

$$Y_N = \sum_{j=1}^{p} \lambda_{0j} W_{j:N} Y_N + X_N \beta_0 + \epsilon_N \tag{3.7}$$

with $Y_N = \begin{pmatrix} Y_{N}^{(o)} \\ Y_{N}^{*} \end{pmatrix}$.

So the system for estimation is

$$\begin{pmatrix} Y_{N}^{(o)} \\ Y_{N}^{*} \end{pmatrix} = \sum_{j=1}^{p} \lambda_{0j} W_{j:N} \begin{pmatrix} Y_{N}^{(o)} \\ Y_{N}^{*} \end{pmatrix} + X_N \beta_0 + \epsilon_N \tag{3.8}$$

Let $J_N = [0_{n' \times (N-n')}, I_{n' \times n'}]$. Then $Y_{N}^{*} = J_N Y_N$.

The reduced form equation implies

$$Y_{N}^{*} = J_N Y_N = J_N (I_N - \sum_{j=1}^{p} \lambda_{0j} W_{j:N})^{-1} (X_N \beta_0 + \epsilon_N) \tag{3.9}$$

and the expectation of $Y_{N}^{*}$ is:

$$E(Y_{N}^{*}) = J_N (I_N - \sum_{j=1}^{p} \lambda_{0j} W_{j:N})^{-1} X_N \beta_0 \tag{3.10}$$

Denote $F_N(\theta_0) = E(Y_{N}^{*})$ which is a function of $X_N$, $W_N$ and unknown parameter $\theta_0$, where $\theta_0 = (\lambda'_0, \beta'_0)'$ and $\lambda_0 = (\lambda_{01}, \ldots, \lambda_{0p})'$.

These imply

$$\begin{pmatrix} Y_{N}^{(o)} \\ F_N(\theta_0) \end{pmatrix} = \sum_{j=1}^{p} \lambda_{0j} W_{j:N} \begin{pmatrix} Y_{N}^{(o)} \\ F_N(\theta_0) \end{pmatrix} + X_N \beta_0 + U_N \tag{3.11}$$

where

$$U_N = \epsilon_N - (I_N - \sum_{j=1}^{p} \lambda_{0j} W_{j:N}) \begin{pmatrix} 0 \\ Y_{N}^{*} - F_N(\theta_0) \end{pmatrix} \tag{3.12}$$

$$= \epsilon_N - (I_N - \sum_{j=1}^{p} \lambda_{0j} W_{j:N}) J_N J_N (I_N - \sum_{j=1}^{p} \lambda_{0j} W_{j:N})^{-1} \epsilon_N$$

$$= (I_N - \sum_{j=1}^{p} \lambda_{0j} W_{j:N}) J_N^{(o)} J_N^{(o)} (I_N - \sum_{j=1}^{p} \lambda_{0j} W_{j:N})^{-1} \epsilon_N$$

87
where $J_N^{(o)}$ is the corresponding selection matrix to $Y_N^{(o)}$, $J_N^{(o)} = [I_{(N-n') \times (N-n')}, 0_{(N-n') \times n'}]$. The last equality comes from the fact that $J_N^{(o)'} J_N^{(o)} + J_N' J_N = I_N$.

Note that $E(U_N) = 0$ even though its variance matrix is complicated. We can use linear moments to do the GMM of this model.

Let $Q_N$ be an $N \times k_x$ IV matrix, and denote $\theta = (\lambda', \beta')'$, where $\lambda = (\lambda_1, ..., \lambda_p)'$ and

$$U_N(\theta) = (I_N - \sum_{j=1}^p \lambda_j W_j N) \left( \begin{array}{c} Y_N^{(o)} \\ F_N(\theta) \end{array} \right) - X_N \beta$$

Suppose $a_N$ is a constant full rank matrix, and let $g_N(\theta) = Q_N' U_N(\theta)$, then under some regularity conditions, the GMM estimator $\hat{\theta}_N$ derived from $\min_{\theta \in \Theta} g_N(\theta) a_N' a_N g_N(\theta)$ is consistent and asymptotically normal. The best choice of the weighting matrix $a_N' a_N$ involves the variance-covariance matrix of the overall disturbances, which can also be estimated consistently.\(^{31}\)

In the next subsection, we will apply this GMM estimation to our data set.

