PREDICTING HOUSE PRICES FOR SINGAPORE CONDOMINIUM RESALE MARKET: A COMPARISON OF TWO MODELS

ZHOU QIN
(B. S. Yunnan University)

A THESIS SUBMITTED
FOR THE DEGREE OF MASTER OF ESTATE MANAGEMENT

DEPARTMENT OF REAL ESTATE

NATIONAL UNIVERSITY OF SINGAPORE

2007

ACKNOWLEDGEMENT

First and foremost, I would like to deeply thank my supervisor, Associate Professor Tu Yong, for her consistent guidance and unceasing patience.

I am also thankful to my friends and colleagues in NUS. Thanks for their kindly suggestions for my research and my life.

Last and most important, I would like to express my thanks and love to my husband, Li Chun, for his deepest love and strongest support to me.

TABLE OF CONTENTS

ACKOWLEDGEMENTS i

TABLE OF CONTENTS ii

SUMMARY v

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER 1: INTRODUCTION

1.1 Background 1

1.2 Objective of the Study 5

1.3 Scope of the Study 5

1.4 Data of the Study 5

1.5 Methodology of the Study 6

1.6 Hypothesis 6

1.7 Organization of the Study 6

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction 8

2.2 Heterogeneity in Hedonic Housing Price Models 9

2.3 Spatial Autocorrelation in Real Estate Market 11

2.4 Solutions to Spatial Autocorrelation in Hedonic Models 14

2.4.1 Lattice Models 15

2.4.2 Geostatistical Models 19

2.5 Summary 25

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction 27

3.2Operationalizing Geostatistical Models 27

3.3 Criteria for Evaluating Model Performance 31

3.4 Operational Considerations 32

CHAPTER 4: RESULTS AND ANANYSIS

4.1 Introduction 34

4.2 Data Collection 34

4.3 Empirical Results of Traditional Hedonic Model 42

4.4 Empirical Results of Geostatistical Models 47

4.5 Comparison of Performance between Geostatistical Model and 2STAR model 51

4.6 Summary 56

CHAPTER 5: MODEL PERFORMANCE IN CONSTRUCTING INDEXES

5.1 Introduction 57

5.2 Index Construction 57

5.3 Discussion on Building Specific Indexes 59

5.4 Discussion on Aggregate Indexes 69

5.5 Summary 74

CHAPTER 6: CONCLUSION, IMPICATIONS AND LMITATIONS

6.1 Conclusion 75

6.2 Implications 76

6.3 Limitations 77

BIBLIGRAPHY 79

APPENDIX A: MATLAB FUNCTION FOR THIS STUDY 86

APPENDIX B: TABLES AND FIGURES FOR CHATER 4 AND 5 91

SUMMARY

Spatial autocorrelations among house prices are more complicated in a multiunit housing market than that of a single family housing market due to the neighborhood effects confined at a particular building. Based on an assumption that the building-related neighborhood effect can be regarded as a nugget effect from the view of goestatistics, we employ the geostatistical model to solve the spatial autocorrelation problem in the hedonic price model and to obtain the accurate predicted house prices in the multiunit housing market. The empirical results based on the transaction data of Singapore condominium resale market from July of 1993 to June of 2003 show that the geostatistical models can effectively capture the building related spatial correlation. Among three geostatistical models, the negative exponential correlogram is the most appropriate method to predict the house price for the Singapore condominium resale market.

Moreover, a comparison of model performance in estimation and house price prediction is carried out between the geostatistical model and the two order spatio-temporal autoregressive (2STAR) model (the 2STAR is exclusively developed to solve the spatial autocorrelation problem in a multiunit housing market based on the lattice approach). We conclude that the geostatistical model can more effectively correct the inefficient OLS estimation resulting from spatial autocorrelation than the 2STAR model does. Regarding to the prediction ability, the geostatistical model outperforms the 2STAR model.

We also derive the building specific indexes and the aggregate indexes both from the

geostatistical model and the 2STAR model for the Singapore condominium resale market. Furthermore, we evaluate the model performance in constructing the specific indexes between the two models. We find that housing price dynamics varies significantly across buildings in the same project, so that the aggregate indexes can not be used as a proxy to reflect an individual building's price dynamics. Moreover, the building specific index derived from the geostatistical model provides a more accurate analysis of the price dynamics of a particular building than the index derived from 2STAR model does. However, the building specific index will be biased if the subject building has few transactions. Finally, we compare the aggregate indexes derived by the geostatistical model with the indexes by the 2STAR model, and with the URA released official indexes. We conclude that although the aggregate index derived from the geostatistical model can reasonably reflect the price change for the whole market, it is not recommended to use this model with the spatial weighted average technique to construct an aggregate index, for it can not provide obvious improvements in reflecting the price dynamics of the whole market compared with the index derived from a simple approach. It is because the spatial weighted average technique which is used to generate the aggregate index may demolish the distinct characteristics related to a particular location or a particular submarket.

LIST OF TABLES

Table 4.1 Variable definition 39

Table 4.2 Statistics of some key variables 40

Table 4.3 Average housing prices and the number of transactions over time 40

Table 4.4 Traditional hedonic model estimates 43

Table 4.5 Empirical results from July of 1995 to June of 1996 49

Table 4.6 Prediction errors of two models 53

Table 5.1 General information of the selected condominium projects 60

Table 5.2 Transaction information of Building 4 in Yong An Park 65

Table 5.3 Average house prices in the Claymore 67

Table 4.1 in Appendix B Empirical results from July of 1992 to June of 1993 91

Table 4.2 in Appendix B Empirical results from July of 1993 to June of 1994 92

Table 4.3 in Appendix B Empirical results from July of 1994 to June of 1995 93

Table 4.4 in Appendix B Empirical results from July of 1996 to June of 1997 94

Table 4.5 in Appendix B Empirical results from July of 1997 to June of 1998 95

Table 4.6 in Appendix B Empirical results from July of 1998 to June of 1999 96

Table 4.7 in Appendix B Empirical results from July of 1999 to June of 2000 97

Table 4.8 in Appendix B Empirical results from July of 2000 to June of 2001 98

Table 4.9 in Appendix B Empirical results from July of 2001 to June of 2002 99

Table4.10 inAppendixB Empirical results from July of 2002 to June of 2003 100

Table 4.11 in AppendixB Estimation and prediction results from OLS 106

Table 4.12 in Appendix B Estimation Results of 2STAR Model 109

Table 4.13 in Appendix B Estimation Results of 2STAR Model for the whole estimation sample 112

LIST OF FIGURES

Figure 4.1 Distribution of available private residential properties 35

Figure 4.2 Property price indices of Singapore condominium market 35

Figure 4.3 Correlograms from July of 1995 to June of 1996 50

Figure 5.1(a) Building specific indexes derived by GEO in Arcadia Garden 62

Figure 5.1(b) Comparison of building specific indexes in the Arcadia Garden 62

Figure 5.2(a) Building specific indexes derived by GEO in Yong An Park 66

Figure 5.2(b) Building specific indexes derived by 2STAR in Yong An Park 66

Figure 5.3 Building specific indexes in the Claymore 69

Figure 5.4 Aggregate indexes derived by GEO 70

Figure 5.5 Aggregate indexes derived by 2STAR 70

Figure 5.6 Aggregate indexes released by URA 70

Figure 5.7 Index of freehold and leasehold resale condos derived by GEO 72

Figure 5.8 Index of freehold and leasehold resale condos derived by 2STAR 72

Figure 5.9 Index of freehold and leasehold condos released by URA 72

Figure 5.10 Price index in Whole Area 73

Figure 4.1 in Appendix B Correlograms from July of 1992 to June of 1993 101

Figure 4.2 in Appendix B Correlograms from July of 1993 to June of 1994 101

Figure 4.3 in Appendix B Correlograms from July of 1994 to June of 1995 102

Figure 4.4 in Appendix B Correlograms from July of 1996 to June of 1997 102

Figure 4.5 in Appendix B Correlograms from July of 1997 to June of 1998 103

Figure 4.6 in Appendix B Correlograms from July of 1998 to June of 1999 103

Figure 4.7 in Appendix B Correlograms from July of 1999 to June of 2000 104

Figure 4.8 in Appendix B	*Correlograms from July of 2000 to June of 2001*	*104*
Figure 4.9 in Appendix B	*Correlograms from July of 2001 to June of 2002*	*105*
Figure 4.10 in Appendix B	*Correlograms from July of 20021 to June of 2003*	*105*
Figure 5.1 in Appendix B	*Price index in Central Area*	*113*
Figure 5.2 in Appendix B	*Price index in East Area*	*113*
Figure 5.3 in Appendix B	*Price index in West Area*	*113*
Figure 5.4 in Appendix B	*Price index in Northeast Area*	*114*

Chapter 1 Introduction

1.1 Background

Nowadays, residential houses are the most common source of accumulating personal wealth, the value of which is the major concern in investment decisions, second mortgages and home owners insurance. However, in most cases we use the predicted price rather than the transaction price. Therefore, it is necessary and important to predict the price of the house as accurate as possible.

Conventionally, the widely used method of predicting house prices is the hedonic model, which is a function of the house price against physical housing characteristics and neighborhood effects. The hedonic model is commonly estimated by the Ordinary Least Square (OLS) technique, and the estimated coefficients are used to produce the predicted house prices. A common problem in estimating the hedonic model with the OLS technique is the spatial autocorrelation between the residuals, which results from the spatial autocorrelation between the house prices caused by neighborhood effects. The spatial autocorrelation in the residuals may violate one of assumptions that there is no correlation between the error terms, which ensure the OLS estimates accurate. As a result, the OLS estimates are unbiased but inefficient, which may cause the predicted house prices to be inaccurate.

In order to solve the problem of spatial autocorrelation and obtain the accurate predicted house prices, many studies have been done in the real estate literature. In general, there are two main approaches concerning how to address spatial autocorrelation in the hedonic residuals. One is the lattice model, which uses a connectivity matrix to define the generating process of the residuals (Pace and Barry, 1997; Pace and Barry, 1998; Pace *et al.*, 1998, 2000; Sun *et al.*, 2005). The other is the geostatistical model, which postulates that the correlation between the residuals is a function of the distance separating their locations (Dubin, 1988, 1992, 1998; Basu and Thibodeau, 1998). Once the spatial pattern between the residuals is estimated, the efficient coefficients of the hedonic model can be estimated by the Generalized Least Square (GLS) and the Maximum Likelihood (ML) techniques.

Although both approaches have made some achievements in literature, most studies only focus on the single family housing market rather than the multiunit housing market. Since the spatial structure in multiunit housing market is more complicated than the single family housing market, the models which are originally developed for the single family housing market may not be appropriate for the multiunit housing market. Therefore, it becomes necessary to exclusively solve the complicated spatial autocorrelation problems and thus obtain the accurate predictions in the multiunit housing market. In the literature, it is Sun *et al.* (2005) who first work on solving the spatial autocorrelation problem and make some success in the multiunit housing market from a context of lattice model. In their study, Sun *et al.* (2005) assume that the spatial structure in the multiunit housing market is complicated by both the uncaptured building-related information and the uncaptured neighborhood information in a multiunit housing market. They argue that the two spatial processes are different and term the former spatial information as *building* *effect* and the latter one as *neighborhood effect*. Therefore, from a context of the lattice model, Sun *et al.* (2005) develop a Two Order Spatial-temporal Autoregressive model (2STAR) to capture the complicated spatial structure for the Singapore condominium market, by splitting the spatial matrix in the Spatial-temporal Autoregressive model (STAR) into a building effect matrix and a neighborhood effect matrix.

Although the 2STAR model outperforms the STAR which is developed for the single family market, several limitations of it still exist. Firstly, the exogenously determined building effect matrix is probably problematic as there is no prior knowledge about the spatial structure of the spatial effect related the building. Secondly the temporal matrix might not reflect the actual temporal process for any particular point as the matrix is determined by the fixed number of prior transactions across the whole area and will reflect the process of the whole area. Thirdly, there is a degree of multicollinearity between the two separate spatial matrices as the so-called building effect matrix capture a certain degree of neighborhood effect. It is because all dwellings in one building must be in one neighborhood. Therefore, the estimates from the 2STAR model might be inaccurate and the predicted house prices thus produced might be inaccurate.

The limitations of the 2STAR model raise two research questions: as an alternative approach to the lattice models, can the geostatistical model be applied to solve the spatial autocorrelation problem for the multiunit housing market? Which model will have a better performance in predicting the house prices? These two questions motivate us to apply the geostatistical model to solve the spatial autocorrelation problems in the multiunit housing market and compare its model performance with the 2STAR model.

In our study, we argue that the building related spatial effect is a small scale of neighborhood effect related to one specific location, that is, a neighborhood effect at the building level. Since the correlation of the price of a dwelling with the price of other dwelling in the same building is very different from one and the distance between them is zero, the building-related spatial effect can be regarded as a nugget effect from the viewpoint of geostatistics. Therefore, we expect the geostatistical approach would effectively identify the spatial autocorrelation structure in a multiunit housing market and the predicted house prices thus obtained would be accurate. We also expect that the geostatistical model performs as well as, if no better than, the 2STAR model in a multiunit housing market.

In our study, we first employ the geostatistical models to model the spatial structure of the Singapore condominium resale market. Next, the evaluation of model performance in prediction is carried out among three geostatistical models, and between the geostatistical model and the 2STAR model. Finally, a comparison of model performance in constructing the building specific indexes and the aggregate indexes is made between the geostatistical model and the 2STAR model. Furthermore, the aggregate indexes generated from these two models are compared with the URA official index.

1.2 Objectives of the Study

The objectives of this study are:

1: To employ the geostatistical model to identify the spatial autocorrelation structure in a multiunit housing market.

2: To evaluate the model performance in house price prediction between the three geostatistical models, between the geostatistical model and the 2STAR model in a multiunit housing market.

3: To evaluate and discuss the model performance in constructing building specific indexes and aggregate indexes between the geostatistical model and the 2STAR model in a multiunit housing market.

1.3 Scope of the Study

Since we aim at predicting accurate housing prices for a multiunit housing market, the empirical study chooses the Singapore housing market, especially limit to the Singapore condominium resale market.

1.4 Data of the Study

The original condominium resale data with hedonic characteristics are obtained from an online real estate transaction database called Reallink, which is maintained by Singapore Institute of Surveyors and Valuers (SISV), the national professional body providing the real estate professional services. The study period spans from July of 1993 to June of

2003. In total, there are 26,446 transactions in our sample.

1.5 Methodology of the study

In our study, three geostatistical models, the negative exponential correlogram, the Gaussian correlogram and the spherical correlogram are employed. The operation procedure of the geostaistical model is following Dubin (2003). The 2STAR model is following Sun *et al.*'s procedure (2005) without any change.

1.6 Hypotheses

Based on the research objectives and methodology, the following hypotheses are formulated in our study:

1. The geostatistical model can effectively identify the spatial autocorrelation structure and thus produce accurate predictions of house prices in a multiunit housing market.
2. The geostatistical model performs as well as, if not better than, the 2STAR model in predicting house prices in a multiunit housing market.
3. The building specific index derived from the goestatistical model can provide a clearer picture of price dynamics of that building than that from the 2STAR model.

1.7 Organization of the Study

This study is organized into six chapters, as stated following:

Chapter 1 provides an overview of this study, including background, objective, scope, data, methodology and hypothesis.

Chapter 2 provides a literature view of issues about the spatial heterogeneity and spatial autocorrelation in the housing market. In what follows is a review of the solutions to the spatial autocorrelation problem in the hedonic housing price model.

Chapter 3 provides an explicitly explanation of the geostatistical models employed and addresses several operational considerations in our study.

Chapter 4 first describes the Singapore condominium market and the dataset used in our study. Next, it presents the empirical results obtained from the traditional hedonic model and then presents the empirical results of three geostatistical models. An evaluation of model performance in estimation and house price prediction between the geostatistical model and the 2STAR model is presented in the followed section.

Chapter 5 first provides a detailed discussion and comparisons of the building specific indexes derived by the geostatistical model and the 2STAR model across three selected condominium projects. In what follows, the aggregate indexes derived from these two models have been discussed and compared with each other and with the URA official index.

Chapter 6 summarizes the research findings and implications together with the existing limitations of this study.

Chapter 2 Literature Review

2.1 Introduction

Over the last thirty years, the hedonic housing price model has been widely used in real estate market, which is a function of the purchase price or rent against physical housing characteristics and neighborhood effects. It is a basis for real estate appraisal and valuation, for estimation of the willingness to pay for different housing attributes and for constructing constant quality price indexes.

The most widely used estimation technique for this model is the Ordinary Least Square (OLS) (Bjorklund *et al.*, 2002). However, in order to estimate the OLS accurately, a number of assumptions have to be made (Gujarati, 2003). In particular, the assumptions about constant variance across the sample data and no correlation between the dependent variables and the error term are often violated in real estate cross-section data set. In terms of spatial econometrics, a violation of the former assumption is named *spatial heterogeneity* and the latter is named *spatial autocorrelation*. In both cases, the OLS estimates are unbiased but inefficient, and the house price prediction based on the estimates may be inaccurate. Furthermore, the variance calculated by the standard OLS procedure is biased, and thus it is difficult to make inference (Fletcher, 2000; Dubin, 1998).

In recent years, there are a large number of studies on the issues of spatial heterogeneity and spatial autocorrelation in real estate literature. In this chapter, we will first review the issue of heterogeneity in hedonic models and then discuss the spatial autocorrelation in real estate market. In what follows, we will review the solutions to the spatial autocorrelation problem in hedonic housing price models. Finally, a summary of literature review is given.

2.2 Heterogeneity in Hedonic Housing Price Models

Heterogeneity has long been treated as a potential problem in hedonic housing price models. Fleming and Nellis (1984) mention this potential problem and examine the residuals graphically when they work on constructing a house price index for Halifax.

Later, Goodman and Thibodeau (1995, 1997) state that the heterogeneity in hedonic models is related to the age of properties. Property age heterogeneity is likely because the magnitude of the error in predicting house price probably increases with property age. The older a property, the more likely the property was significantly upgraded or improved at some time during its life. Since the improvements are typically not recorded in publicly available datasets, there is no way to incorporate these improvements in the hedonic model. In their first paper, they employ an iterative generalized least square procedure to examine the existence of age-related heterogeneity in hedonic model. The empirical result shows the existence of age-related heterogeneity. In their second paper, they extend their work by including more structural variables into the model and controlling for neighborhood effects. As heterogeneity may be a result of mis-specification of the explanatory variables in the model, these refinements are expected to find other induced heterogeneity. However, the age-related heterogeneity continues to be present in the revised model, indicating the earlier result is robust.

Fletcher *et al.* (2000) extend the work of Goodman and Thibodeau and investigate whether heterogeneity is related to other factors as well as age. They find that heterogeneity is related to both the age and the external area of the property.

Stevenson (2004) re-examines the issue of heterogeneity in hedonic housing price models using Boston, MSA data because of its high average age of dwelling. The results greatly support previous findings with the evidence of heterogeneity with respect to the age of dwelling. He also argues that a correction technique for a single variable may not eliminate all heterogeneity from the hedonic model.

In real estate market, it is accepted that housing characteristics are not constant over space. Therefore, spatial heterogeneity is also expected in hedonic housing price models, which reflects a variation in relationship over space (La Sage, 1998). De Graaff *et al.* (2001) point out a number of reasons why spatial heterogeneity should be handled jointly with spatial autocorrelation. Firstly, it may be difficult empirically to separate the two special effects. Secondly, spatial dependence induces a particular form of heterogeneity in real estate market. Finally, there may be no difference between heterogeneity and dependence in an observational sense.

Using Singapore condominium transaction data, Sun *et al.* (2005) deal with the spatial heterogeneity and the spatial autocorrelation problems at the same time with the 2STAR model combined with Bayesian estimation method. They find that there is a trade off between the heterogeneity robustness and the incorporation of spatial correlation information into the model estimation.

Although handling spatial heterogeneity jointly with spatial autocorrelation raises much interest in the real estate literature, we only focus on solving the spatial autocorrelation problem because of limited time and space in our study.

2.3 Spatial Autocorrelation in Real Estate Market

Spatial autocorrelation refers to the possible occurrence of interdependence among the observations from a geographical area. Spatial autocorrelation in house prices is caused by the neighborhood effect which comes from two ways (Basu and Thibodeau,1998): one is that properties in the same neighborhood have similar structural characteristics because neighborhoods tend to be developed at the same time; the other is that neighborhood residential properties share similar accessibility and location amenities, such as public schools, fire departments and so on.

The residuals produced by hedonic housing models are frequently correlated for two reasons (Dubin, 1992). Firstly, it is impossible to specify all the independent variables related to neighborhood so that the residuals appear no pattern over space. For example, Pace *et al.* (2000) estimate a hedonic model using 199 spatial indicator variables. They make a conclusion that the correlation is still above 0.15 for the hedonic residuals of the nearest neighboring housing units. Secondly, even if all the location variables are included in the model, the residuals may still be spatially autocorrelated. It is because the analysts lack the ideal measure for neighborhood services and they are uncertain about how location characteristics are capitalized into house prices (Basu and Thibodeau, 1998; Dubin, 1998a). For instance, it is more difficult to measure the location characteristics such as the quality of public school than the structural characteristics which are typically included in publicly available data.

A lot of researchers have examined empirically the spatial autocorrelation in house prices and in hedonic model residuals in single family housing markets. Assuming the residual correlation between properties is a negative exponential function of the distance between them, Dubin (1988, 1992) estimates the hedonic parameters and the correlation function parameters using a Maximum Likelihood procedure. The result shows that this approach provides a very plausible pattern of housing price variation. By interacting a linear combination of neighborhood variables with traditional structure variables, Can (1990) estimates the effect of neighborhood quality on housing prices. She concludes that implicit prices on the structural attributes vary significantly with neighborhood quality. Can and Megbolugbe (1997) illustrate the importance of spatial dependency for constructing a house price index with comparisons between an often-used hedonic price index model and a hedonic spatial model. Basu and Thibodeau (1998) examine the spatial autocorrelation both in transaction prices and in hedonic residuals for single family properties in Dallas area. They find strong evidence of spatial autocorrelation in transaction prices within all submarkets and the hedonic residuals are spatially autocorrelated only in two of four submarkets.

Sun *et al.*'s study (2005) argue that the spatial autocorrelation is complicated by the building-related spatial effect a multiunit residential market in the following ways. *'First, all the units in one building are located in one location, and hence share similar attributes of that unique location, such as the distance to the main road, which may have an effect on property value. For example, the units in the building that is further away from the main road are likely to fetch a higher price because of the reduction in traffic noise. Second, all the units in one building share similar structural attributes, such as design and layout, which partly determine the property value. Third, the units within one building may fetch different prices due to their specific location within the building. For example, units located at the higher floors are likely to have a better view than those on lower floors, thus enjoying a price premium'.* This kind of building-related spatial effect is termed as *building effect* in the study.