### 3.4.2 The Spatial Econometric Model with Two Weight Matrices and Missing Data

We use a model specification similar to that in section 3.3.1, but with a much larger data set but some missing observations. By setting $p = 2$ in model (3.4), we have the model as follows:

$$Y_{nt} = \lambda_1 W_{1n} Y_{nt} + \lambda_2 W_{2n} Y_{nt} + X_{nt} \beta + \xi_n + \zeta_n + V_{nt},$$

where $t = 1, 2, ..., 124$. The dependent variable $Y_{nt} = (y_{1t}, y_{2t}, ..., y_{it}, ..., y_{nt})'$ is a $n$ by 1 column vector, with $n = 354$. In each $y_{it}$, there are some randomly missing observations. $W_{1n}$ is the spatial weight matrix constructed based on geographic interdependence, and

\(^{31}\)Please refer to Wang (2009)
$W_{2n}$ is the 354 by 354 weight matrix which measures the economic similarities between MSAs. $\lambda_1$ and $\lambda_2$ are the spatial autoregressive coefficients to be estimated which capture the spatial correlation of housing returns among MSAs, geographically and economically, respectively. Values of $\lambda_1$ and $\lambda_2$ should lie between $-1$ and $1$. $X_{nt}$ is an 354 by $k$ matrix of nonstochastic regressors, which, in our regression, $k$ is equal to 2, with the regressors being the quarterly income growth rate and population growth rate. $\beta$ is a $k$ by 1 column vector of parameters to be estimated. $\xi_n$ is 354 by 1 column vector of individual random effects, which does not vary over time. Finally, $V_{nt} = (v_{1t}, v_{2t}, ..., v_{it}, ..., v_{nt})'$ is a 354 by 1 column vector of i.i.d. error terms.

For estimation, we apply the method developed in the previous subsection. For the IV selection, we use $[X_{1,nt}, W_{1n}X_{1,nt}, W_{2n}X_{1,nt}, W_{1n}^2X_{1,nt}, W_{2n}^2X_{1,nt}]$ as we did in the previous section for the random effect model.

### 3.4.3 Estimation Results

The regression results are presented in table (3.4) in the appendix.

$\lambda_1$, the spatial autoregressive coefficient measuring the spatial correlation of housing returns among MSAs geographically, is estimated to be highly significant and positive. This result again implies that controlling for other factors such as income growth, population growth, individual effect, as well as inter-MSA dependence based on economic similarities, there is strong evidence that housing returns are closely and positively related among neighboring MSAs.

$\lambda_2$, the spatial autoregressive coefficient measuring the spatial correlation of housing returns among MSAs based on economic similarities, is also positive and significant. This
gives evidences of positive correlation in housing price dynamics among economically similar MSAs, with other factors being controlled.

Again, we can make a comparison about which effect is more significant by looking at the estimates of $\lambda_1$ and $\lambda_2$. In terms of the values of estimates, $\lambda_1$ is again larger than $\lambda_2$, which might imply that the geographic interrelationship appear to be stronger than the economic interrelationship. In terms of the t-test statistics, the t-ratio corresponding to estimated $\lambda_1$ is also much larger (81.2 vs. 26.9), which suggests that the spatial autoregressive coefficient based on geographic distances is more significant than the coefficient based on economic similarities, when putting them together into a spatial panel regression model.

Furthermore, both income and population turn out to be positively and significantly related to the housing price growth rate.

Compared with the results we got from model (3.1) which used chopped data, we can see that the results are very similar. For both data sets, both $\lambda_1$ and $\lambda_2$ are positive and significant, but $\lambda_1$ consistently appear to be larger and more significant than $\lambda_2$. So the regression results using larger data set again suggest that, by controlling for changes in fundamental factors and MSA-specific factors, both the inter-MSA spatial correlations in housing returns are significant and positive, but with the geographic correlation appearing to be stronger and more significant than the correlation based on economic similarities.

### 3.5 The Spatial Dynamic Panel Model with One Weight Matrix

Previous literature has provided evidences for the argument that housing market is informationally inefficient, which implies that housing prices do not instantaneously reflect all available information. Many paper can be found in the literature which show that
past/lagged price changes could be useful predictors for future price dynamics. Furthermore, evidence also has been presented which supported the fact that housing price changes in one place can be predicted not only by their own lagged changes, but also by housing prices movements in other locations. This research has been conducted by Pollakowski and Ray (1997) at both Census Division level and PMSA level. Clapp and Tirtiroglu (1994) had similar findings using data from submarkets in Hartford, CT. They found that housing price changes in a submarket is correlated not only with their own lagged changes, but also with the past values of price changes in contiguous submarkets.