However, we argue that the *building effect* is a kind of neighborhood effect related to one specific location, that is, a neighborhood effect at the building level. Clearly, the location of a building affects the price of housing units in it in the same way as the location of a project affects the price of houses in it, and the correlation between the housing units within a building is techniquely like the correlation between houses within a neighborhood but of different magnitudes. Since the correlation of the price of a dwelling with the price of other dwelling in the same building is very different from one and the distance between them is zero, the neighborhood effect the building level can be regarded as a nugget effect from the viewpoint of geostatistics. Therefore, we expect the geostatistical approach would effectively identify the spatial autocorrelation structure in a multiunit housing market and the predicted house prices thus obtained would be accurate.

It is noted that the *building effect* manifested in the third way is the really complicated spatial effects in the multiunit housing market. It is a kind of spatial effect manifested in the vertical way rather than the horizontal way. To be honest, the spatial effect in such a vertical way is too complicated to be studied in the current study. Neither the building effect matrix in the 2STAR model nor the nugget effect in the geostatistical model can capture it.

2.4 Solutions to Spatial Autocorrelation in Hedonic Models

A hedonic housing price model can be specified as

$$\begin{aligned} &Y = X\beta + u \\ &E[uu'] = \sigma^2 K = \Omega \end{aligned} \tag{2.1}$$

Where Y is a $n \times 1$ vector of observations of the dependent variables like transaction prices, X is a $n \times m$ vector of housing characteristics, β is a $m \times 1$ vector of parameters to be estimated, u is a $n \times 1$vector of residuals, K and Ω is the $n \times n$ correlation matrix and the $n \times n$ covariance matrix respectively. When the residuals are spatially autocorrelated, both K and Ω have non-zero off diagonal terms which express the dependence of residuals among different locations. Ex-ante, the magnitude of the covariance between any two residuals declines as the distance increases between them. Once the covariance matrix Ω (or equivalently K) is modeled, the efficient parameters can be estimated by an Estimated Generalized Least Squares or a Maximum Likelihood method (Dubin *et al.*, 1999). The means of modeling the covariance matrix or spatial correlation between residuals distinguishes two strands of approaches to solve the spatial autocorrelation problem: the lattice model and the geostatistical model. We will review the lattice model and the geostatistical model in Section 2.4.1 and Section 2.4.2 respectively.

2.4.1 Lattice models

The lattice model, also known as the matrix approach, uses a connectivity matrix to define a process generating the hedonic residuals. The covariance matrix then can be derived from the specified process (Dubin, 1998a). Mathematically,

$$u = \lambda W u + \varepsilon \qquad (2.2)$$

Substitute Equation (2.2) into (2.1) and rearrange it, we get

$$(I - \lambda W)Y = (I - \lambda W)X\beta + \varepsilon \qquad (2.3)$$

Solving (2.2) for u gives

$$u = (I - W)^{-1}\varepsilon \qquad (2.4)$$

and thus

$$\Omega = E(uu') = \sigma^2 (I - \lambda W)^{-1}(I - \lambda W')^{-1} \qquad (2.5)$$

Here, u is defined as in Equation (2.1), ε is a vector of normally distributed and independent error terms (with mean zero and variance σ^2), λ is an unknown autocorrelation parameter, and W is the weight matrix which represents the spatial structure of the data.

The model specified in Equation (2.3) is termed the Spatial Autoregressive Model (SAR) because of its similarity to the time series autoregressive model (Pace and Barry,1997). The SAR model can be expanded by using various lags, just as in the time series autoregressive model (Anselin, 1998). The SAR model performs better than OLS to a large extent both on model fit and statistical inference (Pace and Barry, 1997). Pace and Gilley (1997) demonstrate that the estimated errors from the SAR model falls 44% relative to the traditional OLS model. One problem often met in applying the SAR model is the computation problems when data size is large. Pace and Barry (1997) make contributions on solving this problem by providing a sparse matrix approach. In their study, 20,640 observations of housing prices in California have been analyzed by the SAR model.

Based on the SAR model, Pace *et al.* (1998, 2000) develop a Spatial-Temporal Autoregressive model (STAR). This model takes into account not only spatial but also temporal information by partitioning the weight matrix into a matrix that specifies spatial relationship among observations and another matrix that specifies temporal relations among previous observations. Mathematically, the W matrix in Equation (2.3) can be generalized into a flexible form as:

$$(I - \lambda W) = (I - \lambda_s S - \lambda_T T - \lambda_{ST} ST - \lambda_{TS} TS) \qquad (2.6)$$

Here, S and T refer to the appropriate spatial and temporal weight matrices representing spatial and temporal filtering process separately. ST and TS are the interaction of spatial and temporal matrices indicating the interactive spatial and temporal filtering processes. λ_s, λ_T, λ_{ST} and λ_{TS} are the parameters of the filtering variables. Since the sale price of a neighboring property may influence the subject property only if the neighboring sale is earlier in time, the spatial and temporal weights are supposed to be conditioned on the previous transactions and thus both matrices are lower triangular matrices with diagonal term as zero. Compared with the traditional indicator based model, the STAR model greatly enhances the accuracy in model estimation and reduces the reliance on the number of price determinants.

Based on the STAR model, Sun *et al.* (2005) develop the 2STAR model in order to capture the complicated spatial autocorrelation in the Singapore multiunit housing market, by further splitting the spatial matrix into a neighborhood effect matrix and a building effect matrix. That is, Equation (2.6) can be further specified as

$$(I - \lambda W) = (I - \lambda_{W1} W_1 - \lambda_{W2} W_2 - \lambda_T T - \lambda_{ST} ST - \lambda_{TS} TS) \qquad (2.7)$$

Here, W_1 and W_2 are spatial weight matrices aimed at filtering down the building and neighborhood effects separately, S is a combined spatial matrix that does not differentiate the two effects, T is a temporal matrix as defined in Equation (2.6). λ_{w1}, λ_{w2} , λ_s, λ_T, λ_{ST} and λ_{TS} are the parameters of the filtering variables.

Substitute Equation (2.7) to Equation (2.3), we get

$$\begin{aligned} Y = X\beta - W_1 X\lambda_{W1}\beta - W_2 X\lambda_{W2}\beta - TX\lambda_T\beta - STX\lambda_{ST}\beta - TSX\lambda_{TS}\beta \\ + \lambda_{W1} W_1 Y + \lambda_{W_2} W_2 Y + \lambda_T TY + \lambda_{ST} STY + \lambda_{TS} TSY + \varepsilon \end{aligned} \qquad (2.8)$$

Sun *et al.* (2005) shows the 2STAR model outperforms the STAR model for solving the spatial autocorrelation problem in the Singapore condominium market.

Although the 2STAR model can capture more spatial information than the STAR model for a multiunit housing market, several limitations of it exist.

Firstly, the spatial weight matrix is determined exogenously, i.e., it is based on a prior knowledge or belief of the spatial structure. In a single housing market, although the spatial structure of which has been studied intensively, there is little agreement regarding the best form for the weight matrix, and the results are conditionally upon the specification of the weight matrix (Dubin, 1998). Therefore, the spatial matrices especially the building effect matrix may be problematic in the 2STAR model because there is not so much prior belief about the spatial structure in a multiunit housing market. It might result in an ineffective performance in capturing actual spatial structure in the multiunit market.

Secondly, as specified by the temporal matrix in Equation (2.8), any point on the spatial space is affected by the same prior transactions which may be scattered across the whole market. However, it is recognized by most scholars that urban housing market should be depicted as a set of distinct but interrelated housing submarkets, across which the hedonic housing prices are significantly different. Therefore, this temporal matrix may capture the temporal trend of the whole market, but it is problematic in reflecting the actual temporal process for a single submarket or for any particular point in the market.

Finally, there is a degree of multicollinearity between the two separate spatial weight matrices, and between each separate spatial matrix and the combined spatial effect matrix(S). As we know, all dwellings in one building must be in one neighborhood, the building effect matrix captures a certain degree of neighborhood effect. Furthermore, the combined spatial effect matrix(S) captures a certain degree of building-related and neighborhood-related spatial effect. As a result, the estimates from the 2STAR model may be inaccurate.

2.4.2 Geostatistical models

The geostatistical model postulates that the correlation between observations is a function of the distance separating their locations and thus estimates the covariance matrix Ω of the error term directly[1] , which can be specified in following ways.

Let s_i denote the site of property i , and $u(s_i)$ denote the hedonic price residual for the property located at s_i . If the spatial process is second-order stationary and isotropic, meaning that the mean and variance of each residual distribution are constant at all locations and the correlation between the residuals is a function of the distance separating the properties only. The co-variogram for the distribution of residuals is $C(s_i - s_j) = Cov\{u(s_i), u(s_j)\} = \Omega_{ij}$ for all (s_i, s_j) . $s_i - s_j$ denotes the (Euclidean) distance between locations s_i and s_j and $Cov\{u(s_i), u(s_j)\}$ is the

[1] This is strictly true only for isotropic models

covariance between the two residuals. Note that $C(0)$ is the constant variance for the residual distribution. The semivariogram of the process is

$$\gamma(s_i - s_j) = 0.5Var\{u(s_i) - u(s_j)\} = C(0) - C(s_i - s_j) \qquad (2.9)$$

The semivariogram defined in equation (2.5) is an increasing function of the distance between any two properties. Other features in relation to the semivariogram are: Assuming that $d = s_i - s_j$, clearly $\gamma(-d) = \gamma(d)$ and mathematically $\gamma(0) = 0$. However, sometimes $\gamma(d)$ is discontinuous near the origin and $\gamma(d) \to c_0 > 0$, as $d \to 0$. Matheron (1963) labeled the discontinuity "c_0" as the nugget. This term is from mining geostatistics where nuggets literally exist. Residuals may eventually become spatially uncorrelated as the distance between them increases. Therefore, the semivariogram will stop increasing beyond some threshold and will become a constant. That is, $\gamma(d) \to C*$, as $d \to \infty$. The threshold "$C*$" is called the sill of semivariogram. The range of semivariogram is the value "d_0", which makes $\gamma(d_0) = C*$. So the range of semivariogram is essentially a distance between two residuals beyond which residuals become spatially uncorrelated. Finally, a semivariogram is isotropic if $\gamma(s_i - s_j)$ is a function of only the distance and not the direction between s_i and s_j.

There are three popular isotropic semivariogram models to empirically examine relationships in spatial data (Cressie,1993): Negative exponential, Gaussian and Spherical semivariograms, as defined blow:

Negative exponential semivariogram

$$\gamma(d)=\begin{cases} c_0+c_1(1-\exp(-\frac{d}{d_0})) & d>0 \\ 0 & d=0 \end{cases} \qquad (2.10)$$

Gaussian semivariogram

$$\gamma(d)=\begin{cases} c_0+c_1(1-\exp(-\frac{d}{d_0})^2) & d>0 \\ 0 & d=0 \end{cases} \qquad (2.11)$$

Spherical semivariogram

$$\gamma(d)=\begin{cases} c_0+c_1(1-\frac{3d}{2d_0}+\frac{d^3}{2{d_0}^3}) & 0<d\le d_0 \\ c_0+c_1 & d>d_0 \\ 0 & d=0 \end{cases} \qquad (2.12)$$

Where c_0 is the nugget, c_0+c_1 is the sill, and the range is d_0. Many procedures like ML method and Weighted Least Squares(WLS) method have been used to estimate the three unknown parameters (c_0, c_1, d_0) in semivariograms (Milinino, 2004). Once the parameters are estimated, covariance matrix Ω can be calculated and incorporated into a regression model to estimate the regression parameters.

Similarly, the correlation matrix K in equation (2.1) also can be estimated directly. Suppose the sample size is N, the correlation matrix K will be of order N and have N^2 elements. In order to estimate these elements, a functional form must be assumed first because they can not be estimated from the N observations. The parameters of this function can then be estimated from the sample observations and the elements of K can be determined by the estimated function. This function, known as a correlogram, expresses the correlation between two points as a function of distance separating them. Ex ante, this function is such that the correlation between two points increases as the distance between them declines. When the distance decreases to zero, the correlogram is the correlation of the point with itself. One might think the correlation of an observation with itself should be one. However, the correlation function may be discontinuous at the origin, that is, at very small separation distance the correlation may be very different from one due to the nugget effect. Clearly, nugget effect exists in real estate market for a number of reasons.

Firstly, it is likely that house prices contain error (Dubin, 1998). The sale price of a house represents a compromise between the buyer and the seller. Many factors may cause the same property to sell for a different amount at a different point in time even though nothing about the house has changed. Among these factors are seller's need for cash, the strength of the buyer's desire for the property, and the relative bargaining skills of the two parties. The presence of measurement error means that the correlogram should return a correlation smaller than one when the separation distance is zero.

Secondly, a house on a busy street will probably have a lower value than a nearby house located in the quiet interior of the block. Theoretically, the correlation between them should be very different from one even if the distance between them is close to zero.

Moreover, the nugget effect is self-evident in a multi unit housing market. House units in the same building share the same X-Y coordinate and the Euclidean distance between these transactions is zero. But the correlation between them is definitely not one.

It is Dubin (1988) who firstly introduces the geostatistical technique to the literature of real estate. To examine the spatial autocorrelation in the hedonic house price residuals for Baltimore homes, Dubin (1988) estimates hedonic parameters using 221 transactions of properties sold in 1978. She assumes that the residual correlation between properties is a negative exponential function of the distance between them and estimates the hedonic parameters using the ML procedure. The empirical result shows that ML method is useful for analyzing the spatial correlated cross-sectional data set.

As an extension of her earlier work, Dubin (1992) presents an alternative approach for modeling spatial autocorrelation in the hedonic house price residual. In her study, she omits all neighborhood and accessibility measures from the set of explanatory variables and examines residual autocorrelation using exponential correlogram. Moreover, she predicts market values using kriging. Kriging is a statistical technique borrowed from geostatistic. It predicts house prices based on the structural characteristics of properties and on an average of the hedonic residuals for nearby properties.

Using a semilog hedonic house price equation and a spherical semivariogram with data for over 5000 transactions of homes sold between the fourth quarter of 1991 and the first of quarter of 1993, Basu and Thibodeau (1998) examine the spatial autocorrelation in transaction prices of single family properties in Dallas, Texas. The hedonic and spherical semivariogram parameters are estimated separately for each submarket using the Estimated Generalized Least Square (EGLS). They find strong evidence of spatial autocorrelation in transaction prices within all submarkets. As to the hedonic equation residuals, there is evidence of spatial autocorrelation for properties located within a 1200 meter radius in four of eight submarkets. In two submarkets, the hedonic residuals are spatially autocorrelated throughout the submarket, while there is no spatial autocorrelation in hedonic residuals in the remaining two submarkets. The kriged EGLS performs better than OLS in six of eight submarkets where the hedonic residuals are spatial autocorrelated, while OLS has smaller prediction errors in submarket where the residuals are spatially uncorrelated.

Using Baltimore multiple listing data to estimate hedonic model and to predict future prices, Dubin (1998) compares the model performance between OLS and the negative exponential and Gaussian correlograms, between these two correlograms, and between distance separated two houses measured in houses and measured in feet. The results demonstrate that both forms of the correlograms provide an improvement over OLS and the negative exponential is better. In addition, distance measured in feet appears to be better than measured in houses.

Although both of these two approaches have made some achievements in real estate literature, there is little theoretical justification underlying the choice of the model, the researcher never knows which model has generated the error terms (Dubin, 2003). It is possible that the researcher is estimating a misspecified model because of lack of theoretical guidance. To examine how well the models perform when the error structure has been incorrectly specified through a series of Monte Carlo experiments, Dubin (2003) finds that all the spatial models are robust with respect to standard statistical inference, and all predict better than OLS. With respect to house price prediction, there is a clear difference between the spatial models: the geostatistical models dominate the lattice models.

2.5 Summary

In summary, there are some issues which have not been solved and require further exploration:

1. Most studies on the spatial autocorrelation except Sun *et al.* (2005) focus on the single family market rather than a multiunit housing market. Due to the popularity of multiunit residence in Asian countries, little work has been done for our understanding the spatial structure in the multiunit housing market.

2. Although a self-evident nugget effect exists in the multiunit housing market, the geostatistical model has never been applied to model its spatial structure.

3. Although the 2STAR model effectively capture the spatial information in a multiunit housing market, there still are some limitations of the model as discussed in Section 2.3.1.

The above mentioned problems motivate this study to fill the gaps and serve as a contribution to both international and local literature. In our study, the geostatistical model is employed to model the spatial structure in the Singapore condominium resale market. The comparisons of model performance in estimation and house price prediction are carried out among the geostatistical models, and between the geostatistical model and the 2STAR model. Moreover, the comparisons of model performance in constructing building specific indexes and aggregate indexes are made between the geostatistical model and the 2STAR model. Further, the aggregate indexes generated from these two models are compared each other and with the URA official index.

Chapter 3 Research Methodology

3.1 Introduction

In this chapter, we will explicitly illustrate the procedure of operating geostatistical models following Dubin (2003). In what follows, we will introduce some criteria for evaluating model performance and address some operational issues in our study

3.2 Operationalizing Geostatistical models

Hedonic Housing Price Model

In order to make the variance of disturbance with respect to price constant in relative terms, a semi-logarithmic hedonic function is used in our study. Rewrite equation (2.1) as

$$\begin{aligned} Y &= Log(P) = X\beta + u \\ E[uu'] &= \sigma^2 K \end{aligned} \tag{3.1}$$

Where P is a vector of dwelling transaction prices, X , β , u , K are of the same definition as in equation(2.1). In our study, I exclude all the variables related to neighborhood characteristics so that the spatial relationship information among the house unit prices can be included in the residual as much as possible. In geostatistical terminology, y and u are known as regionalized variables. A regionalized variable is simply a random variable that is tied to a particular location. For example, the price of a house is tied to its location, or the house's coordinate. The regionalized variable has a probability distribution at each location. It is necessary to make some assumptions regarding these distributions to draw inference from the geographically scattered data. For example, u should be second-order stationary to draw inference, which requires that the mean and the variance of the regionalized variable be the same at all locations and the correlation between the regionalized variables depend on the distance between them. Here, we assume u is second-order stationary.

Correlation Function

There are several functional forms that are valid for correlograms(Christensen, 1991). In our study, we will use three popular forms of correlation function to find the most appropriate one for the multiunit housing market. These functional forms are,

the spherical correlation function

$$K_{ij} = \begin{cases} b_1\left(1 - \frac{3d_{ij}}{2b_2} + \frac{d_{ij}^{\,3}}{2b_2^{\,3}}\right) & \text{for} \quad 0 < d_{ij} \le b_2 \\ 1 & \text{for} \quad d_{ij} = 0 \\ 0 & \text{for} \quad d_{ij} > b_2 \end{cases} \tag{3.2}$$

the Gaussian correlation function

$$K_{ij} = \begin{cases} b_1 \exp\left(-\left(\frac{d_{ij}}{b_2}\right)^2\right) & \text{for} \quad d_{ij} > 0 \\ 1 & \text{for} \quad d_{ij} = 0 \end{cases} \tag{3.3}$$

And the negative exponential correlation function

$$K_{ij} = \begin{cases} b_1 \exp\left(-\left(\frac{d_{ij}}{b_2}\right)\right) & \text{for} \quad d_{ij} > 0 \\ 1 & \text{for} \quad d_{ij} = 0 \end{cases} \tag{3.4}$$

Where K_{ij} is an element of the correlation matrix K , representing the correlation between two observations separated by distance d_{ij}. Clearly, K is a symmetric matrix. b_1 and b_2 are the parameters to be estimated. One might think that b_1 should be restricted to be one as the correlation of an observation with itself should be one. As discussed in Section 2.4.2, b_1 will be less than one due to the nugget effect.

Estimation

The parameters to be estimated are: the regression parameters β and the correlation parameters b_1 and b_2. These parameters can be estimated using the Maximum Likelihood method simultaneously, and Equation (3.5) shows the concentrated log likelihood function. Once estimates of b_1 and b_2 are obtained, K can be determined and β can thus be estimated immediately.

Here we choose the correlation parameters (b_1 and b_2) which maximize the concentrated log likelihood function with a grid-search procedure(Dubin,2003): b_1 is allowed to vary between 0.1 and 1, in increments of 0.1, and b_2 is allowed to vary between 0.5km and 5km[2], in increments of 0.5km. For each b_1, b_2 pair, K is computed first, which allow us evaluate β, which in turn allow us evaluate L. For the b_1, b_2 pair which maximize the L, the search procedure is repeated by reducing the increments by half and evaluating the pairs on either side. The search procedure is

[2] Empirical correlograms indicate the correlation between house prices starts fluctuating around zero at around 4.5 km or less in most cases for the Singapore condominium market.

repeated one more time to produce the final estimates.

$$L = -\frac{N}{2} \times \ln\left[(Y - X\tilde{\beta})'\tilde{K}^{-1}(Y - X\tilde{\beta})\right] - \frac{1}{2}\ln\left|\tilde{K}\right|$$
$$\tilde{\beta} = (X'\tilde{K}^{-1}X)^{-1}X'\tilde{K}^{-1}Y \qquad (3.5)$$

Prediction

When the errors are spatially autocorrelated, it is possible to make a prediction for the error term using kriging. Kriging is a minimum mean squared error statistical procedure for spatial prediction that assign differential weights to observations that are spatially close to the dependent variable's location(Goldberger,1962).

This predicted error is then added to the standard $X\tilde{\beta}$ to make an improved prediction. The prediction at site s_0 is

$$\hat{Y}(s_0) = x_0\tilde{\beta} + \hat{u}(s_0) \qquad (3.6)$$

where $\hat{u}(s_0)$ is the predicted error at site s_0, $\hat{Y}(s_0)$ is the predicted value at site s_0, x_0 is the vector of independent variable at site s_0, $\tilde{\beta} = \left(X'\tilde{K}^{-1}X\right)^{-1}X'\tilde{K}^{-1}Y$ is the maximum likelihood estimate of the regression coefficients.

In kriging, the predicted error at any site s_0 is a weighted average of the regression residuals u from the estimation sample.

$$\hat{u}(s_0) = w'u \qquad (3.7)$$

The weights are chosen to minimize the expected value of the squared difference between the actual and predicted errors at site s_0.

$$MinF = E\{[u(s_0) - \hat{u}(s_0)]^2\} = E[u(s_0)^2 - 2u(s_0)\hat{u}(s_0) + \hat{u}(s_0)^2]$$

$$\begin{aligned} &= \sigma^2 - 2E[u(s_0)w'u] + E[w'uu'w] = \sigma^2 - 2w'q(s_0) + w'\Omega w \\ &= \sigma^2[1 - 2w'k(s_0) + w'\tilde{K}w] \end{aligned} \tag{3.8}$$

Where $u(s_0)$ is the actual error at site s_0, $\hat{u}(s_0)$ is the predicted error at site s_0, w is the vector of weights, $q(s_0)$ is the covariance vector between the observation at site s_0 and the estimation data, $k(s_0)$ is the correlation vector between the observation at site s_0 and the estimation data. The correlation vector can be obtained by the correlation functions. Take the partial derivatives of Equation (3.8) with respect to the vector w, and set these equal to zero, and solve for w which will minimize F.

$$\frac{\partial F}{\partial w} = \sigma^2[-2k(s_0) + 2\tilde{K}w] = 0 \tag{3.9}$$

Then we have $w = \tilde{K}^{-1}k(s_0)$ (3.10)

Here, $\tilde{K}^{-1}$ can be obtained by substituting the ML estimate of b_1 and b_2 into the correlation function and applying this function to the data locations.

3.3 Criteria for Evaluating Model Performance

First, the value of log likelihood function (VOF, Equation (3.5)) is used to determine the significance of the spatial correlations: twice the difference between VOF of the OLS and ML results is distributed as a chi-squared random variable with one degree of freedom.

In our study, one of the objectives is to evaluate model performance in modeling spatial structure for the multiunit Singapore condominium market among the three correlograms, and between the 2STAR model and the geostatistical model. The model performance is assessed primarily on the basis of out of sample prediction accuracy, which is determined by four statistics: the Mean Square Error (MSE), Theil's U, Mean of Absolute Value of Errors(Mean ABE), and Median of Absolute Values(Median ABE).

MSE is calculated as

$$MSE = \frac{\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}{N} \tag{3.11}$$

And Theil's U is calculated as

$$U = \frac{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} Y_i^2} + \sqrt{\frac{1}{N}\sum_{i=1}^{N} \hat{Y}_i^2}} \tag{3.12}$$

Where, N is the number of transactions in prediction sample. Y_i is the price value, and $\hat{Y}_i$ is the predicted price value.

3.4 Operational Considerations

In our study, the whole data set is partitioned into eleven subsets from July of a subject year to June of the next year for two reasons. One is to avoid the heavy computational burden. As equation (3.5) shows, the log likelihood function contains both the inverse and the determinant of the correlation matrix K. The sample size in our study is 26,446, and the correlation matrix will be of order 26,446. Without using sparse matrix technique (and the correlation matrix is not sparse), it is computationally intensive to invert such a large matrix. The other is that it is not reasonable to expect the implicit prices and the spatial relationship to be constant over the entire study period, so we expect that the partition might be helpful to reflect variations in implicit prices and in spatial relationship over time.

Compared with the 2STAR model, one limitation of the geostatistical model is that the model itself can not handle with spatial autocorrelation and temporal autocorrelation in the same time. In order to capture the temporal trend of the whole housing market, quarterly time dummy variables are included in equation (3.1) as independent variables.

The 2STAR model employed in the comparison procedure follows the Sun *et al.*'s (2005) procedure without any change, please refer it for details.

Since Matlab is powerful software on mathematics and scientific computing, especially for matrix, all the programs are written with Matlab in our study. The programs for the geostatistical model are presented at Appendix A.

Chapter 4 Results and Analysis

4.1 Introduction

In this chapter, we will first introduce the Singapore condominium market and then describe the working data used in our study, which is followed by the empirical results from the traditional hedonic model. An evaluation of model performance will be carried out among three geostatistical models in Section 4.4, and a comparison of performance between the geostatistical and 2STAR models is conducted in Section 4.5.

4.2 Data Collection

Singapore Condominium Market

In Singapore, the 86 per cent of the population lives in public housing units built by the Housing and Development Board(HDB), while the 14 per cent stays in private residence. The condominiums account for the biggest part of the private residential properties. Figure 4.1 shows the distribution of available private residential properties, which is based on the first quarter of 2006 URA statistics. From this figure, we can see that condominium units account for 45% of the total available private residential stock in Singapore. Apartments, terraced houses and semi-detached houses or bungalows are 26%, 16% and 9% of the stock respectively. Therefore, it is necessary and important to do a detailed research on the condominium market, for it can represent the Singapore private housing market.

Figure 4.1: Distribution of available private residential properties (As of the end of the 1st quarter 2006)[3]

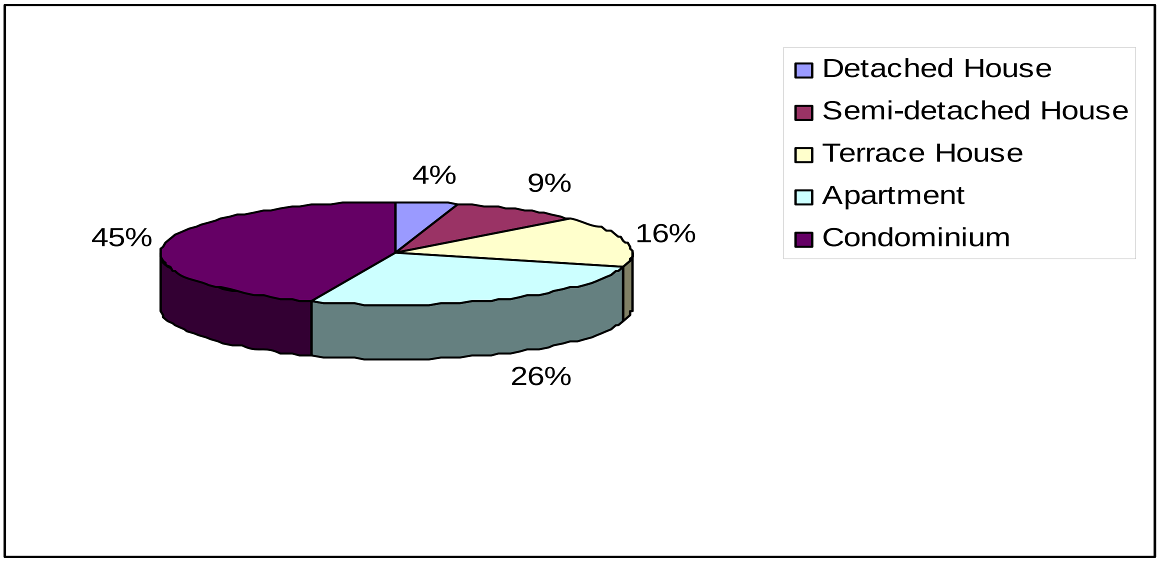
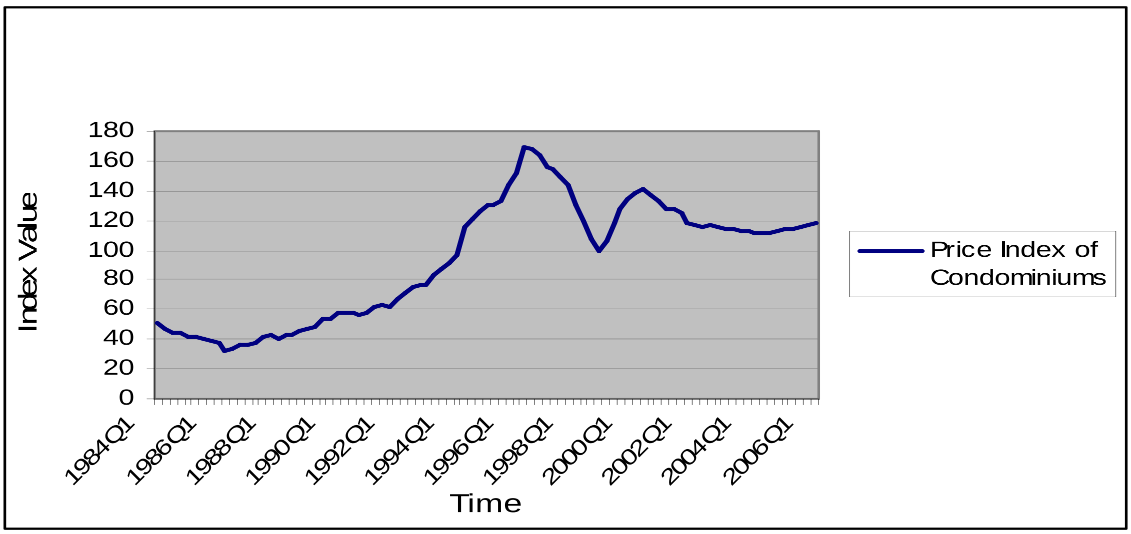


Figure 4.2 Property price index of the Singapore condominium market



Singapore condominium price is closely related to Singapore economy and government

[3] Condominium and apartment are two property types that make up the non-landed private residential property sector in Singapore. A condominium is a strata-titled development with a full-range of facilities provided within a minimum land area of 0.4 hectares, apartment are also multi-unit project built on smaller parcel of land.

macro-control as well. Figure 4.2 shows a property price index for the Singapore condominium market released by URA, the national planning authority of Singapore.

From early 1984 to the middle of 1986, Singapore economy suffered a recession and a negative GDP growth rates for the first time. The figure shows the condominium price experienced downward trend during this period. In order to stimulate the troubled real estate market, Singapore government took some measures in 1986. So the condominium price ceased to decline and showed signals of recovery from then on. In the next ten years, along with rapid economic growth of Singapore and investment and speculation from foreigners, the condominium market experienced the strongest boom and the condominium price reached a historic high record in the second quarter of 1996. Facing the over-heated real estate market, Singapore government took serious measures to deflate the speculative bubble in the whole real estate market. As a result, the condominium prices began to fall from the third quarter of 1996. The downward trend was worsen by the Asian Financial Crisis happened in July 1997. After Singapore government introduced some measures to activate the slump real estate market, the condominium prices began to soar from the beginning of 1999. However, the over supply of new houses in Singapore, the fallout of the stock market in the world led to the condominium price downward again from the middle of 2000. From the third quarter of 2004, the condominium market showed signals of recovery and entered a period of adjustment.

Transaction of condominium units could take place through directly from developers before its completion as well as through a secondary market, the resale market. Presale activities are quite common in Singapore. Data shows the average development managed to sell 54.3% of the units released for sale within three months if its initial launch (Ooi, 2005). Apart from structural and neighborhood characteristics, the prices of presale condominium units are affected by the behavior of the developers (Ooi, 2005). At the same time, the resale condominium market in Singapore is also active, with more than seven transactions per day from July of 1992 to June of 2003. Different from the price of presale condominium units, which is partly affected by the characteristics of developers, the price of resale units is determined by the market itself. Therefore, it is necessary to estimate the presale condominium market and the resale condominium market separately. In our study, we focus on the resale condominium market.

Data Description

The original data are obtained from an online real estate database called Reallink. They are five types of information:

1: Transaction Price;

2: Details of the transaction property, including House number (indicating the Level and Unit), Street Name, Project Name and Postal Code;

3: Floor Area and Construction Data;

4: Land Tenure of Property;

5: Contract and Transfer Dates of the transaction.

After deleting the transactions which did not have accurate information on the postal code, transaction price, age, floor level of the unit etc, we get a dataset of 26,446 transactions from July of 1992 to June of 2003. These transactions come from 1223 buildings in 695 condominium projects.

Next, our dataset is extended by adding information on condominium and neighborhood attributes. Of the adding information, detailed information of condominium facilities (whether the project has gym, swimming pool, etc) is obtained from “Guide to Private Residential Properties in Singapore”, published by Jones Lang Lasalle, a worldwide property research and consultant company. In addition, the X-Y Cartesian coordinates of transactions and other location-based neighborhood facilities are obtained from Virtual Map Singapore Pte., Ltd by referring to postcode for each building block. In Singapore, one building block corresponds to one postcode. The X-Y Cartesian coordinate for each house unit is the geostatistical information in our dataset, which allow us to compute linear distance between two buildings, between each building to various neighborhood facilities such as MRT station, CBD, schools and so on. As a result, each transaction is related to the variables indicating the full address, hedonic characteristics, condominium project characteristics, neighborhood amenities as well as the details associated with the sales of the transaction. Table 4.1 provides the definition of hedonic variables which are used in our study, and Table 4.2 presents the statistics of some key variables.

Out of all the condominium units in the dataset, 64.4% is either 999-year leasehold or freehold, 55.2% has barbecue facilities, 78% has car park, 49.8% has Gymnasium, 12.4% has Jacuzzi, 35.8% has fitness, 77.6% has playground, 91.9% has swimming pool , 75.1% has tennis courts and 94.2% has security facility.

Table 4.1 Variable definition

Variable	Description
Price	Dwelling transaction price
FLRAREA	Floor area in each condominium flat(unit: sqm)
AGE	The age of the condominium project(unit:year)
LEVEL	The floor level of the flat
FREEHOLD	Dummy variable, 1 indicating freehold or 999 leasehold; 0 indicating else
Dis_PRI1	Linear distance to 1st nearest top 30 primary school(unit:km)[4]
Dis_PRI2	Linear distance to 1st nearest top 30 primary school(unit:km)
Dis_SEC	Linear distance to 1st nearest top 10 secondary school(unit:km)[5]
Dis_JC	Linear distance to 1st nearest top 10 junior college(unit:km)[6]
MRT	Linear distance to the nearest MRT station(unit:km)[7]
CBD	Linear distance to the CBD(unit:km)[8]
DIS_RC	Linear distance to the nearest regional center[9]
BBQ	Dummy variable, 1 if project has barbecue lots and 0 of not
CAR PARK	Dummy variable, 1 if project has car park and 0 of not
GYM	Dummy variable, 1 if project has gym and 0 of not
JACUZZI	Dummy variable, 1 if project has jacuzzi and 0 of not
FITNESS	Dummy variable, 1 if project has fitness center and 0 of not
MINIMART	Dummy variable, 1 if project has minimart and 0 of not
MPH	Dummy variable, 1 if project has mph and 0 of not
PLAYGROU	Dummy variable, 1 if project has playground and 0 of not
SAUNA	Dummy variable, 1 if project has sauna and 0 of not
SWIMMING	Dummy variable, 1 if project has swimming pool and 0 of not
TENNIS	Dummy variable, 1 if project has tennis courts and 0 of not
SQUASH	Dummy variable, 1 if project has squash courts and 0 of not
WADING	Dummy variable, 1 if project has wading 0 of not
SECURITY	Dummy variable, 1 if project has security and 0 of not
Others	Dummy variable, 1 if project has other facilities and o of not
TOTALUNI	Total Number of dwelling units in the project

[4] The top 30 primary schools are defined by Singapore Ministry of Education and rank remains constant from 1990 to 2003. The coordinate of these schools are supplied by Virtual Map Singapore Pte.,Ltd.

[5] The top 10 secondary schools are defined by Singapore Ministry of Education and rank remains constant from 1990 to 2003. The coordinate of these schools are supplied by Virtual Map Singapore Pte.,Ltd.

[6] The top 10 Junior colleges are defined by Singapore Ministry of Education and rank remains constant from 1990 to 2003. The coordinate of these schools are supplied by Virtual Map Singapore Pte.,Ltd.

[7] The coordinate of all MRT station are supplied by by Virtual Map Singapore Pte.,Ltd

[8] The coordinate of Centerpoint at CBD area is used as reference, which supplied by Virtual Map Singapore Pte.,Ltd.

[9] Besides CBD, there are three regional centers in Singapore .

Table 4.2 Statistics of some key variables

Variable[10]	Mean	Std[11]
PRICE(S$)	992132.4	673834.6
LEVEL	7.634765	6.200157
AGE	8.437363	5.395435
FLRAREA(sqm)	156.4547	73.95235
DIS_PRI1(km)	1.352626	1.166419
DIS_PRI2(km)	2.133206	1.264307
DIS_SEC(km)	2.033095	1.301755
DIS_JC(km)	3.652817	2.30795
MRT(km)	1.477166	0.833358
CBD(km)	8.176599	3.803499
TOTALUNI	331.4643	299.2598

Table 4.3 presents the quarterly average transaction prices by per square meter and the number of transactions between July of 1992 and June of 2003. From this table, it is apparent that the average price of condominium increased from the third quarter of 1992 and reached the peak value at the second quarter of 1996. Then the price dropped downward and reached trough at the fourth quarter of 1998. The price increased one quarter later and then decreased five quarters later. In one word, the dynamics of housing price is consistent with the previous discussion of phases of the condominium price index.

Table 4.3 Average housing prices and the number of transactions over time

Quarter	Ave.Price(S$)[12] (Std)	Number of Transactions	Quarter	Ave.Price(S$) (Std)	Number of Transactions
Q3 1992	3823.1 (955.6)	160	Q1 1998	6575.7 (1869.6)	176
Q4 1992	3845.1 (942.0)	371	Q2 1998	6016.8 (1939.4)	383

[10] The definition of all variables is given in Table 3.1
[11] It denotes standard deviation
[12] Singapore Dollar per square meter

Q1 1993	4080.1 (976.6)	447	Q3 1998	5669.6 (1898.6)	296
Q2 1993	4330.7 (1008.6)	795	Q4 1998	5376.3 (1672.4)	523
Q3 1993	4486.3 (1058.0)	916	Q1 1999	5611.8 (1717.1)	824
Q4 1993	4645.4 (1152.4)	763	Q2 1999	6398.6 (2022.5)	1810
Q1 1994	5103.5 (1271.9)	656	Q3 1999	6940.6 (2161.2)	1182
Q2 1994	5580.9 (1450.3)	1004	Q4 1999	7355.2 (2505.9)	860
Q3 1994	6135.2 (1600.4)	716	Q1 2000	7736.6 (2628.9)	947
Q4 1994	6523.2 (1906.4)	783	Q2 2000	7344.4 (2412.4)	745
Q1 1995	6658.4 (1809.3)	437	Q3 2000	6991.7 (2377.4)	671
Q2 1995	7490.9 (2880.2)	850	Q4 2000	7576.6 (2743.3)	588
Q3 1995	7073.3 (2398.0)	680	Q1 2001	6445.3 (2164.0)	391
Q4 1995	7008.6 (2068.5)	649	Q2 2001	6226.3 (2182.4)	469
Q1 1996	7665.7 (2411.8)	770	Q3 2001	6193.9 (2372.9)	394
Q2 1996	8357.4 (2986.3)	861	Q4 2001	5819.6 (1846.7)	123
Q3 1996	7752.3 (2729.2)	331	Q1 2002	6015.3 (1975.9)	605
Q4 1996	7887.0 (2848.0)	306	Q2 2002	6168.0 (2243.1)	634
Q1 1997	7929.1 (3060.1)	387	Q3 2002	5912.9 (3771.8)	532
Q2 1997	7736.2 (2610.8)	437	Q4 2002	5789.3 (1869.1)	544
Q3 1997	7610.0 (2513.6)	378	Q1 2003	5242.3 (1650.5)	372
Q4 1997	6963.5 (2105.0)	267	Q2 2003	5310.4 (1771.7)	413
			Whole Period	6372.4 (2430.3)	26,446

4.3 Empirical Results of Traditional Hedonic model

In the traditional hedonic model, the dependent variable is the log price of housing unit and the independent variables consist of four types. The first type is structural characteristics of the housing unit, such as level, age, floor area and freehold. The second type is dummy variables indicating characteristics related to the whole condominium project. For example, whether there is a facility such as swimming pool or tennis court in this condominium project. The third is location-related variables such as the distance to primary school, the distance to CBD and so on. The fourth is quarterly time dummy variables reflecting house price changes along with time.

Among those structural variables, AGE is expected to relate negatively to property value as age of a property is used as a proxy for residential depreciation in terms of deterioration and obsolescence. In a tropical island like Singapore, house units of higher levels command a premium because of better view and ventilation, so that the sigh of LEVEL is expected to be positive. Similar to many house price studies, the expected sign of FLRAREA is positive. The expected sign for Freehold is positive as usually people prefer freehold properties to leasehold properties. House units situated at a condo with more facilities fetch higher selling prices, so that the expected signs for all project-related dummy variables except TOTALUNIT are positive. The public school, MRT station, CBD, regional center are all considered as positive externalities and the expected signs for the distance to them are negative.

At first, we run the hedonic model including all listed independent variables. Then we run the model again after omitting the independent variables with wrong signs or insignificant value in the prior run. This repeat procedure can not stop until all the independent variables are significant and of expected signs. The results from the first run and the final run are presented on Table 4.4

Table 4.4 Traditional hedonic model estimates

Variable[13]	**Expected Sign**	**Beta (First Run)**	**Beta (Last Run)**
CON	+	12.1090** [14] (0.0233) [15]	12.3960** (0.0135)
LEVEL	+	0.0049** (0.0002)	0.0074** (0.0002)
AGE	-	-0.0093** (0.0003)	-0.0115** (0.0003)
FLRAREA	+	0.0040** (2.04E-5)	0.0043** (2.17E-5)
FREEHOLD	+	0.3102** (0.0034)	0.3228** (0.0033)
DIS_PRI1	-	0.0269** (0.0023)	
DIS_PRI2	-	0.0109** (0.0027)	
DIS_SEC	-	-0.0321** (0.0015)	-0.0385** (0.0013)
DIS_JC	-	-0.0282** (0.0008)	-0.0205** (0.0008)
MRT	-	-0.0297** (0.0018)	-0.0551** (0.0018)
CBD	-	-0.0008 (0.0010)	
DIS_RC	-	0.0344** (0.0011)	
BBQ	+	0.0120** (0.0033)	0.0219** (0.0033)
CARPARK	+	0.0035 (0.0036)	0.0217** (0.0037)
GYM	+	0.0745**	

[13] The definition of all variables are given by Table 3.1
[14] ** denotes the significance at 0.05 level.
[15] Standard error is in parentheses.