Here, we can use a newly developed spatial dynamic panel model to test this spillover mechanism at the MSA level, using the chopped panel data set. The ideal way of model specification is to add lagged housing returns directly to model (3.1) and compare our estimation results with what we found in the previous section. However, according to Lee and Yu (2009), econometric theory corresponding to this particular model (with two weight matrices and also lagged dependant variables) has not been developed yet. As for now, the best and latest model we can employ is the spatial dynamic panel model with only one weight matrix at a time, based on the econometric theory developed by Lee and Yu (2009).

Although most of the related studies that have been done in the literature focussed on the spillover/diffusion mechanism based on geographic interrelationships between locations, it is also recognized that the diffusion can happen along a number of different dimensions, including political jurisdiction, demographic factors and economic similarities. So here, we check the spillover pattern using both weight matrices, but with only one weight matrix each time.

The Model

We first look at the case using the distance based weight matrix. The model specification is as follows:

\[ Y_{nt} = \lambda_1 W_1 Y_{nt} + \gamma Y_{n,t-1} + \rho_1 W_1 Y_{n,t-1} + X_{nt} \beta + c_n + \alpha_t l_n + V_{nt}, \quad (3.15) \]

where \( Y_{n,t-1} \) is the real housing returns of one quarter earlier, \( \gamma \) is the parameter to be estimated, which measures the effect of own lagged housing returns on the housing price changes of the current period. \( \rho_1 \) is the spatial autoregressive coefficient which captures the effect of the lagged housing returns of neighboring MSAs. \( \alpha_i \) is a scalar of time effect and \( l_n \) is a column vector of ones. The time effect can capture the effects of those factors that do not vary across MSAs, for example, the mortgage interest rates. \( \lambda_1, \beta, c_n, \) and \( V_{nt} \) have the same interpretations as they do in model (3.1). Furthermore, the length of lag is set to be 1 in the model, constrained by the latest advances in related econometric theory.

The model using weight matrix based on economic similarities is very similar:

\[ Y_{nt} = \lambda_2 W_2 Y_{nt} + \gamma Y_{n,t-1} + \rho_2 W_2 Y_{n,t-1} + X_{nt} \beta + c_n + \alpha_t l_n + V_{nt}. \quad (3.16) \]

So basically, we just replace \( W_1 \) by \( W_2 \) and use the same methodology to estimate all the parameters.

Estimation Methods

With both time and individual fixed effects, Lee and Yu (2009) have shown that the direct MLE method will yield a bias of order \( O(\max(1/n, 1/T)) \) for \( \theta = (\gamma, \rho, \beta', \lambda, \sigma^2)' \), where \( \sigma^2 \) is the variance for i.i.d. error term \( V_{nt} \). To avoid the bias of order \( O(1/n) \), they
have proposed a transformation approach that estimates the within model ignoring the time effects by the MLE. By premultiplication of $J_n = I_{n} - \frac{1}{n}l_n l_n'$, the transformed model is

$$J_n Y_{nt} = \lambda J_n W_n J_n Y_{nt} + \gamma J_n Y_{n,t-1} + \rho J_n W_n J_n Y_{n,t-1} + J_n X_{nt} \beta + J_n c_n + J_n V_{nt}. \quad (3.17)$$

Let $(F_{n,n-1}, l_n/\sqrt{n})$ be the orthonormal matrix of eigenvectors of $J_n$ where $F_{n,n-1}$ corresponds to the eigenvalue of one. As $F_{n,n-1}' J_n = F_{n,n-1}'$ and $F_{n,n-1} F_{n,n-1}' = J_n$, premultiplication of (3.17) by $F_{n,n-1}'$ gives

$$Y_n^* = \lambda W_n^* Y_{nt} + \gamma Y_{n,t-1}^* + \rho W_n^* Y_{n,t-1}^* + X_{nt}^* \beta + c_n^* + V_{nt}^*, \quad (3.18)$$

where $Y_{n,t}^* = F_{n,n-1}' Y_{nt}$, $X_{nt}^* = F_{n,n-1}' X_{nt}$, $c_n^* = F_{n,n-1}' c_n$, $V_{n,t}^* = F_{n,n-1}' V_{nt}$, and $W_n^* = F_{n,n-1}' W_n F_{n,n-1}$. Suppose that $V_{nt}$ is normally distributed $N(0, \sigma_0^2 I_n)$. The log likelihood function of the transformed model is