		(0.0034)	
JACUZZI	+	0.0465^{**} (0.0047)	
FITNESS	+	0.0067 (0.0039)	
MINIMART	+	-0.0008 (0.0045)	
MPH	+	0.0232^{**} (0.0034)	0.0597^{**} (0.0034)
PLAYGROUD	+	-0.0076 (0.0039)	
SAUNA	+	-0.0063 (0.0034)	
SQUASH	+	0.0310^{**} (0.0041)	0.0386^{**} (0.0042)
SWIMMING	+	0.0806^{**} (0.0067)	0.0592^{**} (0.0071)
TENNIS	+	0.0226^{**} (0.0042)	-0.0060^{**} (0.0043)
WADING	+	0.0319^{**} (0.0037)	0.0244^{**} (0.0038)
SECURITY	+	0.0608^{**} (0.0073)	0.0566^{**} (0.0078)
OTHERS	+	-0.0165^{**} (0.0041)	
TOTALUNI	-	0.0000 (7.16E-6)	
1992Q4		0.0097 (0.0200)	
1993Q1		0.0761^{**} (0.0195)	0.0765^{**} (0.0149)
1993Q2		0.1366^{**} (0.0183)	0.1312^{**} (0.0131)
1993Q3		0.1842^{**} (0.0181)	0.1901^{**} (0.0127)
1993Q4		0.2010^{**} (0.0184)	0.2089^{**} (0.0132)
1994Q1		0.2799^{**} (0.0187)	0.2901^{**} (0.0136)
1994Q2		0.3579^{**} (0.0180)	0.3720^{**} (0.0125)
1994Q3		0.4447^{**} (0.0185)	0.4618^{**} (0.0133)
1994Q4		0.4967^{**} (0.0184)	0.5059^{**} (0.0131)
1995Q1		0.5057^{**}	0.5241^{**}

		(0.0196)	(0.0151)
1995Q2		0.5961** (0.0183)	0.6179** (0.0129)
1995Q3		0.5881** (0.0186)	0.5943** (0.0135)
1995Q4		0.5982** (0.0187)	0.6017** (0.0136)
1996Q1		0.6525** (0.0184)	0.6685** (0.0132)
1996Q2		0.7192** (0.0183)	0.7544** (0.0129)
1996Q3		0.6876** (0.0204)	0.6981** (0.0163)
1996Q4		0.6657** (0.0207)	0.6688** (0.0167)
1997Q1		0.7106** (0.0199)	0.7131** (0.0156)
1997Q2		0.6758** (0.0196)	0.6784** (0.0151)
1997Q3		0.6737** (0.0200)	0.6827** (0.0157)
1997Q4		0.6134** (0.0212)	0.6139** (0.0175)
1998Q1		0.5092** (0.0232)	0.5176** (0.0203)
1998Q2		0.4301** (0.0200)	0.4378** (0.0157)
1998Q3		0.3165** (0.0208)	0.3189** (0.0169)
1998Q4		0.2610** (0.0192)	0.2723** (0.0144)
1999Q1		0.3131** (0.0183)	0.3257** (0.0130)
1999Q2		0.4419** (0.0175)	0.4494** (0.0116)
1999Q3		0.5253** (0.0179)	0.5339** (0.0123)
1999Q4		0.5740** (0.0183)	0.5890** (0.0129)
2000Q1		0.6523** (0.0182)	0.6654** (0.0127)
2000Q2		0.5677** (0.0185)	0.5864** (0.0133)
2000Q3		0.5458** (0.0187)	0.5649** (0.0136)
2000Q4		0.6132**	0.6171**

		(0.0189)	(0.0140)
2001Q1		0.4803** (0.0199)	0.4819** (0.0156)
2001Q2		0.4540** (0.0195)	0.4638** (0.0149)
2001Q3		0.4406** (0.0199)	0.4487** (0.0156)
2001Q4		0.3660** (0.0255)	0.3929** (0.0234)
2002Q1		0.3983** (0.0189)	0.4166** (0.0140)
2002Q2		0.4311** (0.0189)	0.4475** (0.0139)
2002Q3		0.3866** (0.0192)	0.4070** (0.0144)
2002Q4		0.3956** (0.0192)	0.4133** (0.0144)
2003Q1		0.3186** (0.0201)	0.3395** (0.0159)
2003Q2		0.3120** (0.0199)	0.3377** (0.0155)
$\overline{R}^2$		0.8174	0.7787
Mean absolute error		0.1534	0.1706
Median absolute error		0.1152	0.1322
Sum square of error		1179.6	1430.2
First lag autocorrelation[16]		0.7546	0.7906
Number of observations		26,446	26,446
Number of variables		71	58

The empirical results shows the coefficients of all structural variables are significant and of expected signs in all runs. Some project-related variables are either insignificant or have unexpected signs in the first run, such as car park, minimart, playground etc, indicating it may be not reasonable to treat these project-related variables as dummies.

With regard to the neighborhood-related variables, it is unsuccessful to relate them to housing prices in the models. The coefficients of distance to primary schools have

[16] It refers to the correlation of the residual with its nearest residual

positive signs indicating condos with further distance to primary schools fetch higher price, which may be caused by the multicollinearity between the distance to primary schools and the distance to MRT as most primary schools are near MRT stations in Singapore. In addition, the distance to CBD has expected sign but is not significant in the first run. It is probably due to the fact that Singapore is a polycentric city rather than a homocentric city. As we know, besides the CBD, there are three regional centers which play an important role to decentralize the functions and services of CBD in Singapore.

In both runs, the first lag spatial autocorrelations among the residuals are more than 0.75, indicating the residuals may still be spatially auto correlated and some spatial information may not be captured by the independent variables in the model. As a result, the OLS estimates of hedonic coefficients are inefficient and the statistics inferences based on them are invalid. Predicted house price based on these estimates will be also inaccurate. However, efficient coefficients can be obtained through both lattice models and geostatistical models as we discussed in Chapter 2.

4.4 Empirical Results of Geostatistical Models

When applying the geostatistical models, we divide the whole data set into eleven subsets, from July of a subject year to June of the next year respectively. In each subject, randomly selected 500 observations are used as a prediction sample, and the remaining observations consist of an estimation sample. In total, we have eleven pairs of estimation and prediction samples.

Table 4.5 presents the estimation and prediction results from the three geostatistical models and the hedonic model (OLS) during the period from July of 1995 to June of 1996. The empirical results of other subsets are presented in Table 4.1 to Table 4.10 in Appendix B. Note that the OLS model listed here includes the same independent variables as the three geostatistical models. Strictly speaking, it is not correct to compare the geostatistical model estimates except the VOFs with the OLS results in this fashion since this kind of hedonic model is obviously mis-specified. In all subsets, the estimated coefficients of LEVEL, FLRAREA, and FREEHOLD are significant with constant positive signs as they are all desired attributes of a house unit. The estimation coefficients of AGE are negative as what it is expected.

From the VOFs in the tables, all three estimated correlograms show significant spatial correlations. Among them, the negative exponential correlogram has the largest VOF in all subsets, indicating it must outperform other two correlograms in estimating the spatial correlation of prices for the Singapore condominium resale market.

Figure 4.2 shows a graph of the three estimated correlograms and the empirical correlograms[17] in the period from July of 1995 to June of 1996. The correlogram graphs of other subsets are shown in Figure 4.1 to Figure 4.10 in Appendix B. These graphs illustrate the fitness of estimated correlograms to the empirical correlogram. Again, the negative exponential correlogram fits the empirical correlogram best among three correlograms across all subsets.

---

[17] The experimental correlogram is produced by finding the correlation between all observations separated by a distance of 0-0.5 km. This is one point. Then you find the correlation between all observations separated by.5-1 km.

Table 4.5 Empirical results from July of 1995 to June of 1996

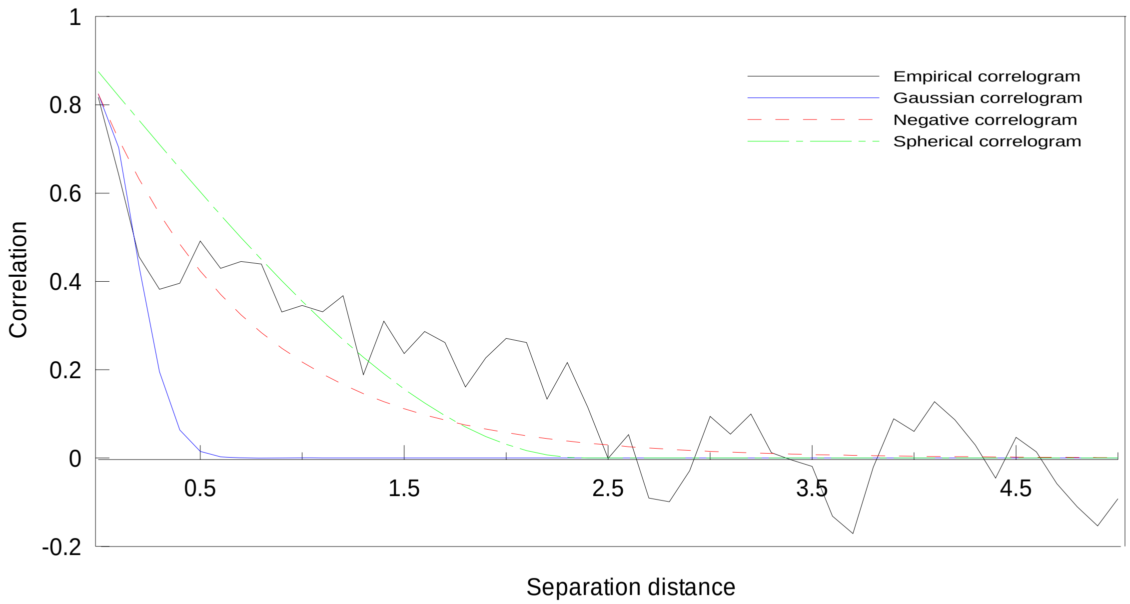
Period		July 1995 – June1996			
In sample size		2460			
Out sample size		500			
	VARIABLES	OLS	GSS	NEG	SPH
ESTIMATION	CONSTANT	13.1680** (0.0209)	13.0190** (0.0012)	12.9840** (0.0017)	12.9800** (0.0023)
	LEVEL	0.0040** (0.0009)	0.0065** (2.8E-7)	0.0066** (2.6E-7)	0.0066** (2.6E-7)
	AGE	-0.0251** (0.0013)	-0.0106** (2.9E-6)	-0.0123** (2.7E-6)	-0.0128** (2.4E-6)
	FLRAREA	0.0057** (0.0001)	0.0038** (2.9E-9)	0.0037** (2.7E-9)	0.0037** (2.7E-9)
	FREEHOLD	0.2688** (0.0052)	0.3248** (0.001)	0.2950** (0.0008)	0.3000** (0.0007)
	T2[18]	0.0070 (0.0166)	0.0207** (0.0001)	0.0200** (0.0001)	0.0210** (0.0001)
	T3	0.0846** (0.0159)	0.0546** (4.9E-5)	0.0552** (4.6E-5)	0.0561** (5.2E-5)
	T4	0.1600** (0.0155)	0.1232** (4.8E-5)	0.1235** (4.5E-5)	0.1245** (4.7E-5)
	VOF	-6416.7	-4685.8	-4576	-4590.8
	b1		0.825	0.8250	0.875
	b2		0.25	0.7500	2.375
PREDICTI-ON	MSE	0.1053	0.0157	0.0148	0.0151
	Theil's U	0.0038	0.0006	0.0005	0.0005
	Mean ABE	0.2310	0.0775	0.0752	0.0761
	Median ABE	0.1778	0.0538	0.0531	0.0532

From theses figures, it is apparent that the spherical correlogram has the highest intercept and falls off rapidly with separation distance till zero at a certain distance and afterwards. The Gaussian correlagram gives more weight to close house units and falls off most rapidly with separation distance. The negative exponential correlogram falls off more

[18] T2, T3, T4 are the quarterly dummy variables.

slowly than other two correlograms.

Figure 4.3 Correlograms of period from July of 1995 to June of 1996



The prediction parts of Table 4.2 and of Table 4.1 to Table 4.9 in Appendix B present four statistics for out of sample predictions obtained from the three geostatistical models. As statistics obtained from the OLS are not appropriate to be compared with the geostatistical models, we present the prediction results (Table 4.11 in Appendix B) obtained from the OLS which includes all location-related independent[19] variables. From these tables, we find that the four statistics of all the geostatistical models are constantly much lower than those of the OLS across four diagnostic statistics across all subsets, indicating the former predict much better than the latter. Among three geostatistical models, the value of mean absolute residuals and the Theil's U of the negative exponential correlogram is constantly lower than those of the other two geostatistical models. The value of median absolute value and the MSE of the negative

[19] After series of trial and error experiments, we find that the hedonic model including more spatial independent variables predict better.

exponential correlogram is lower than those of other two models in nine and ten of eleven subsets respectively. These imply that the negative exponential correlogram predicts better than the Gaussian and spherical correlograms.

In conclusion, all geostatistical models predicts much better than the traditional hedonic model. Among them, the negative exponential correlogram outperforms the other two correlograms both in estimating the spatial structure and in predicting house prices for the condominium resale market in Singapore. Therefore, we allow the negative exponential correlogram as a representative for the geostatistical models to compare with the 2STAR model in next section.

4.5 Comparison of Performance between Geostatistical model and 2STAR model

Geostatistical models and lattice models are alternative approaches in modeling spatial autocorrelation in real estate markets. In order to evaluate the performance of two models for a multiunit housing market, 2STAR specified by Sun *et al.* (2003) has been applied to the same estimation and prediction samples. The spatial and temporal lag is 16 and 20 respectively, both of which are determined by Geographically Weighted Regression (GWR). The regression results from the 2STAR model are presented in Table 4.12 in Appendix B.

Comparison of Estimations

Regarding the estimation, a comparison between the negative exponential estimates and the 2STAR estimates shows a consistent result that estimates from the former produce smaller standard errors across four structural coefficients and across all subsets. This fact seems to imply that the negative exponential estimation can more effectively correct the inefficient OLS estimation resulting from spatial autocorrelation than the 2STAR model.

One of the STAR's advantages is that it can handle with spatial and temporal autocorrelation at the same time. Moreover, the 2STAR model can be applied to the whole data set with a sparse matrix technique. So we estimate the eleven estimation samples together with the 2STAR model and then to predict each prediction sample. The estimation result is presented in Table 4.13 in Appendix B. A comparison of this table with the tables which present the estimation results from geostatistical models shows that the standard error of all coefficients of the structure variables except FLRAREA are still larger than those of the geostatistical model. And the standard error of the coefficient of FLRAREA is zero and smaller than that in the geostatistical model, which seems to imply some inefficient OLS estimates with an spatial autocorrelated error term may be corrected by the 2STAR with a larger estimation sample.

Comparison of predictions

The performance of the two spatial models is primarily assessed by their prediction ability, which is determined by the statistics presented in Table 4.6. From this table, we find that the four statistic of the 2STAR with estimation sample of the whole dataset are constantly lower than those of the 2STAR with estimation sample of each dataset except in the periods from July of 1996 to June of 1997 and from July of 1998 to June of 1999, indicating that increasing the sample size will improve the prediction ability of the 2STAR model.

As compared to the 2STAR model with estimation sample of the whole dataset, the negative exponential correlogram at least results in a 34% reduction in the mean of squared errors, a 30% reduction in the Theil's U, a 24% reduction in the mean of absolute errors, and a 21% reduction in the median of absolute errors in each subset.

As compared to the 2STAR model with estimation sample of each dataset, the negative exponential correlogram at least results in a 24% reduction in the mean of squared errors, a 18% reduction in the Theil's U, a 18% reduction in the mean of absolute errors, and a 20% reduction in the median of absolute errors in each subset.

Table 4.6(a) Prediction errors of two spatial models

	MSE			Theil's U		
Period	NEG	2STAR	2STAR(1)	NEG	2STAR	2STAR(1)
Jul92-Jun 93	0.0132	0.0246	0.0703	0.0005	0.0009	0.0026
Jul93-Jun94	0.0136	0.0321	0.0315	0.0005	0.0012	0.0012
Jul 94-Jun95	0.0159	0.0353	0.0340	0.0006	0.0013	0.0012
Jul 95-Jun96	0.0148	0.0359	0.0371	0.0005	0.0013	0.0013
Jul 96-Jun97	0.0182	0.0367	0.0354	0.0007	0.0013	0.0013
Jul 97-Jun98	0.0237	0.0461	0.0315	0.0009	0.0017	0.0011
Jul 98-Jun99	0.0234	0.0355	0.0491	0.0009	0.0013	0.0018
Jul 99-	0.0223		0.0330	0.0008		0.0012

Jun00		0.0373			0.0014	
Jul 00-Jun01	0.0201	0.0335	0.0317	0.0007	0.0012	0.0012
Jul 01-Jun02	0.0208	0.0358	0.0328	0.0008	0.0013	0.0012
Jul 02-Jun03	0.0211	0.0334	0.0332	0.0008	0.0012	0.0012

Table 4.6(b) Prediction errors of two spatial models

	Mean Absolute Error			Median Absolute Error		
Period	NEG	2STAR	2STAR(1)	NEG	2STAR	2STAR(1)
Jul 92-Jun93	0.0804	0.1179	0.1432	0.0577	0.0949	0.1026
Jul 93-Jun94	0.0870	0.1346	0.1213	0.0664	0.1011	0.0872
Jul 94-Jun95	0.0870	0.1400	0.1269	0.0649	0.1126	0.0966
Jul 95-Jun96	0.0752	0.1401	0.1351	0.0531	0.1088	0.0986
Jul 96-Jun97	0.0966	0.1382	0.1339	0.0704	0.1002	0.0936
Jul 97-Jun98	0.1071	0.1661	0.1294	0.0784	0.1365	0.0982
Jul 98-Jun99	0.0944	0.1436	0.1494	0.0631	0.1134	0.1177
Jul 99-Jun00	0.0939	0.1395	0.1285	0.0639	0.0968	0.0956
Jul 00-Jun01	0.0930	0.1309	0.1288	0.0656	0.0930	0.0995
Jul 01-Jun02	0.0976	0.1454	0.1344	0.0681	0.1243	0.0985
Jul 02-Jun03	0.1011	0.1343	0.1398	0.0773	0.0991	0.1165

Discussion

In a multi unit housing market, the building-related spatial autocorrelations may be stronger than the ones caused by the neighborhood effect as these units are confined to one specific location. Sun *et al.* (2005) argue that spatial effects can be divided into building effects and neighborhood effects in a multiunit housing market. Base on this argument, they develop the 2STAR model and find that it outperforms the STAR model in modeling the spatial structure for the Singapore condominium market. Their empirical result in turn provides some evidence for the argument.

However, we argue that the building effect termed by Sun *et al.* is not a kind of different spatial effect from the neighborhood effect. On the contrary, it is a kind of neighborhood effect related to one specific location, that is, a neighborhood effect at the building level. Since the correlation of a transaction with the transaction in the same building is very different from one but the distance between them is zero, the neighborhood effect at building level can be regarded as a nugget effect from the point view of geostatistics. Since the geostatistical model has a good performance both in estimation and in house price prediction, it is convincing that the nugget effect is captured effectively by the geostatistical model.

The nugget effect might be captured to some extent by a separate spatial weight matrix in the 2STAR model (the building effect matrix) as the 2STAR model performs better than the STAR model which is originally developed for the single family housing market (Sun *et al.*, 2005). However, a certain degree of multicollinearity is likely to be introduced into the model, for the building effect is a neighborhood effect at the building level. As a result, the estimates from the model might be inaccurate.

4.6 Summary

The analysis concludes that all the three geostatistical models can effectively model the spatial autocorrelation structure in the Singapore condominium resale market. Among the three geostatistical models, the negative exponential correlogram performs best. Compared with the 2STAR model, the negative correlogram can more effectively correct the inefficient OLS estimation resulting from spatial autocorrelation. Regarding the house price prediction, the negative correlogram has a better performance than the 2STAR model, which is consistent with the Monte Carlo experiment results (Dubin, 2003).

In addition, we argue that the building effect (Sun *et al.*, 2005) is a neighborhood effect at the building level, which can be regarded as a nugget effect from the point view of geostatistics. This argument is convincing, because the geostatistical model has a better performance both in estimation and in house price prediction than the 2STAR model.

Chapter 5 Model Performance in Constructing Indexes

5.1 Introduction

In this chapter, we will first compare and discuss the model performance in constructing building specific indexes between the geostatistical model and the 2STAR model. In what follows, we will discuss the performance of these two models in constructing aggregate indexes and compare them with the URA official indexes.

5.2 Index Construction

Both the geostatistical model and the 2STAR model can be used to construct building specific indexes to uncover the price dynamics at building level. When the dummy variables are included in the geostatistical model, it is easy to separate out a price index over time at any point on the spatial surface. Based on the empirical results from the negative exponential correlogram in Table 4.2 and Table 4.1-10 in Appendix B, quarterly building specific indexes of 1223 blocks have been constructed. A standard housing unit with an average size ($156m^2$), level (7) and age (8 years) is chosen. The tenure of the standard housing unit depends on the leasehold status of the subject building. The value appreciation between the base quarter (the third quarter of 1992) and the subject quarter is calculated and converted into indexes.

Similarly, the spatial-temporal filtering process in the 2STAR model enables it to construct building specific indexes (Sun *et al.*, 2005). Since the 2STAR model with estimation sample of the whole dataset predicts better than with estimation sample of each dataset, we use the former to construct building specific indexes. The standard unit is the same as used by the geostatistical model. Its value at the middle day of each quarter between the third quarter of 1992 and the second quarter is predicted by the estimation results presented in Table 4.13.

The building specific indexes derived from the two models can be used to generate an aggregate index, which is the weighted average of all building specific indexes in the subject area, with a weight of the number of transactions in each building.

The URA index is constructed using average transaction prices and the fixed base weighted Laspeyres formula. Prior to the fourth quarter of 1998, the index is a price relative of the current price per square meter compared with that of base year (=1990). From the fourth quarter of 1998, the price indexes were computed with weights based on the moving average method over the last twelve quarters, which means the weights in the price index are updated quarterly. Including the condominium price index of the whole Singapore Island, sub-indexes of four planning regions are also constructed in the same way. That is, Central region, East region, North-east region and West region.

Before we discuss and compare different indexes, it is necessary to introduce some criteria first for assessing the usefulness of these indexes. Mark and Goldberg (1984) state that a usable index would embody three characteristics:

- First, the index should have a sound theoretical and conceptual foundation.

- Second, the index should be administratively simple and not rely on expensive and awkward sampling or survey procedures. In other wordss, the index should rely on readily available data.
- Third, the index should be reasonably insensitive to temporal variations in the nature of transactions.

Since the building specific indexes in our study are derived from a combination of the hedonic model and the spatial model, which satisfy the first criteria. Moreover, the data for constructing indexes comes from Reallink database released by URA, the statutory body responsible for the urban planning and development of Singapore, which indicates that the second criteria is satisfied. Thirdly, all the building specific indexes derived have the quality-control characteristics, which meet with the third criteria. In conclusion, the building specific indexes derived in our study are reasonable.

5.3 Discussion on Building Specific Indexes

In order to illustrate the price level of a specific building across the Singapore condominium resale market, three projects located at the central region representing the luxury, moderate and economical condominium respectively have been chosen in our study. The general information of these three projects is presented in Table 5.1.

Table 5.1 General information of the selected condominium projects

Project Name	Tenure	Total Unit	Year of TOP	Average Price[20]	Block	No.of transactions[21]
Arcadia Garden	99 Year	164	1983	4739	B1	8
					B2	91
					B3	80
Yong An Park	Freehold	288	1986	7973	B1	82
					B2	68
					B3	48
					B4	7
					B5	38
The Claymore	Freehold	146	1986	14694	B1	66
					B2	158

Arcadia Garden

The Arcadia Garden is a relatively low price condominium, the average price per square meter of which is 4739 Singapore dollars. Figure 5.1(a) displays building specific indexes derived by the geostatistical model (GEO) in Arcadia Garden. For comparison, we also plot the URA Condominium Price Index for the whole area and the central region as well.

A careful examination of Figure 5.1(a) tells us three interesting findings:

Firstly, both the building specific indexes and the aggregate indexes reflect the major market cycles consisting of upward movements from the third quarter of 1992 to the third quarter of 1996 and from the first quarter of 1999 to the second quarter of 2000; downward movements from the fourth quarter of 1996 to the fourth quarter of 1998 and from the third quarter of 2000. Although the major cycles indicated by these indexes are similar, they are of different magnitudes. Prior to the third quarter of 1995, the released

[20] The unit is Singapore dollars per square meter
[21] The transactions occurred in our study period

URA indexes and the building specific indexes show close price movement. Afterwards, the URA released indexes indicate a generally higher price level than the building-specific indexes, no matter during a declining market or a recovery market, indicating unique neighborhood effects closely related to the project. We attribute this difference to the leasehold status of the Arcadia. As presented in Table 5.1, the Arcadia Garden is a 99-year leasehold condominium. As a result, if an appraiser or investor predicts the price movements of the Arcadia based on the URA index, he might get unsatisfactory results. So that it is necessary to construct building specific indexes to help appraisers and investors to better predict the price movements of a particular building rather than using a general index as proxy.

Secondly, the indexes of the three buildings show different price movements. The price movement of B3 is different from the price movement of B1 and B2, which can be considered as evidence of building-related neighborhood effect in a multiunit housing market. Other things being equal, the block with different view or orientation will command a premium or discount over other blocks in the same project.

Thirdly, the price movement of B1 is much closer to the price movement of B2 than to that of B3. Since there are only 8 transactions in B1 during the study period, the predicted value of the standard unit in this building will be primarily conditioned on the transactions in other buildings. The linear distance between B1 and B2 is 0.1013km, and the distance between B1 and B3 is 0.1960. According to the formula of the negative correlation function, the transactions in B2 may contribute more than those in B3 for constructing the price index of B1. Since the index of a building with few transactions is primarily affected by the transactions in the nearest building, the index might be biased. It is because two buildings with a short distance may follow different price dynamic paths with the neighborhood effect at the building level.

Figure 5.1(a) Building specific indexes derived by GEO in Arcadia Garden

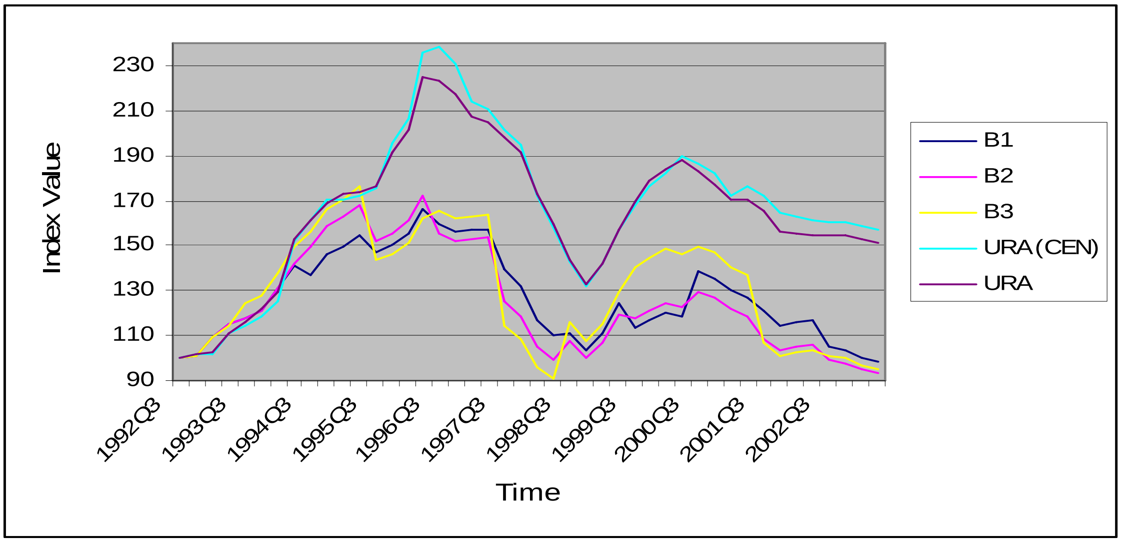


Figure 5.1(b) Building specific indexes in the Arcadia Garden

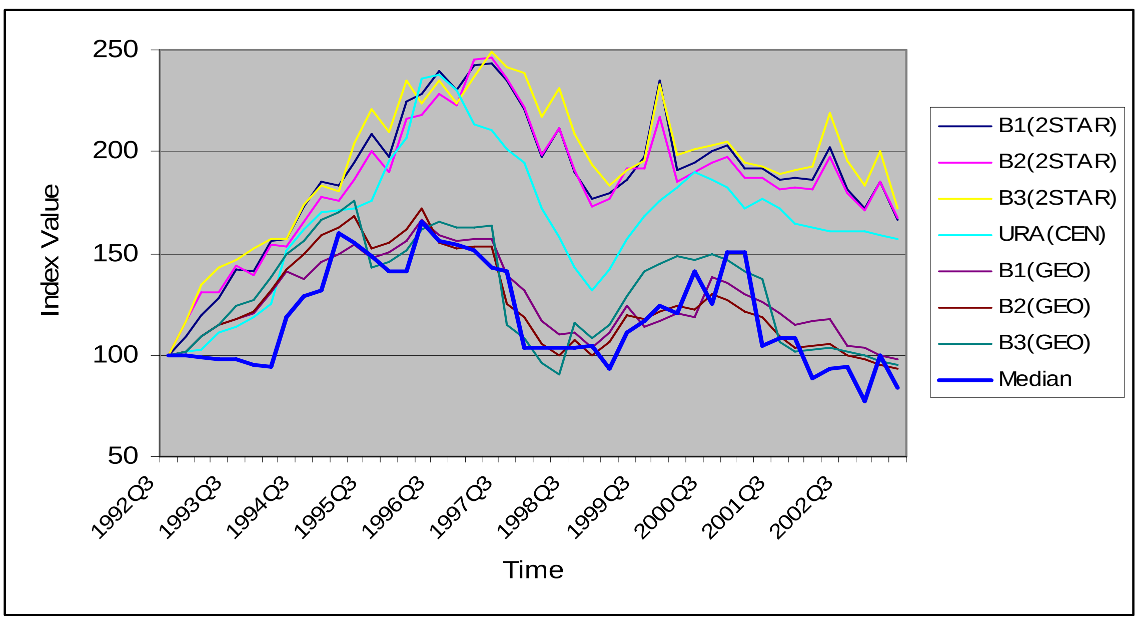


Figure 5.1(b) shows the building specific indexes derived from the geostatistical model and the 2STAR model. For comparison, we plot the URA index and the index derived from the quarterly median prices of the Arcadia as well. Compared with the URA released index, the indexes constructed by the 2STAR indicate a generally higher price level from the second quarter of 1997. Compared with the indexed from the geostatistical model, the indexes constructed by the 2STAR indicate a generally higher price level from the fourth quarter of 1994. However, the geostatistical model derived indexes fluctuate around the index from the quarterly median prices, indicating that they reflect the actual price movement of the buildings in the Arcadia. It is because the geostatistical model always has better performance in prediction than the 2STAR model.

Yong An Park

The Yong An Park is one of the moderate price condominiums in Singapore, the average price per square meter of which is 4739 Singapore dollars. Figure 5.2(a) and Figure 5.2(b) show the building specific indexes in Yong An Park derived by geostatistical and 2STAR model respectively. From these two figures, we find that all building specific indexes reflect a generally higher price level than the URA index, and that different buildings in the same project follow different price dynamic paths.

However, there are still some differences between the indexes derived by the two models. Firstly greater and more similar volatilities are exhibited among the indexes derived from the 2STAR than that from the geostatistical model. We attribute this difference to the temporal weight matrix in the 2STAR model. In particular, the fixed temporal lag equals 20, meaning that the predicted value of the standard unit at a particular quarter is conditioned on the prior 20 transactions cross the whole area. Since there are 7 transactions per day in average in the Singapore condo resale market, the 20 transactions may happen in a short period of 3 or 4 days. In other words, the temporal price movement in each quarter is represented by the transactions which were done in 3 or 4 days and across the whole area, which might result in great volatilities. Moreover, the predicted values of the standard unit in all buildings are conditioned on the same prior transactions, which certainly results in the similar volatilities among the building indexes.

Secondly, the index derived by the geostatistical model is more sensitive to outliers than that from the 2STAR model. As presented at Table 5.2, Building 4(B4) consists of town houses with floor area of more than 700 square meters. There are only 7 transactions in this building during our study period, and a transaction with abnormally low price occurs in the third quarter of 1998. This outlier is reflected in the price index of B4 derived by the geostatistical model. As observed in Table 5.2(a), the price movement of B4 is close to the price movements of displays close other four buildings over time except from the third quarter of 1998 to the first quarter of 1999. During this period, the index of B4 drops abruptly to an abyss. On the other hand, as observed in Figure 5.2(b), the index of B4 derived from the 2STAR does not exhibit this outlier and shows similar directions but much higher price level compared with the other building specific indexes.

Table 5.2 Transaction information of Building 4 in Yong An Park

No.	Price (S$)	Floorarea	Level	Age	Contract Date
1	2100000	717	1	5.52	8-Jul-92
2	2480000	720	1	6.59	31-Jul-93
3	2970000	717	1	7.19	9-Mar-94
4	3150000	717	1	7.19	10-Mar-94
5	**1000000**	**717**	**1**	**11.62**	**11-Aug-98**
6	3100000	720	1	11.73	19-Sep-98
7	3400000	774	1	12.44	7-Jun-99

The reason for the second difference lies in the different processes to predict the value of the standard unit. In the geostatistical model, as specified by Equation (3.6), the price of the standard unit in a particular building at a particular quarter is determined by the standard $x_0\tilde{\beta}$ (x_0 is a vector consist of standard structural variables and quarterly time dummy variables) plus the predicted error $\hat{u}_0$ which is a weighted average of the hedonic residuals for nearby properties. The weights are determined by the correlation matrix and the correlation vector which is obtained by the correlogram. From the negative correlogram in Equation (3.4), clearly the transactions of the nearest distance will give the largest weights to the correlation function and the nearest distance between transactions in a multiunit market is zero. In other wordss, the value of the standard unit is primarily conditioned on the prices of transactions in the same building and secondly on the transactions in the nearest building and so on. As a result, a building specific index will exhibit abnormity when there is an outlier in that building.

Figure 5.2(a) Building specific indexes derived by GEO in Yong An Park

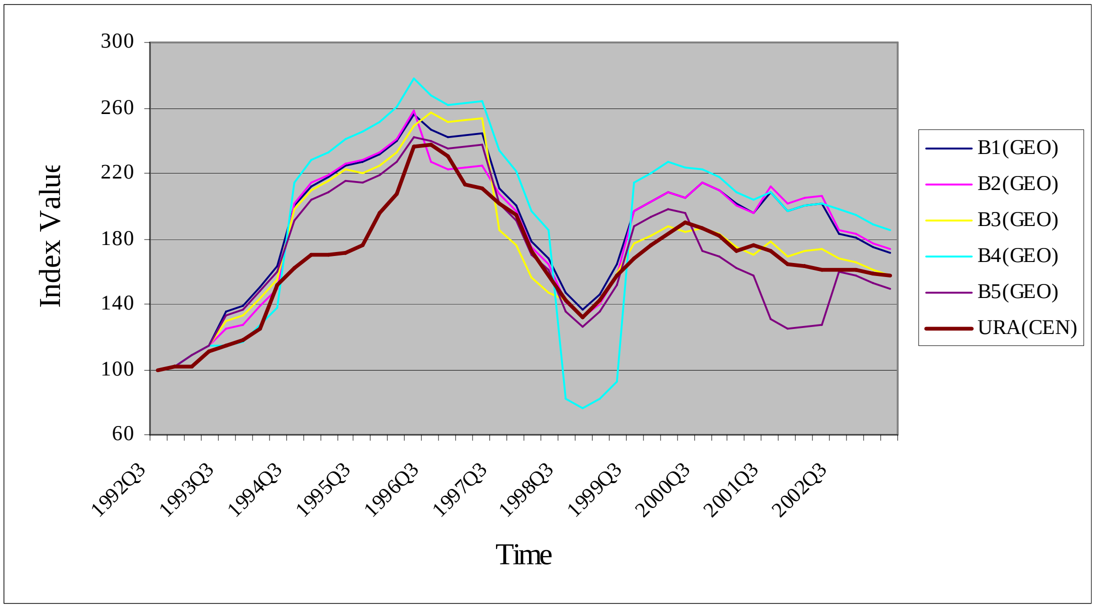
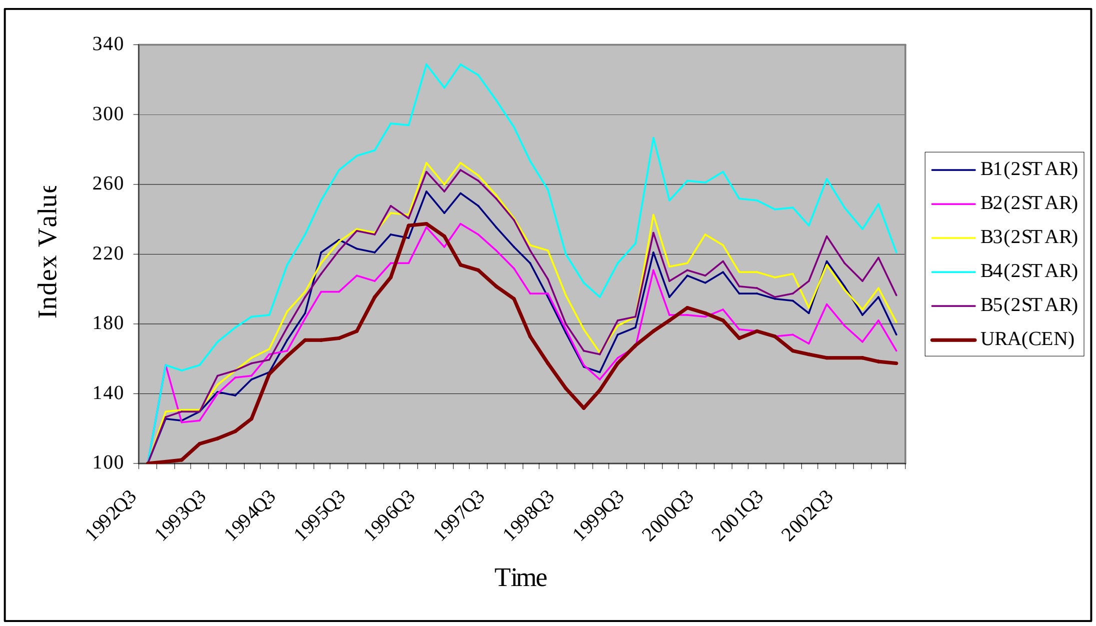


Figure 5.2(b) Building specific indexes derived by 2STAR in Yong An Park



In the 2STAR model, as specified by Equation 2.8, the predicted value of the standard unit in a particular building at a particular quarter is conditioned on the prior transactions in the same building, in the same neighborhood and across the whole area. So there are two possible reasons why the building specific index derived from the 2STAR can not exhibit abnormity when there is an outlier. Firstly, since the index value at a particular quarter only reflect the price trend of the transactions which were done in 3 or 4 days of that quarter, it will not exhibit abnormity if the outliers occurs outside this period. Secondly, even if the outlier happens during the period, its effect on the value of standard unit will be lowered by the prior transactions in the same building. For example, in B4, the effect of the outlier on the predicted value will be lowered by the four previous transactions of it.

The Claymore

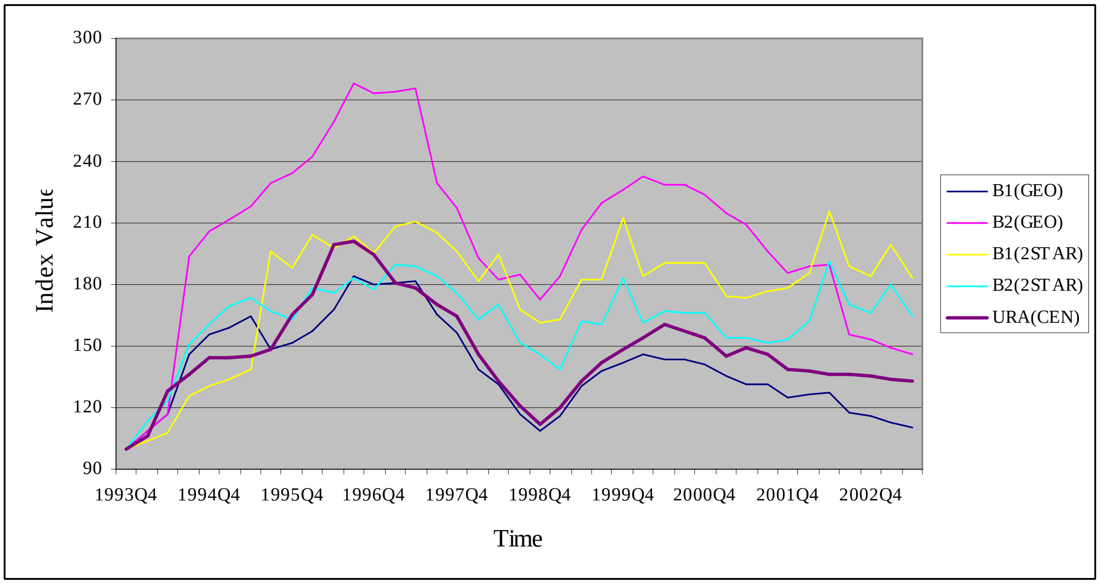
Table 5.3 Average house prices in the Claymore

Quarter	Average Price(S$)		Quarter	Average Price(S$)	
	B1	B2		B1	B2
Q4 1993	8407		Q4 1998	9679	
Q1 1994	9687		Q1 1999	11019	12651
Q2 1994	12262		Q2 1999	12266	14482
Q3 1994	12721		Q3 1999	11703	15663
Q4 1994	15148		Q4 1999	15691	16107
Q1 1995	12851		Q1 2000	15870	16770
Q2 1995	16238	14171	Q2 2000		
Q3 1995	16064	17511	Q3 2000	14366	16559
Q4 1995		19054	Q4 2000	12952	14900
Q1 1996	14257	17190	Q1 2001	13419	15863
Q2 1996	16270	17785	Q2 2001	12035	
Q3 1996	16867	17328	Q3 2001		

Q4 1996	16145		Q4 2001	11044	
Q1 1997	17235	19474	Q1 2002		17505
Q2 1997		17994	Q2 2002	11120	12851
Q3 1997	17068		Q3 2002	10611	
Q4 1997	18810	17269	Q4 2002	11369	
Q1 1998	10772		Q1 2003		
Q2 1998		14317	Q2 2003	10442	
Q3 1998				9679	

The Claymore is one of the most luxury condominiums located in the city central, with average price per square meter of which is 14,694 Singapore dollars. Table 5.2 lists quarterly average transaction prices by per square meter of each building in the Claymore. From this table, it is easy to note that the average house prices of Building 2(B2) are generally higher than Building 1(B1), indicating B2 command a premium over B1 because of its building-related characteristics such as view, orientation and geomantic omen. This building-related neighborhood effect is also reflected by the specific indexes derived from the geostatistical model. As shown in Figure 5.3, the price index of B2 indicates much higher price level than the index of B1. The price indexes from the 2STAR, on the other hand, seems not be able to capture the significant difference of price level between the two buildings.

Figure 5.3 Building specific indexes in the Claymore[22]



5.4 Discussion on Aggregate Index

Figure 5.4 and Figure 5.5 show the indexes for the whole area and four planning regions derived from the geostatistical model and the 2STAR model respectively, and Figure 5.6 shows the URA released indexes.

In general, all three figures show that the price index for the whole market seems to be an average of the sub-indexes of four planning areas. Compared with the sub-indexes which are derived from the spatial models, the URA released sub-indexes reflect more difference in the directions of price movement across four planning areas. For example, the index of the central region released by the URA increases and decreases more sharply than the other three sub-indexes do between the third quarters of 1993 and the fourth quarter of 1998.

[22] As the first transaction in Claymore occurred in the fourth quarter of 1993 during this study period, we take the fourth quarter as a basement.

Figure 5.4 Aggregate indexes derived by GEO

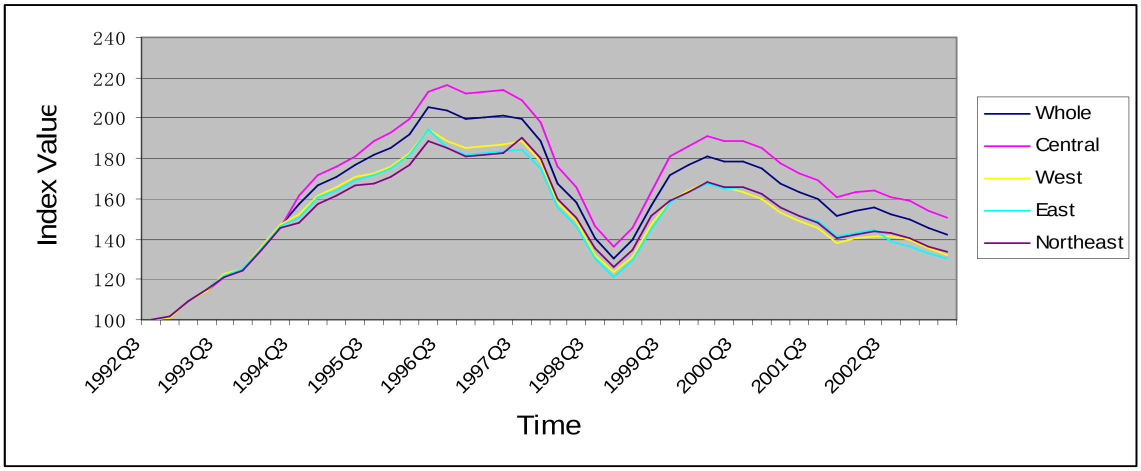


Figure 5.5 Aggregate indexes derived by 2STAR

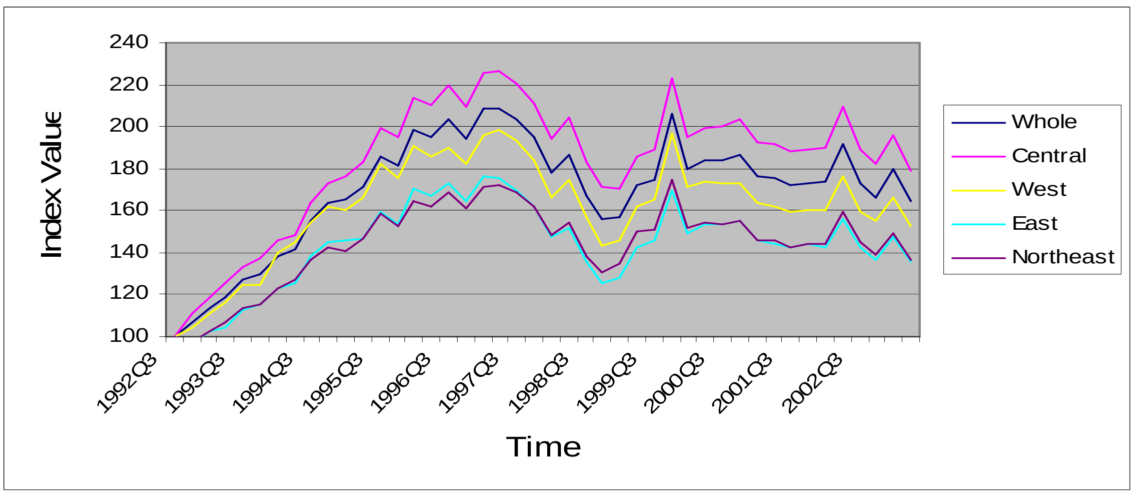
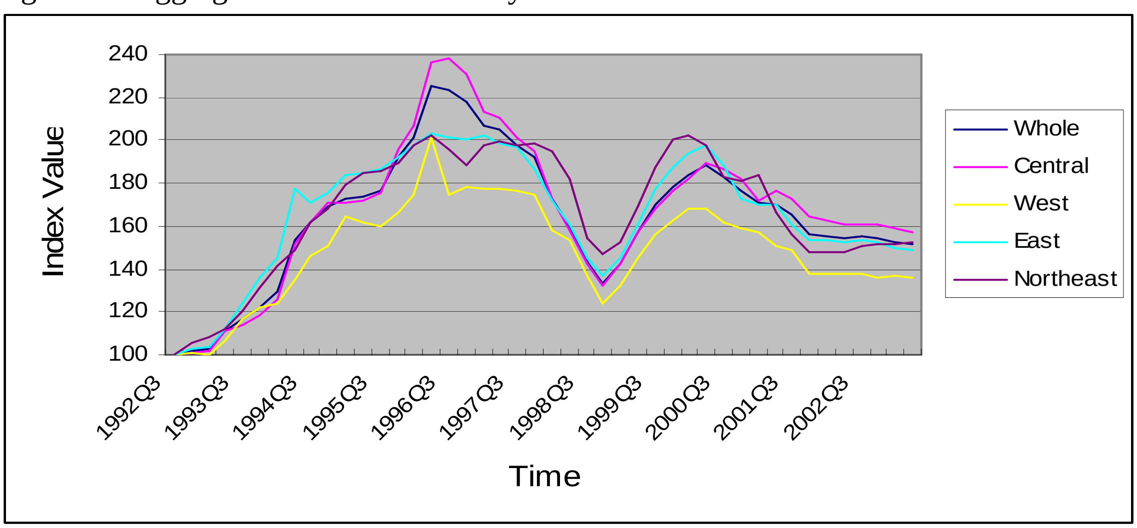


Figure 5.6 Aggregate indexes released by URA



Moreover, the indexes which are derived from the 2STAR model show greater and similar volatilities than the other two kinds of indexes during our study period. As discussed earlier, the predicted values of the standard unit of all buildings in a particular quarter are affected by the same prior transactions happened in a short period and across the whole area. This temporal effect certainly remains after the spatial weighted average. Therefore, the aggregate indexes exhibit much similar volatilities across the submarkets.