$$\ln L(\theta, c_n^*) = -\frac{(n-1)T}{2} \ln 2\pi - \frac{(n-1)T}{2} \ln \sigma^2 + T \ln |I_n - \lambda W_n^*| - \frac{1}{2\sigma^2} \sum_{t=1}^{T} V_{n,t}^*(-\theta) V_{n,t}^*(\theta),$$

where $V_{n,t}^*(\theta) = (I_{n-1} - \lambda W_n^*) Y_{n,t}^* - \gamma Y_{n,t-1}^* - \rho W_n^* Y_{n,t-1}^* - X_{nt}^* \beta - c_n^*$. As shown in Lee and Yu (2009), when $T$ is relatively larger than $n$, the ML estimator is consistent and asymptotically normal; when $T$ is proportional to $n$, the estimator is consistent and asymptotically normal, but the limit distribution is not centered around zero; when $T$ smaller than $n$, the estimator is consistent and has a degenerated limit distribution. We correct the asymptotic bias of the estimator by the bias-adjustment procedure in Lee and Yu (2009).

**Estimation Results**

The regression results are presented in table (3.5) in the appendix.

We get very similar regression results from the model with distance based weight matrix and the model with economic similarity based weight matrix. For both models, we obtain
positive and significant estimates for all the parameters. This implies that the MSA level real housing returns are significantly and positively related not only to their own lagged changes (estimates of $\gamma$ are positive and significant), but also to the present and lagged housing prices changes in their neighboring or economically similar MSAs (we get positive and significant estimates of $\lambda_1 & \rho_1$, $\lambda_2 & \rho_2$).

Income growth is found to be significantly and positively related to the housing prices changes, for both cases. This result complies with findings from previous researches, and it is also quite intuitive given the fact that changes in income is one of the major fundamental demand side factors for housing price dynamics.

Population growth also appears to be significantly and positively related to the housing prices changes, this is consistent with the result we got from the static model with two weight matrices.

3.6 Conclusion

Previous research argued that spillover effects of an economic shock on the housing market would not necessarily occur between neighboring areas, but would be based on economic interrelationships. This paper uses spatial panel models to investigate the spillover mechanism in the housing market. In particular, we are interested in finding out whether housing returns are more highly correlated among economically similar cities (in various geographic locations) or among nearby Cities.

Two different weight matrices are constructed to measure the geographic dependence and economic correlation respectively. Advanced econometric theories and methods are applied in our empirical work. More specifically, spatial panel models with two weight matrices are used to find out in which dimension are MSA level housing returns more
correlated. A newly developed spatial dynamic panel model with both time and individual effect is employed to check the effects of lagged housing returns. And additionally, a brand new econometric theory of missing data imputation specifically for spatial econometric models is applied in our paper to tackle the unbalanced panel problem with our data.

Our empirical regression results showed that both the inter-MSA spatial correlations in housing returns are significant and positive, but with the geographic correlation appearing to be stronger and more significant than the correlation based on economic similarities. Regression results from the spatial dynamic panel model provided evidence that housing prices changes in one MSA area are correlated not only to their own lagged changes, but also correlated to the present and lagged housing prices changes in their neighboring or similar MSAs. This result supports the hypothesis of informational inefficiency in the housing market.

A more ideal model set up for this research question is to have a spatial dynamic panel model with multiple weight matrices and more flexible choice of lags. However, econometric theories corresponding to this model specification are not available in the literature yet. This may be one potential topic for future research. Improvements on the construction of weight matrices can also be made in the future. For example, the geographic weight matrix based on distances could be constructed using the actual geographic distances between MSAs, so that we can get a weight matrix with continuous elements, which is more comparable with the economic similarity weight matrix than the current zero-one matrix. We can then re-run the regressions and check the robustness of our empirical results.
3.7 Tables

To calculate the weight matrix for economic dependence, we decompose the MSA level GDP into following 20 industries:

1. Agriculture, forestry, fishing, and hunting
2. Mining
3. Utilities
4. Construction
5. Manufacturing
6. Wholesale trade
7. Retail trade
8. Transportation and warehousing, excluding Postal Service
9. Information
10. Finance and insurance
11. Real estate and rental and leasing
12. Professional and technical services
13. Management of companies and enterprises
14. Administrative and waste services
15. Educational services
16. Health care and social assistance
17. Arts, entertainment, and recreation
18. Accommodation and food services
19. Other services, except government
20. Government
Table 3.1: Descriptive Statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>obs.</th>
<th>Mean</th>
<th>Median</th>
<th>Stdev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>real housing returns</td>
<td>43,896</td>
<td>0.0169</td>
<td>0.0141</td>
<td>0.0559</td>
<td>-0.3274</td>
<td>0.4117</td>
</tr>
<tr>
<td>income growth rate</td>
<td>43,896</td>
<td>0.0264</td>
<td>0.0252</td>
<td>0.0377</td>
<td>-2.4449</td>
<td>1.8542</td>
</tr>
<tr>
<td>population growth rate</td>
<td>43,896</td>
<td>0.0122</td>
<td>0.0102</td>
<td>0.0157</td>
<td>-0.2791</td>
<td>0.1773</td>
</tr>
<tr>
<td>economic similarities</td>
<td>43,896</td>
<td>0.6123</td>
<td>0.6500</td>
<td>0.2312</td>
<td>-0.2240</td>
<td>0.9990</td>
</tr>
</tbody>
</table>

Table 3.2: Two Stage Least Squares Results–Fixed Effects

<table>
<thead>
<tr>
<th>Variables/Coefficient</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.8567 (40.7014)***</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.3527 (5.1056)***</td>
</tr>
<tr>
<td>income growth</td>
<td>0.0590 (1.9025)*</td>
</tr>
<tr>
<td>population growth</td>
<td>0.2929 (10.7217)***</td>
</tr>
<tr>
<td>$n$</td>
<td>14,536</td>
</tr>
<tr>
<td>R-square</td>
<td>0.7195</td>
</tr>
</tbody>
</table>

dependent variable: real housing returns
t statistics in parentheses
* significant at 10%, ** significant at 5%, *** significant at 1%

Table 3.3: Two Stage Least Squares Results–Random Effects

<table>
<thead>
<tr>
<th>Variables/Coefficient</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.8474 (40.7014)***</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.1242 (4.9797)***</td>
</tr>
<tr>
<td>income growth</td>
<td>0.0389 (2.9062)***</td>
</tr>
<tr>
<td>population growth</td>
<td>0.1256 (3.4674)***</td>
</tr>
<tr>
<td>$n$</td>
<td>14,536</td>
</tr>
<tr>
<td>R-square</td>
<td>0.7084</td>
</tr>
</tbody>
</table>

dependent variable: real housing returns
t statistics in parentheses
* significant at 10%, ** significant at 5%, *** significant at 1%
### Table 3.4: GMM Estimation Results —with Missing Data Imputation

<table>
<thead>
<tr>
<th>Variables/Coefficient</th>
<th>Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.5334 (81.2411)***</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.3153 (26.9334)***</td>
</tr>
<tr>
<td>income growth</td>
<td>0.7226 (102.6267)***</td>
</tr>
<tr>
<td>population growth</td>
<td>0.1628 (40.6245)***</td>
</tr>
<tr>
<td>$n$</td>
<td>43,896</td>
</tr>
</tbody>
</table>

Dependent variable: real housing returns

T statistics in parentheses

* significant at 10%, ** significant at 5%, *** significant at 1%

### Table 3.5: Maximum Likelihood Estimation Results

<table>
<thead>
<tr>
<th>Variables/Coefficient</th>
<th>Model with $W_1$</th>
<th>Model with $W_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$</td>
<td>0.2668 (35.7766)***</td>
<td></td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>0.0732 (8.4438)***</td>
<td>0.9247 (5.2854)***</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td></td>
<td>0.8221 (4.9504)***</td>
</tr>
<tr>
<td>$\rho_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lagged housing returns</td>
<td>0.7552 (133.3420)***</td>
<td>0.8950 (224.7173)***</td>
</tr>
<tr>
<td>income growth</td>
<td>0.0301 (5.5050)***</td>
<td>0.0388 (6.5103)***</td>
</tr>
<tr>
<td>population growth</td>
<td>0.1919 (9.1502)***</td>
<td>0.1736 (7.6116)***</td>
</tr>
<tr>
<td>$n$</td>
<td>14,536</td>
<td>14,536</td>
</tr>
<tr>
<td>Maximum Likelihood</td>
<td>40,487</td>
<td>39,585</td>
</tr>
</tbody>
</table>

Dependent variable: quarterly real housing returns, t statistics in parentheses

* significant at 10%, ** significant at 5%, *** significant at 1%