Similar conclusions can be reached by examining the price indexes of the freehold and leasehold resale condominiums derived by two spatial models, as shown at Figure 5.7, Figure 5.8 and Figure 5.9[23].

It is reasonable that the price index of the whole market is an average of the price indexes of submarkets or of different types of properties in the market. However, it may raise some doubts for discussion that the indexes derived from the two spatial models respectively indicate similar directions of price movement across submarkets or across different types. One possible reason for that is the similar average process generating the aggregate index. That is, the weighted average method with weights of the number of transactions in each building. This approach of averaging may demolish the distinct characteristics related to a particular submarket or a particular type of property in the whole market. Therefore, the aggregate indexes thus generated may exhibit similar directions of price movement across submarkets or across different types.

[23] As the URA leased price index for both freehold and leasehold condominiums starts from the fourth quarter of 1998 and is not appropriate for comparison, we show them only for reference.

Figure 5.7 Index of freehold and leasehold resale condos derived by GEO

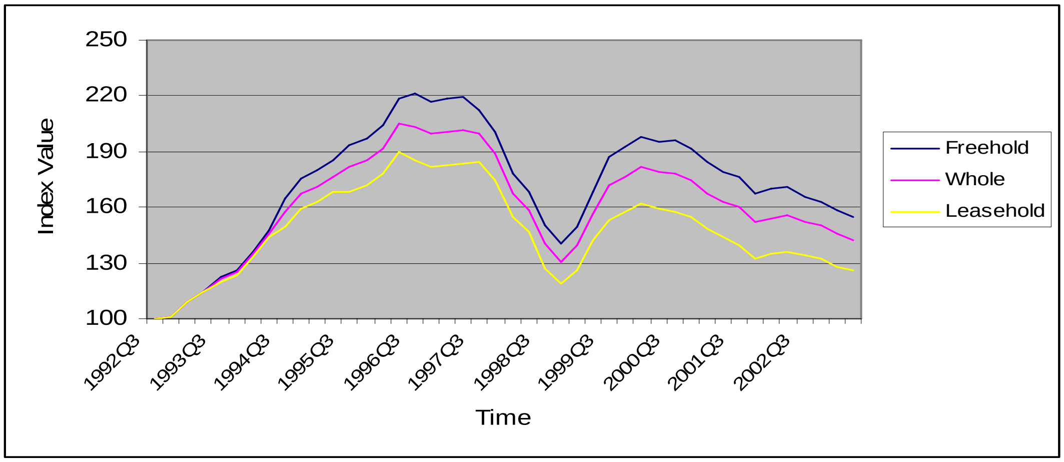


Figure 5.8 Index of freehold and leasehold resale condos derived by 2STAR

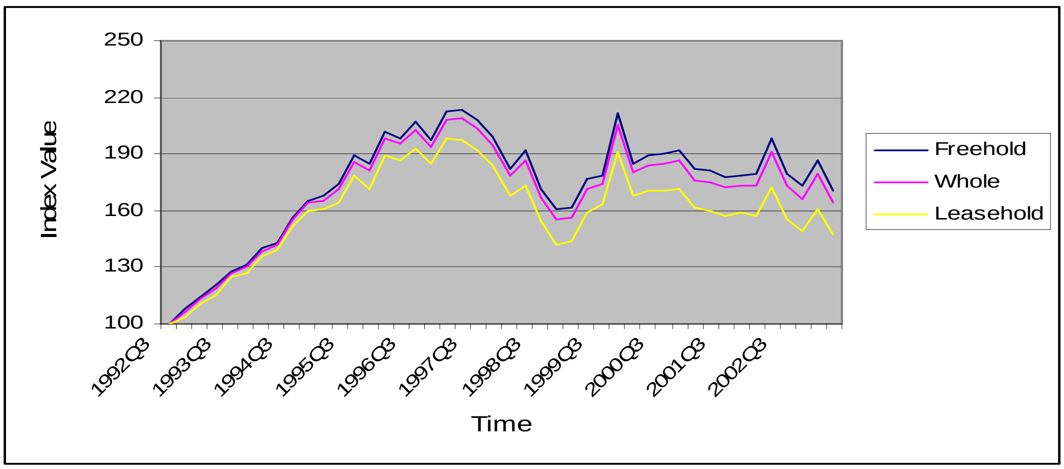


Figure 5.9 Price Index of freehold and leasehold condos released by URA

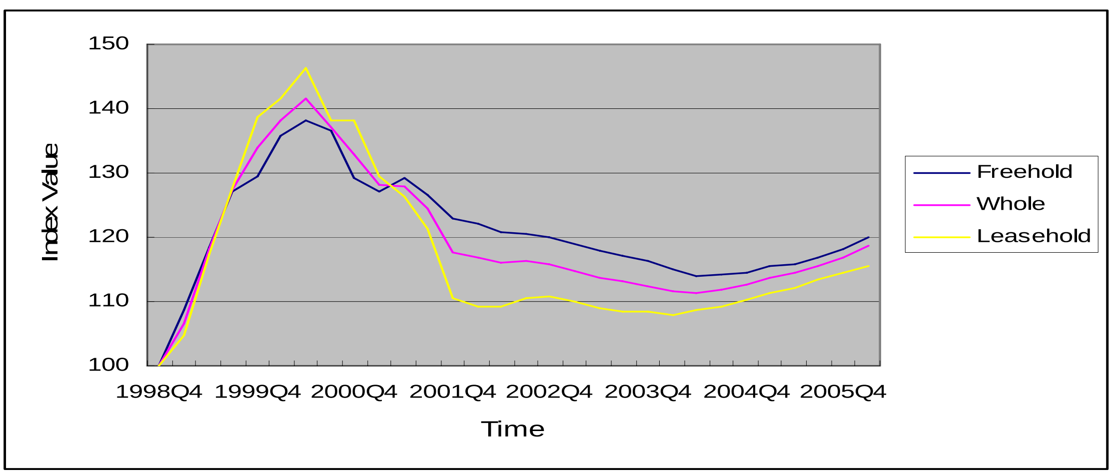
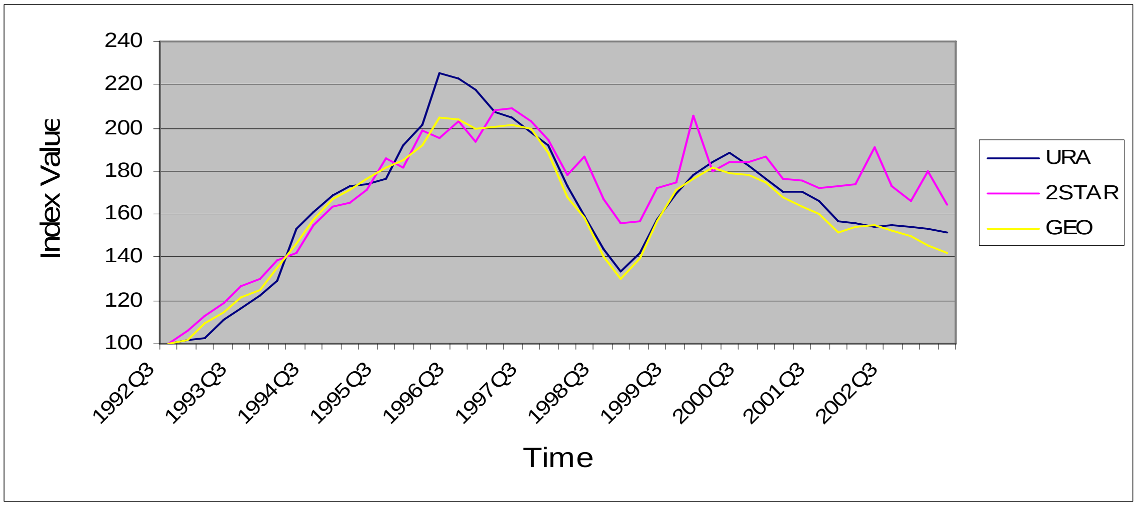


Figure 5.10 Price index in whole area



In order to further compare between aggregate indexes derived from the spatial models and the URA indexes, we examine the indexes of the whole area, as shown in Figure 5.10. The indexes for four planning areas are shown at Figure 5.1-5.4 in Appendix B respectively.

As shown in Figure 5.10, all the indexes capture the same broad price movement which has been explicitly discussed in Chapter 4. The price index derived from the geostatistical model moves more closely to the URA index than to the index derived from the 2STAR. Because the URA index is commonly regarded as reasonable, we conclude that the aggregate index derived from the geostatistical model can be regarded as reasonable. However, we will not recommend the geostatistical model with this spatial weighted average to construct an aggregate index. Because the aggregate index derived from such a complicated model does not present a more accurate picture of the changes in transaction prices than the index based on the simple Laspeyres technique, at least in our study. It is because the spatial weighted average technique which is used to generate the aggregate index may demolish the distinct characteristics related to a particular location or a particular submarket.

Although the 2STAR derived index exhibits greater volatilities than other two indexes, the magnitude of it varies less than others from the second quarter of 1998. One possible reason for this difference is that the price level of transactions happened in a short period can not fully represent the price level of transactions of the whole quarter.

5.5 Summary

Discussions and comparisons of both disaggregate and aggregate price indexes have been made in this chapter. Several interesting conclusions can be reached from our results.

1. Housing price dynamics varies significantly cross buildings, so that the aggregate index can not be used as a proxy to reflect individual building’s price dynamics.
2. A building specific index which is derived from the geostatistical model can provide a more accurate analysis of the price dynamics of a particular building than a building index which is derived from the 2STAR model. However, the specific index will be biased if the subject building has few transactions.
3. Although the aggregate index derived from the geostatistical model can reasonably reflect the price change for the whole market, it is not recommended to use this model with the spatial weighted average technique to construct an aggregate index, for it does not provide obvious improvement for reflecting the price dynamics of a general market compared with the index derived from a simple approach.

Chapter 6 Conclusion, Implications and Limitations

6.1 Conclusion

Based on an assumption that the building-related neighbor effects can be regarded as a nugget effect in a multiunit housing market, we employ the geostatistical model to solve the spatial autocorrelation problem in the hedonic price model and to obtain the accurate predicted house prices for the Singapore condominium resale market. The study period spans from July of 1993 to June of 2003. The comparisons of model performance in house price prediction are carried out among three geostatistical models, and between the geostatistical model and the 2STAR model. The empirical results show that the geostatistical models can effectively identify the spatial autocorrelation structure in the multiunit housing market. Among the three geostatistical models, the negative exponential correlogram has the best performance in house price prediction. Compared with the 2STAR model, the negative correlogram can be more efficient to correct the inefficient OLS estimation resulting from spatial autocorrelation. Regarding the house price prediction, the negative correlogram has a better performance than the 2STAR model, which is consistent with the Monte Carlo experiment results (Dubin, 2003).

We also derive the building specific indexes and the aggregate indexes both from the geostatistical model and the 2STAR model for the Singapore condominium resale market. Furthermore, we evaluate the model performance in constructing the specific indexes between these two models. We find that housing price dynamics varies significantly across building in the same project, so that the aggregate indexes can not be used as proxy to reflect individual building's price dynamics. Moreover, the building specific index derived from the geostatistical model can provide a more accurate analysis of the price dynamics of a particular building than the index derived from 2STAR model does. However, the building specific index will be biased if the subject building has few transactions. Finally, we compare the aggregate indexes derived by the geostatistical model with the indexes by the 2STAR model, and with the URA official indexes. We conclude that although the aggregate index derived from the geostatistical model can reasonably reflect the price change for the whole market, it is not recommended to use this model with the spatial weighted average technique to construct an aggregate index, for it can not provide obvious improvements in reflecting the price dynamics of the whole market compared with the index derived from a simple approach.

6.2 Implications

This study has three implications:

Firstly, it will serve as a contribution to both international and local real estate literature. From the modeling framework perspective, we are the first to employ the geostatistical model to capture the complicated spatial autocorrelation in a multiunit housing market. We argue that the building effect (Sun *et al.*, 2005) is a neighborhood effect at the building level, which can be regarded as a nugget effect from the point view of geostatistics. This argument is convincing, because the geostatistical model has a better performance both in estimation and in house price prediction than the 2STAR model does.

Secondly, it illustrates that the goestatistical model can be applied to predict the house prices in a multiunit housing market like Singapore and Hong Kong. Once the model is estimated, the price of a house can be produced by adding two portions. One is due to the structure, which can be calculated by multiplying the vector of structural attributes by the estimated coefficients. The other is due to neighborhood effect, which can be calculated by a weighted average of the estimation residuals. Moreover, the portion of the price due to the neighborhood effect gives an assessment of the neighborhood effects and eliminates the need to find comparable property.

Finally, since the price index at building level is different across buildings and significantly different from the aggregate index in a general market, it is necessary to derive building specific indexes to help investors make better investment decisions.

6.3 Limitations

In our study, there might be three limitations.

Firstly, we assume the error term of the hedonic model is second-order stationary which is usually examined empirically in a single family housing market. If the error term is not stationary, it will make the error term stationary by adding trends into the model. However, in a multiunit housing market, location of a transaction is determined by more than three dimensions. For example, there probably exist at least two transactions at the same level and with same coordinates. Therefore, in a multiunit housing market, it is difficult to examine the stationarity of the error term; and it is more difficult to make the error term stationary by adding trends into the model.

Secondly, although the geostatistical model can effectively capture the spatial autocorrelation structure in the multiunit housing market, it lacks ability to capture temporal autocorrelation structure in house prices. Further research is expected to explore the possibility of applying the geostatistical space-time model to model the spatial and temporal structure for a multiunit housing market.

The last limitation of this study is that although the geostatistical model can provide an improvement to solve the spatial autocorrelation problem in the multiunit housing market, it is based on the empirical study in the Singapore condominium resale market. Further research on this issue is expected from a context of Monto Carlo simulation.

Bibliography

Anselin,L. (1988) Spatial Econometrix: Methods and Models, Dordrecht: Kluwer.

Basu, Sabyasachi and Thibodeau, Thomas G. (1998) Analysis of Spatial Autocorrelation in House Prices, *Journal of Real Estate Finance and Economics*, 17(1): 61-85

Bjorklund, K, Soderberg, B and M, Wilhellmsson (2002) An Investigation of Property Price Studies, ARES Monographs, *Real Estate Valuation Theory* Ed. K Wang & M L Wolverton, Kluwer, Boston.

Blackley, D.M., Follain, J.R. and Lee, H. (1986) An Evaluation of Hedonic Price Indexes for Thirty-four Large SMSAs, AREUEA Journal, 14(2):179-205

Can, A. (1990) The Measurement of Neighborhood Dynamics in Urban Housing Prices, *Economic Geography*, 66(3), 254-272

Can, A. and Megbolugbe,I. (1997) Spatial Dependence and House Price Index Construction, *Journal of Real Estate Finance and Economics*, 14(1/2):203-222

Can, A. (1992) Specification and Estimation of Hedonic Housing Price Models, *Regional Science and Urban Economics*, 22:453-477

Can, A and Megbolugbe, I. (1997) Spatial Dependence and House Price Index

Construction, *Journal of Real Estate Finance and Economics*, 14:203-222

Case, B, Clapp,J, Dubin R and Rodriguez, M (2004) Modeling Spatial and Temporal House Price Patterns: A Comparison of Four Models, *Journal of Real Estate Finance and Economics,* 29(2):167-191

Clayton, J.D., Geltner, D.M. and Hamilton, S.W (2001) Smoothing in Commercial Property Valuations: Evidence from Individual Appraisals. *Real Estate Economics* 29(3): 337-360

Cressie, N.A. (1993) Statistics for Spatial Data, New York: Wiley

De Graaff, T., Florax, R. J. G. M., Nijkamp, P and A. Reggiani (2001) A General Misspecification Test for Spatial Regression Model: Dependence, Heterogeneity, and Nonlinearity, *Journal of Regional Science* 41: 255-276

Dubin, Robin A. (1992) Spatial Autocorrelation and Neighborhood Quality, Regional *Science and Urban Economics*, 22:433-452

Dubin, Robin A. (1998) Spatial Autocorrelation: A Primer, *Journal of Housing Economics,* 7: 304-327

Dubin, Robin A. (1998) Predicting House Prices Using Multiple Listing Data, *Journal of Real Estate Finance and Economics,* 17: 35-59

Dubin, Robin A. (2003) Robustness of Spatial Autocorrelation Specifications: Some Monte Carlo Evidence, *Journal of Regional Science*, 43: 221-248

Dubin, Robin A, R. K. Pace and T. G. Thibodeau. (1999) Spatial Autoregression Techniques for Real Estate Data, *Journal of Real Estate Literature* 7:70-95

Englund, Evan J. (2002) Spatial Autocorrelation: Implications for Sampling and Estimation, U.S. Environmental Protection Agency, P.O. Box 93478, Las Vegas, NV 89193-3478

Fleming, M.C. and Nellis, J.G. (1984) The Halifax House Price Index technical Details, Halifax Building Society, Halifax

Fletcher, M., P. Gallimore and J. Mangan (2000) Heteroscedasticity in hedonic house price models, *Journal of Property Research*, 17(2): 93-108

Forrest, D. (1991) An Analysis of House Price Differentials between English Regions, *Regional Studies*, 25, 231-8

Gelfand, A. E, M.D. Ecker, J.R. Knight and C.E. Sirmans (2004) The Dynamics of Location in Home Price, *Journal of Real Estate Finance and Economics,* 29(2):149-166

Gillen, K, T. Thibodeau and S. Wachiter (2004) Anisotropic Autocorrelation in House Prices, *Journal of Real Estate Finance and Economics* 23(1) : 5-30

Gneiting, T. (2002) Nonseparable, Stationary Covariance Functions for Space-Time Data, *Journal of the American Statistical Association,* 97(458):590-600

Gneiting T., Genton G. M. and Guttorp P. (2003) Geostatistical Space-Time Models, Stationarity, Separability and Full symmetry, Technical Report no.475

Goodman, A.C. and Thibodeau, T.G. (1995) Age-related Heteroskedasticity in Hedonic House Price Equations, *Journal of Housing Research* 6, 25-42

Goodman, A.C. and Thibodeau, T.G. (1997) Dwelling-age-related Heteroskedasticity in Hedonic House Price Equations: an Extension, *Journal of Housing Research* 8, 299-317

Goodman, A.C. and Thibodeau, T.G. (1997) Dwelling Age Heteroskedasticity in Repeat Sales House Price Equations. Real Estate Economics 26, 151-71

Gujarati, D.N. (2003) Basic Econometrics, *Mc Graw Hill Press,* 335-336

Haurin, D. R. and D. Brasington (1996) School Quality and Real House Prices: Inter- and Intrametropolitan Effects, *Journal of Housing Economics* 5: 351–368

Kwame, A.D. and Y.M. Tan (2003) Effect of Good Quality School on the Price of Residential Properties in Singapore. *A Research Bulletin of the Centre for Real Estate Studies (CRES), Department of Real Estate*, *NUS,* 3(1):2

Kuo, C.L.(1997) A Bayesian Approach to the Construction and Comparison of Alternative House Price Indices, *Journal of Real Estate Finance and Economics*, 14(1/2): 113-132

Kyriakidis P. C. and Journel A.G. (1999) Geostatistical Space-time Model: A Review, *Mathematical Geology,* 31(6):651-684

LeSage,J.P. (1998) Spatial Econometrics. Unpublished Manuscript. (*www.Spatial-econometrics.com*)

Lesage, J.P. and R.K. Pace (2004) Models for Spatial Dependent Missing Data, *Journal of Real Estate Finance and Economics,* 29(2): 233-254

Mark, J.H. and Goldberg, M.A. (1984) Alternative Housing Price Indices: Evaluation, *AREUEA Journal*, 12(Spring):30-49

Matheron, G. (1963) Principles of Geostatistics, *Economic Geology* 58: 1246-1266

Militino,A. F, M.D. Ugarte and L.Garcia-Reinaldos (2004) Alternative Models for

Describing Spatial Dependence among Dwelling Selling Price, *Journal of Real Estate Finance and Economics*, 29(2): 193-209

Oden, L.N. and Sokal, S.R (1986) Directional Autocorrelation: An Extension of Spatial Correlograms to Two Dimensions, *Systematical Zoology* 35(4):608-617

Pace, P.K. and Gilley, O.W. (1997) Using the Spatial Configuration of the Data to Improve Estimation, *Journal of Real Estate Finance and Economics*, 14, 333-340.

Pace, P.K. and Barry, R.(1997) Sparse Spatial Autoregressions, *Statistical and probability Letters*, 33, 291-297.

Pace, R.K., R. Barry, J.M. Clapp and M. Rodriquez (1998) Spatiotemporal Autoregressive Models of Neighborhood Effects, *Journal of Real Estate Finance and Economics*, 17(1):14-33

Pace, R.K., Barry, R., Gilley, O.W. and Sirmans, C.F.(2000) A Method for Spatial-temporal Forecasting with an Application to Real Estate Prices, *International Journal of Forecasting*, 16: 229-246

Pace, P.K and Barry, R. (2004) Spatial Statistics and Real Estate, Louisiana State University

Pace, R.K. and J.P. Lesage (2004) Spatial Statistics and Real Estate, *Journal of Real Estate Finance and Economics*, 29(2): 147-148

Phang, S. and W. Wong (1997) Government Policies and Private Housing Prices in Singapore, *Urban Studies*, 34(11), 1819 – 1830

Stevenson,S. (2004) New Empirical Evidence on Heteroscedasticity in Hedonic Housing Models, *Journal of Housing Economics*, 13(2): 136-153

Stoer, J and Bulirsch, R. (1980) Introduction to Numerical Analysis. New York: Springer-Verlag

Tu, Y, Yu,S.M and Sun.H (2004) Transaction-Based Office Price Indexes: A Spatiotemporal Modeling Approach. *Real Estate Economics* 32(2):297-328

Sun, H, Tu, Y and Yu, S.M (2005) A Spatio-temporal Autoregressive Model for Multiunit Residential Market Analysis, *Journal of Real Estate Finance and Economics* 31(2): 155-187

Tse, R.Y.C and J.R. Webb (1999) Dynamic Equilibrium of the Housing Market, *Urban Studies* 36(16): 2361 – 2373

Appendix A

Matlab Functions for This Study

Function 1: negkrig.m

```
function results=negkrig(y,x,xco,yco,yp,xp,xpco,ypco);
% PURPOSE: to estimate hedonic and spatial pameters and then predict house
% price using kriging with negative correlation function
%-------------------------------------------------
% USAGE: results = negkrig(y,x,xco,yco,yp,xp,xpco,ypco)
% where: y = dependent variable vector    in estimation sample    (nobs x 1)
%        x = independent variables matrix in estimation sample (nobs x nvar)
%        xco = x-coordinate vector in estimation sample (nobs x 1)
%        yco = y-coordinate vector in estimation sample (nobs x 1)
%        yp = dependent variable vector    in house price prediction sample    (mobs x 1)
%        xp = independent variables matrix in house price prediction sample (mobs x nvar)
%        xpco = x-coordinate vector in house price prediction sample (mobs x 1)
%        ypco = y-coordinate vector in house price prediction sample (mobs x 1)
%-------------------------------------------------
% RETURNS: a structure
%        results.meth    = 'negkrig'
%        results.b1=b1
%        results.b2=b2
%        results.vof    =value of maximum likelihood function
%        results.beta    = bhat         (nvar x 1)
%        results.tstat = t-stats    (nvar x 1)
%        results.sd        = standard error (nvar x 1)
%        results.yphat = predicted values (mobs x 1)
%        results.residp = house price prediction residuals(error) (mobs x 1)
%        results.mse    = mean square of house price prediction error      scalar
%        results.tu      = Theil's U    scalar                              scalar
%         results.mean = mean of absolute value of house price prediction error scalar
%        results.median = median of absolute value of house price prediction error scalar
%-------------------------------------------------------------
  b1=[0.1:0.1:0.9]; %b1 varies between 0.1 and 1, in increments of 0.1
  b2=[0.5:0.5:5];      %    b2 varies between 0.5 and 5, in increments of 0.5
  [nobs nvar] = size(x);
  [mobs mvar]=size(xp);
results.meth = 'negkrig';
```

```
dist=zeros(nobs,nobs);

for i=1:nobs          % calculating the distance between two transactions
    for j=1:nob
        dist(i,j)   =   sqrt((xco(i,1)-xco(j,1)).*(xco(i,1)-xco(j,1))   +   (yco(i,1)-
         yco(j,1)).*(yco(i,1)-yco(j,1)))/1000;
    end
end
% the first search %
timefirst=fix(clock)
first=negsearch(y,x,dist,b1,b2);
aa=first.b1hat;
b1f=[aa-0.05 aa aa+0.05];
bb=first.b2hat;
b2f=[bb-0.25 bb bb+0.25];
clear aa bb
% the second search
timesecond=fix(clock)
second=negsearch(y,x,dist,b1f,b2f);
aa=second.b1hat;
b1s=[aa-0.025 aa aa+0.025];
bb=second.b2hat;
b2s=[bb-0.125 bb bb+0.125];
clear aa bb b1 b2
% the last search
timelast=fix(clock)
last=negsearch(y,x,dist,b1s,b2s);
b1=last.b1hat;
b2=last.b2hat;
results.b1=b1;
results.b2=b2;
results.vof=last.maxvof;
% The following precedure aims to estimate    hedonic regression
timeestimate=fix(clock)
  K=zeros(nobs,nobs);
      for i=1:nobs
          for j=1:nobs
              if dist(i,j)==0
                  if i==j
                      K(i,j)=1;
                  else
                      K(i,j)=b1;
                  end
              else
                  K(i,j)=b1*exp(-(dist(i,j)/b2));
              end
```

```
        end
    end

de=det(K);
inve=inv(K);
temp=inv(x'*inve*x);
results.beta=temp*x'*inve*y;
yhat = x*results.beta;
resid = y - yhat;
sigu = resid'*inve*resid;
sige = sigu/(nobs-nvar);
tmp = sige*diag(temp);
results.sd=sqrt(tmp);
results.tstat = results.beta./(sqrt(tmp));
%-----------------------------------------------------------
% The following procedure performs kriging
timekrig=fix(clock)
distp=zeros(mobs,nobs);
for i=1:mobs
    for j=1:nobs

distp(i,j)=sqrt((xpco(i)-xco(j)).*(xpco(i)-xco(j))+(ypco(i)-yco(j)).*(ypco(i)-yco(j)))/1000;
        if distp(i,j)==0
          co(i,j)=b1;
        else

            co(i,j)=b1*exp(-(distp(i,j)/b2));

        end
    end
end
yphat=xp*results.beta+co*inve*resid;
results.yphat=yphat;
residp=yp-yphat;
results.residp=residp;
results.mse=residp'*residp/mobs; % Mean Square Error
results.tu=(results.mse)/(sqrt(yp'*yp/mobs)+sqrt( yphat'* yphat/mobs)); % Theil's U
results.mean=mean(abs(residp)); % Mean of absolute value of house price prediction
error
results.median=median(abs(residp));% Median of absolute value of house price
prediction error
```

Function 2: negkrig.m

```
function results=negsearch(y,x,dist,b1,b2)
% PURPOSE: grid search procedure to find the value of spatial parameter(b1,b2) of
% negative correlation function that maximize the likelihood function
%-------------------------------------------------
% USAGE: results = negsearch(y,x,dist,b1,b2)
% where: y = dependent variable vector       (nobs x 1)
%              x = independent variables matrix (nobs x nvar)
%              dist=distance matrix(nobs x nobs, symmetric)
%              b1= b1 grid vector
%              b2=b2    grid vector
%-------------------------------------------------
% RETURNS: a structure
%              results.meth    = 'negsearch'
%              results.maxvof = value of maximum likelihood function
%              results.b1hat = b1max
%              results.b2hat    = b2max
%-------------------------------------------------

  [nobs nvar] = size(x);
results.meth = 'negsearch';
num1=length(b1);
num2=length(b2);
L=zeros(num1,num2);
for k=1:num1
      for l=1:num2
         K=zeros(nobs,nobs);
        for i=1:nobs
            for j=1:nob
                    if dist(i,j)==0
                          if i==j
                                K(i,j)=1;
                          else
                                K(i,j)=b1(k);
                          end
                    else
                          K(i,j)=b1(k)*exp(-(dist(i,j)/b2(l))); % negative correlation function

                    end
                end

        end

de=det(K);
inve=inv(K);
temp1=inv(x'*inve*x);
```

```
beta=temp1*x'*inve*y;
temp2=log((y-x*beta)'*inve*(y-x*beta));
if de==0
        q=k*0.5;
        dd(k,l)=det(K*q);
        k0=k;
        l0=l;
        L(k,l)=-nobs*temp2-log(dd(k,l))+nobs*log(q);
          if dd(k,l)>=1.7976e+308
                      m=1;
                      dd(k,l)=det(K*(q-m));
                      kmax=k;
                      lmax=l;
                      while dd(k,l)>=1.7976e+308
                          m=m+1;
                          dd(k,l)=det(K*(q-m));
                      end
                      L(k,l)=-nobs*temp2-log(dd(k,l))+nobs*log(q-m);
                      clear m
          else
                  if    dd(k,l)<=2.2251e-308
                          n=1;
                          kmin=k;
                          lmin=l;
                          dd(k,l)=det(K*(q+n));
                          while dd(k,l)<=2.2251e-308
                          n=n+1;
                          dd(k,l)=det(K*(q+n));
                          end
                          L(k,l)=-nobs*temp2-log(dd(k,l))+nobs*log(q+n);
                          clear n
                  end
            end
else
    L(k,l)=-(nobs*temp2+log(de));
end
end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
MM=max(max(L));
[i,j]=find(L==MM);
results.L=L;
results.maxvof=MM;
  results.b1hat = b1(i);
  results.b2hat = b2(j);
```

Appendix B

Tables and Figures for Chapter 4 and 5

Table 4.1 Empirical results from July of 1992 to June of 1993

Period		July 1992 – June 1993			
In sample size		1273			
Out sample size		500			
	VARIABLES	OLS	GSS	NEG	SPH
ESTIMATION	CONSTANT	12.5540** (0.0242)	12.6180** (0.0014)	12.6230** (0.0019)	12.627** (0.002)
	LEVEL	0.0052** (0.0009)	0.0044** (0.0000)	0.0045** (0.0000)	0.0045** (0.0000)
	AGE	-0.0188** (0.002)	-0.0155** (0.0000)	-0.0169** (0.0000)	-0.0170** (0.0000)
	FLRAREA	0.0039** (0.0001)	0.0031** (0.0000)	0.0031** (0.0000)	0.0031** (0.0000)
	FREEHOLD	0.1763** (0.0113)	0.2349** (0.0008)	0.1968** (0.0008)	0.1944** (0.0007)
	T2	0.0195 (0.0207)	0.0132 (0.0001)	0.0129 (0.0001)	0.0127 (0.0001)
	T3	0.1057** (0.0201)	0.0861** (0.0001)	0.0837** (0.0001)	0.0843** (0.0001)
	T4	0.1645** (0.0188)	0.1411** (0.0001)	0.1375** (0.0001)	0.1383** (0.0001)
	VOF	-2414.3	-1849.7	-1827	-1832.8
	b1		0.775	0.775	0.825
	b2		0.375	0.875	2.375
PREDIC-TION	MSE	0.0370	0.0145	0.0132	0.0134
	Theil's U	0.0014	0.0005	0.0005	0.0005
	Mean A	0.1508	0.0845	0.0804	0.0811
	Median A	0.1252	0.0576	0.0577	0.0578

Table 4.2 Empirical results from July of 1993 to June of 1994

Period		July 1993 – June 1994			
In sample size		2839			
Out sample size		500			
	VARIABLES	OLS	GSS	NEG	SPH
ESTIMATION	CONSTANT	12.735** (0.0147)	12.793** (0.0011)	12.8000** (0.0013)	12.808** (0.0014)
	LEVEL	0.0073** (0.0006)	0.0056** (0.0000)	0.0055** (0.0000)	0.0056** (0.0000)
	AGE	-0.0153** (0.0012)	-0.0102** (0.0000)	-0.0107** (0.0000)	-0.0104** (0.0000)
	FLRAREA	0.0037** (0.0000)	0.0028** (0.0000)	0.0028** (0.0000)	0.0028** (0.0000)
	FREEHOLD	0.2169** (0.0000)	0.2693** (0.0000)	0.2285** (0.0000)	0.2356** (0.0000)
	T2	0.0362** (0.0109)	0.0298** (0.0000)	0.0302** (0.0000)	0.0303** (0.0000)
	T3	0.1257** (0.0114)	0.1111** (0.0000)	0.1118** (0.0000)	0.1118** (0.0000)
	T4	0.1974** (0.0101)	0.1855** (0.0000)	0.1864** (0.0000)	0.1864** (0.0000)
	VOF	-6778	-5515.6	-5462	-5475
	b1		0.775	0.75	0.8
	b2		0.25	0.5	1.125
PREDIC-TION	MSE	0.0455	0.0155	0.0136	0.0137
	Theil's U	0.0017	0.0006	0.0005	0.0005
	Mean A	0.1565	0.0908	0.087	0.0872
	Median A	0.1227	0.0695	0.0664	0.0659

Table 4.3 Empirical results from July of 1994 to June of 1995

Period		July 1994 – June 1995			
In sample size		2286			
Out sample size		500			
	VARIABLES	OLS	GSS	NEG	SPH
ESTIMATION	CONSTANT	12.8150** (0.0212)	12.9360** (0.0015)	12.9510** (0.0019)	12.9480** (0.0031)
	LEVEL	0.0097** (0.0009)	0.0061** (0.0000)	0.0060** (0.0000)	0.0061** (0.0000)
	AGE	-0.0143** (0.0015)	-0.0065** (0.0000)	-0.0083** (0.0000)	-0.0085** (0.0000)
	FLRAREA	0.0044** (0.0001)	0.0031** (0.0000)	0.0030** (0.0000)	0.0031** (0.0000)
	FREEHOLD	0.3005** (0.013)	0.3276** (0.0012)	0.2689** (0.0011)	0.270** (0.0011)
	T2	0.0654** (0.0156)	0.0635** (0.0001)	0.0616** (0.0001)	0.0618** (0.0001)
	T3	0.0657** (0.0183)	0.0916** (0.0001)	0.0901** (0.0001)	0.0902** (0.0001)
	T4	0.1763** (0.0153)	0.1193** (0.0001)	0.1205** (0.0001)	0.1206** (0.0001)
	VOF	-5854.9	-4417.2	-4331	-4347.6
	b1		0.825	0.825	0.9
	b2		0.25	0.625	2.25
PREDIC-TION	MSE	0.0697	0.0164	0.0159	0.0158
	Theil's U	0.0025	0.0006	0.0006	0.0006
	Mean A	0.2048	0.0889	0.087	0.0872
	Median A	0.1654	0.0656	0.0649	0.0649

Table 4.4 Empirical results from July of 1996 to June of 1997

Period		July 1996 – June 1997			
In sample size		961			
Out sample size		500			
	VARIABLES	OLS	GSS	NEG	SPH
ESTIMATION	CONSTANT	13.0140** (0.0305)	13.2320** (0.0017)	13.2050** (0.0027)	13.2110** (0.0036)
	LEVEL	0.0091** (0.0016)	0.0056** (0.0000)	0.0053** (0.0000)	0.0053** (0.0000)
	AGE	-0.0171** (0.0019)	-0.0147** (0.0000)	-0.0128** (0.0000)	-0.0134** (0.0000)
	FLRAREA	0.0045** (0.0001)	0.0033** (0.0000)	0.0031** (0.0000)	0.0031** (0.0000)
	FREEHOLD	0.3748** (0.0192)	0.2974** (0.0007)	0.2959** (0.0009)	0.2914** (0.0008)
	T2	-0.0140 (0.027)	-0.0258** (0.0002)	-0.0278** (0.0002)	-0.0283** (0.0002)
	T3	0.0162 (0.0256)	-0.0243** (0.0002)	-0.0215 (0.0002)	-0.0221 (0.0002)
	T4	-0.0329 (0.0242)	-0.0266** (0.0002)	-0.0241** (0.0002)	-0.0252** (0.0002)
	VOF	-2051.4	-1571.2	-1528	-1533.1
	b1		0.75	0.8000	0.85
	b2		0.75	1.1250	3.5
PREDIC-TION	MSE	0.0762	0.0213	0.0182	0.0188
	Theil's U	0.0027	0.0008	0.0007	0.0007
	Mean A	0.2104	0.1050	0.0966	0.0984
	Median A	0.1721	0.0758	0.0704	0.0712

Table 4.5 Empirical results from July of 1997 to June of 1998

Period		July 1997 – June 1998			
In sample size		704			
Out sample size		500			
	VARIABLES	OLS	GSS	NEG	SPH
ESTIMATION	CONSTANT	12.9200** (0.0312)	13.1650** (0.0015)	13.1490** (0.0027)	13.1530** (0.0022)
	LEVEL	0.0078** (0.0016)	0.0038** (0.0000)	0.0044** (0.0000)	0.0043** (0.0000)
	AGE	-0.0150** (0.0018)	-0.0165** (0.0000)	-0.0152** (0.0000)	-0.0154** (0.0000)
	FLRAREA	0.0053** (0.0002)	0.0036** (0.0000)	0.0036** (0.0000)	0.0036** (0.0000)
	FREEHOLD	0.3199** (0.0203)	0.2833** (0.0005)	0.2790** (0.0007)	0.2795** (0.0006)
	T2	-0.0931** (0.0252)	-0.0575** (0.0003)	-0.0585** (0.0003)	-0.0587** (0.0003)
	T3	-0.1498** (0.0286)	-0.1877** (0.0004)	-0.1832** (0.0004)	-0.1872** (0.0004)
	T4	-0.2437** (0.0227)	-0.2399** (0.0003)	-0.2400** (0.0002)	-0.2405** (0.0002)
	VOF	-1296.9	-1092.3	-1079	-1079.7
	b1	0	0.55	0.6500	0.65
	b2	0	1.25	1.6250	3.875
PREDIC-TION	MSE	0.0690	0.0250	0.0237	0.0239
	Theil's U	0.0025	0.0009	0.0009	0.0009
	Mean A	0.1817	0.1107	0.1071	0.1072
	Median A	0.1380	0.0791	0.0784	0.0785

Table 4.6 Empirical results from July of 1998 to June of 1999

Period		July 1998 – June 1999			
In sample size		2953			
Out sample size		500			
	VARIABLES	OLS	GSS	NEG	SPH
ESTIMATION	CONSTANT	12.5850** (0.0198)	12.7710** (0.0008)	12.7560** (0.0011)	12.7570** (0.0015)
	LEVEL	0.0102** (0.0008)	0.0060** (0.0000)	0.0060** (0.0000)	0.0060** (0.0000)
	AGE	-0.0187** (0.0008)	-0.0138** (0.0000)	-0.0151** (0.0000)	-0.0157** (0.0000)
	FLRAREA	0.0050** (0.0001)	0.0038** (0.0000)	0.0038** (0.0000)	0.0038** (0.0000)
	FREEHOLD	0.3359** (0.0097)	0.3196** (0.0006)	0.2781** (0.0005)	0.2789** (0.0004)
	T2	-0.0150 (0.0189)	-0.0635** (0.0001)	-0.0682** (0.0001)	-0.0677** (0.0001)
	T3	0.0302 (0.0176)	-0.0031 (0.0001)	-0.0051 (0.0001)	-0.0048 (0.0001)
	T4	0.1520** (0.0163)	0.1134** (0.0001)	0.1120** (0.0001)	0.1121** (0.0001)
	VOF	-7563.9	-6169.8	-6056	-6073
	b1	0	0.775	0.7500	0.825
	b2	0	0.25	0.7500	2.25
PREDIC-TION	MSE	0.0576	0.0254	0.0234	0.0240
	Theil's U	0.0021	0.0009	0.0009	0.0009
	Mean A	0.1670	0.0956	0.0944	0.0952
	Median A	0.1206	0.0634	0.0631	0.0639

Table 4.7 Empirical results from July of 1999 to June of 2000

Period		July 1999 – June 2000			
In sample size		3234			
Out sample size		500			
	VARIABLES	OLS	GSS	NEG	SPH
ESTIMATION	CONSTANT	12.7940** (0.0148)	12.9070** (0.0007)	12.8970** (0.0011)	12.8930** (0.002)
	LEVEL	0.0093** (0.0008)	0.0057** (0.0000)	0.0056** (0.0000)	0.0056** (0.0000)
	AGE	-0.0158** (0.0008)	-0.0096** (0.0000)	-0.0122** (0.0000)	-0.0129** (0.0000)
	FLRAREA	0.0050** (0.0001)	0.0039** (0.0000)	0.0039** (0.0000)	0.0039** (0.0000)
	FREEHOLD	0.3563** (0.01)	0.3304** (0.0005)	0.2996** (0.0004)	0.3029** (0.0004)
	T2	0.0503** (0.0127)	0.0295** (0.0000)	0.0305** (0.0000)	0.0310** (0.0000)
	T3	0.1106** (0.0123)	0.0529** (0)	0.0540** (0)	0.0548** (0)
	T4	0.0475** (0.0132)	0.0434** (0.0000)	0.0462** (0.0000)	0.0467** (0.0000)
	VOF	-8732.7	-6793.6	-6674	-6694.1
	b1	0	0.8	0.8000	0.875
	b2	0	0.25	0.7500	2.625
PREDIC-TION	MSE	0.0737	0.0232	0.0223	0.0224
	Theil's U	0.0027	0.0008	0.0008	0.0008
	Mean A	0.1989	0.0969	0.0939	0.0945
	Median A	0.1503	0.0663	0.0639	0.0643

Table 4.8 Empirical results from July of 2000 to June of 2001

Period		July 2000 – June 2001			
In sample size		1273			
Out sample size		500			
	VARIABLES	OLS	GSS	NEG	SPH
ESTIMATION	CONSTANT	12.7600** (0.0215)	13.0310** (0.001)	12.9650** (0.0018)	12.9660** (0.0046)
	LEVEL	0.0050** (0.0012)	0.0050** (0.0000)	0.0048** (0.0000)	0.0048** (0.0000)
	AGE	-0.0109** (0.0011)	-0.0148** (0.0000)	-0.0156** (0.0000)	-0.0159** (0.0000)
	FLRAREA	0.0051** (0.0001)	0.0039** (0.0000)	0.0039** (0.0000)	0.0039** (0.0000)
	FREEHOLD	0.4288** (0.0144)	0.2500** (0.0006)	0.2681** (0.0006)	0.2585** (0.0005)
	T2	0.0530** (0.0174)	-0.0212** (0.0001)	-0.0174** (0.0001)	-0.0171 (0.0001)
	T3	-0.0951** (0.0195)	-0.0595** (0.0001)	-0.0561** (0.0001)	-0.0568** (0.0001)
	T4	-0.1094** (0.0187)	-0.0914** (0.0001)	-0.0880** (0.0001)	-0.0895** (0.0001)
	VOF	-3854.4	-3071.4	-2999	-3012.1
	b1	0	0.8	0.8250	0.9
	b2	0	0.375	0.8750	3.75
PREDIC-TION	MSE	0.0834	0.0213	0.0201	0.0203
	Theil's U	0.0030	0.0008	0.0007	0.0007
	Mean A	0.2204	0.0969	0.0930	0.0940
	Median A	0.1733	0.0721	0.0656	0.0685

Table 4.9 Empirical results from July of 2001 to June of 2002

Period		July 2001- June 2002			
In sample size		1256			
Out sample size		500			
	VARIABLES	OLS	GSS	NEG	SPH
ESTIMATION	CONSTANT	12.6450** (0.0244)	12.8630** (0.0009)	12.8450** (0.0015)	12.8510** (0.0025)
	LEVEL	0.0104** (0.0012)	0.0063** (0.0000)	0.0060** (0.0000)	0.0060** (0.0000)
	AGE	-0.0140** (0.0012)	-0.0142** (0.0000)	-0.0145** (0.0000)	-0.0149** (0.0000)
	FLRAREA	0.0048** (0.0001)	0.0038** (0.0000)	0.0037** (0.0000)	0.0038** (0.0000)
	FREEHOLD	0.4440** (0.0156)	0.3395** (0.0007)	0.3163** (0.0006)	0.3102** (0.0006)
	T2	-0.0356 (0.0319)	-0.0738** (0.0004)	-0.0727** (0.0003)	-0.0746** (0.0003)
	T3	-0.0164 (0.0198)	-0.0354** (0.0001)	-0.0385** (0.0001)	-0.0375** (0.0001)
	T4	-0.0028 (0.0197)	-0.0269** (0.0001)	-0.0303** (0.0001)	-0.0293** (0.0001)
	VOF	-2783	-2298.8	-2218	-2229.5
	b1		0.75	0.7750	0.85
	b2		0.375	0.8750	3
PREDIC- TION	MSE	0.0681	0.0234	0.0208	0.0208
	Theil's U	0.0025	0.0009	0.0008	0.0008
	Mean A	0.1938	0.1038	0.0976	0.0983
	Median A	0.1473	0.0733	0.0681	0.0691

Table 4.10 Empirical results from July of 2002 to June of 2003

Period		July 2002 – June 2003			
In sample size		1361			
Out sample size		500			
	VARIABLES	OLS	GSS	NEG	SPH
ESTIMATION	CONSTANT	12.5620** (0.0224)	12.7710** (0.0010)	12.7500** (0.0020)	12.7610** (0.0020)
	LEVEL	0.0113** (0.0011)	0.0042** (0.0000)	0.0044** (0.0000)	0.0045** (0.0000)
	AGE	-0.0108** (0.0011)	-0.0167** (0.0000)	-0.0156** (0.0000)	-0.0159** (0.0000)
	FLRAREA	0.0050** (0.0001)	0.0043** (0.0000)	0.0043** (0.0000)	0.0043** (0.0000)
	FREEHOLD	0.4237** (0.0141)	0.3254** (0.0003)	0.3076** (0.0004)	0.3080** (0.0004)
	T2	0.0078 (0.0177)	-0.0132 (0.0001)	-0.0185 (0.0001)	-0.0173 (0.0001)
	T3	-0.0652** (0.0199)	-0.0453** (0.0002)	-0.0474** (0.0002)	-0.0475** (0.0002)
	T4	-0.0584** (0.0193)	-0.0728** (0.0002)	-0.0771** (0.0001)	-0.0763** (0.0001)
	VOF	-3016.7	-2493.4	-2463	-2466.2
	b1		0.65	0.7	0.75
	b2		1	1.5	3.75
PREDIC-TION	MSE	0.0631	0.0234	0.0211	0.0214
	Theil's U	0.0023	0.0009	0.0008	0.0008
	Mean A	0.1910	0.1082	0.1011	0.1019
	Median A	0.1385	0.0804	0.0773	0.0769

Figure 4.1 Correlograms from July of 1992 to June of 1993

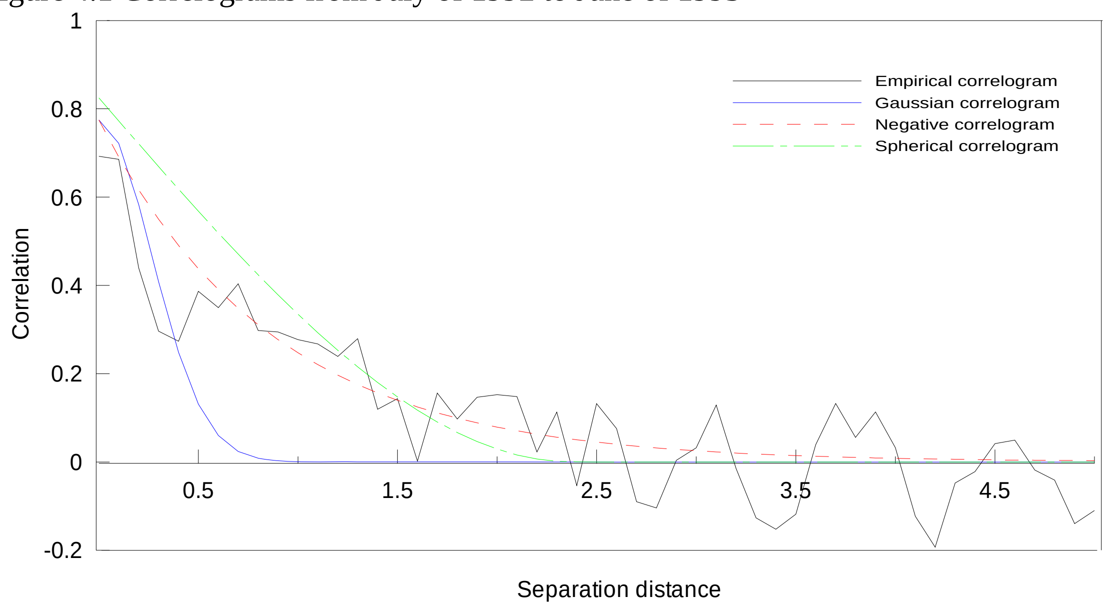


Figure 4.2 Correlograms from July of 1993 and June of 1994

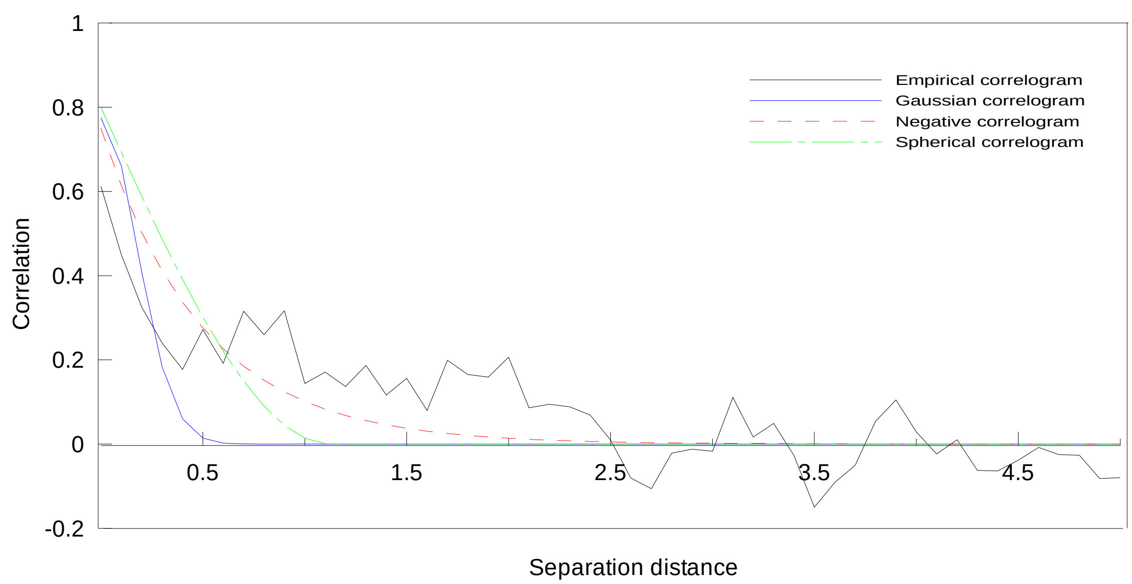


Figure 4.3 Correlograms from July of 1994 to June of 1995

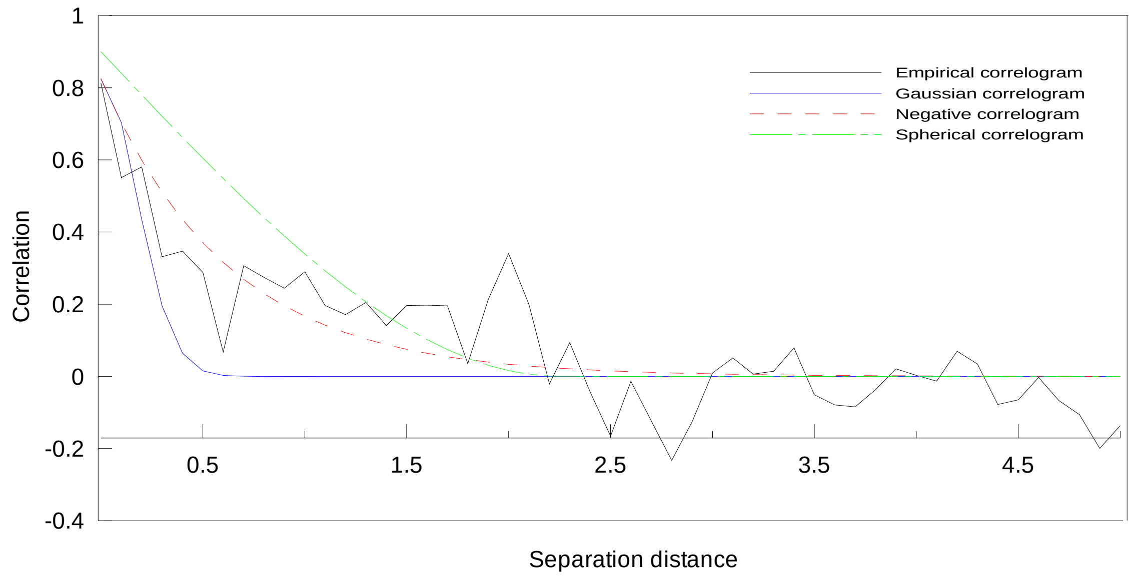


Figure 4.4 Correlograms from July of 1996 to June of 1997

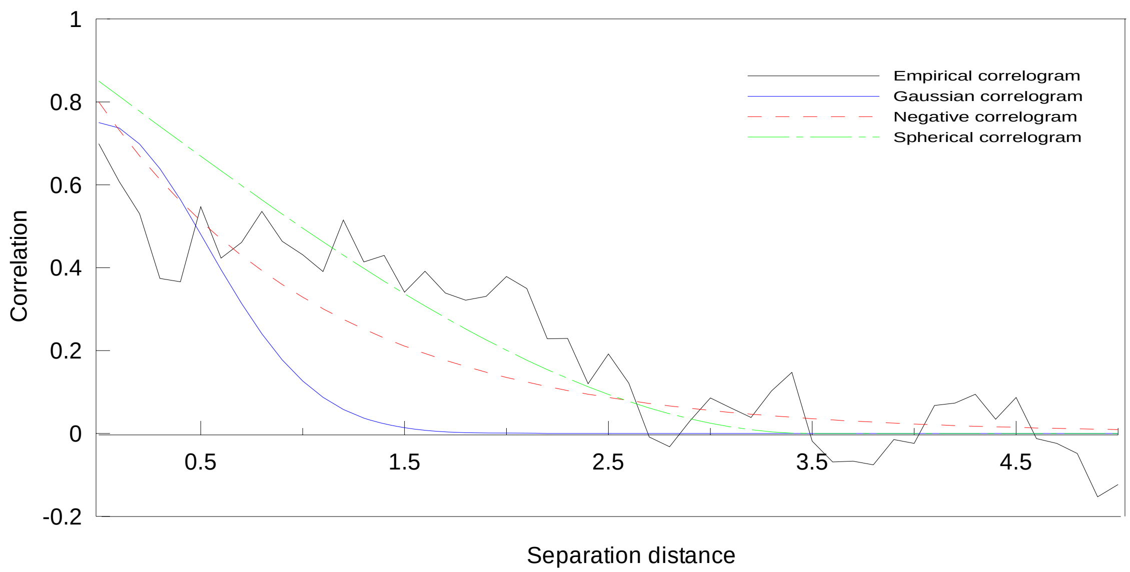


Figure 4.5 Correlograms from July of 1997 to June of 1998

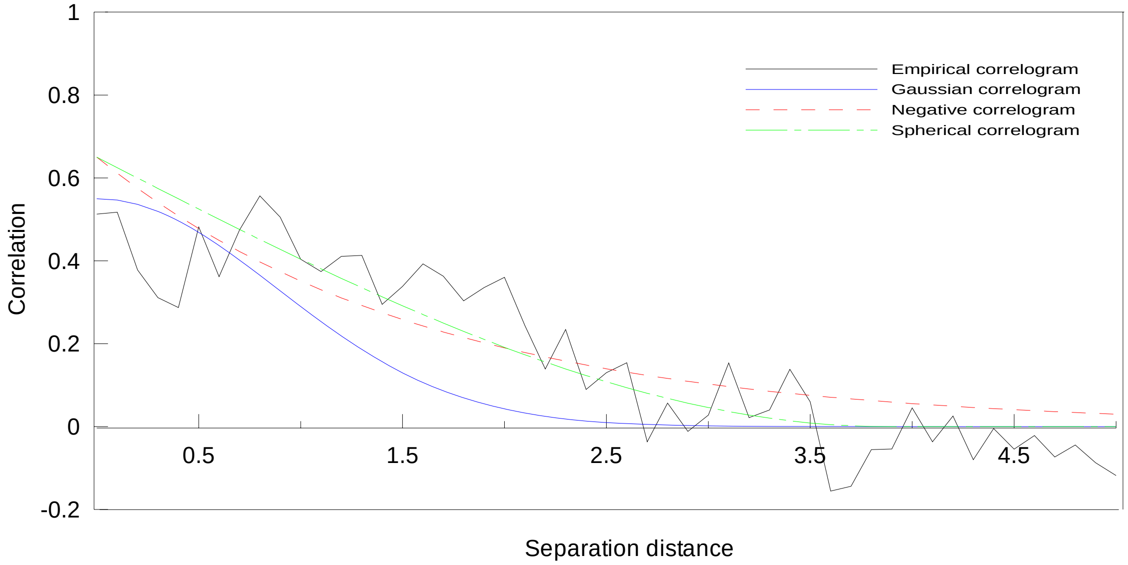


Figure 4.6 Correlograms July of 1998 to June of 1999

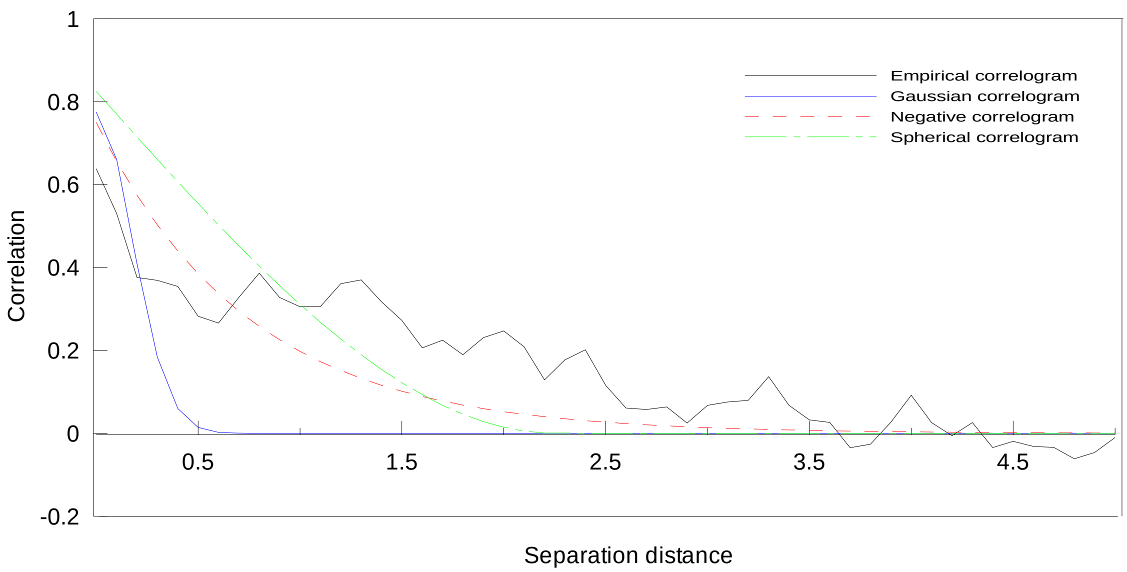


Figure 4.7 Correlograms from July of 1999 to June of 2000

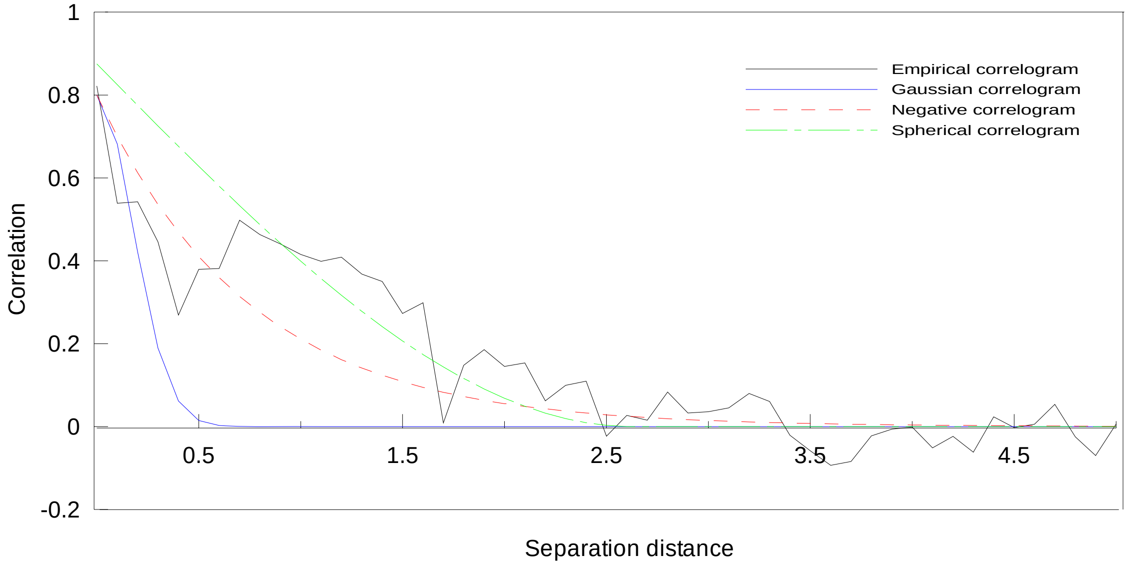


Figure 4.8 Correlograms from July of 2000 to June of 2001

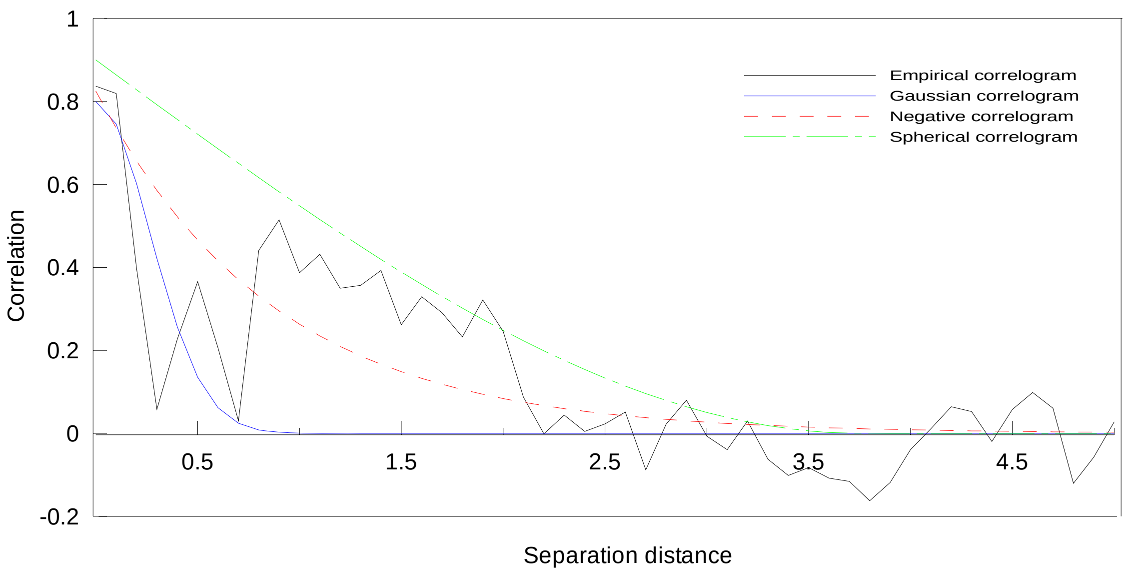


Figure 4.9 Correlograms from July of 2001 to June of 2002

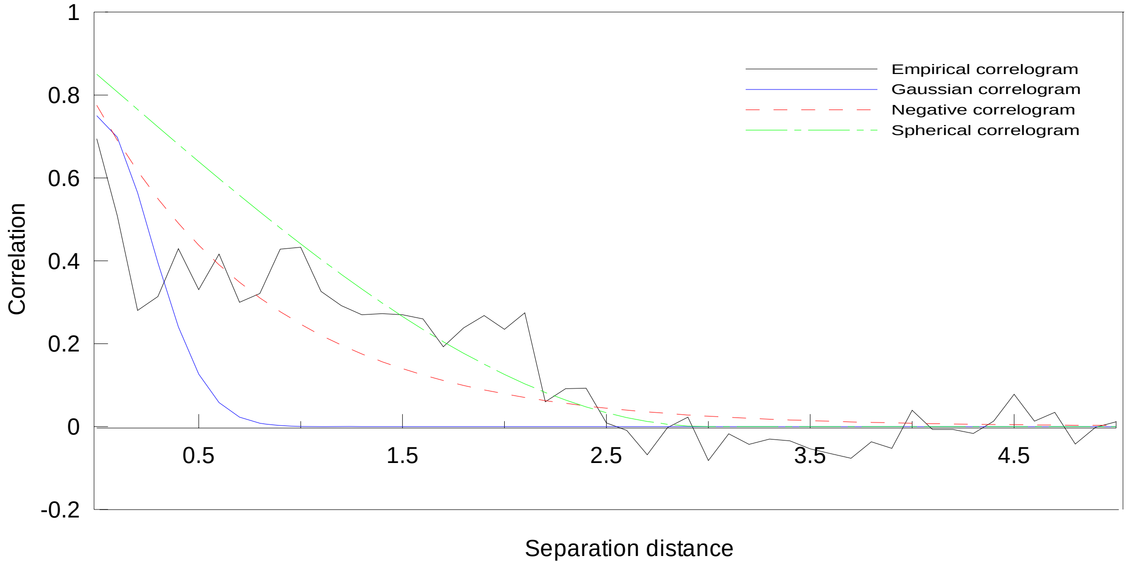


Figure 4.10 Correlograms from July of 2002 to June of 2003

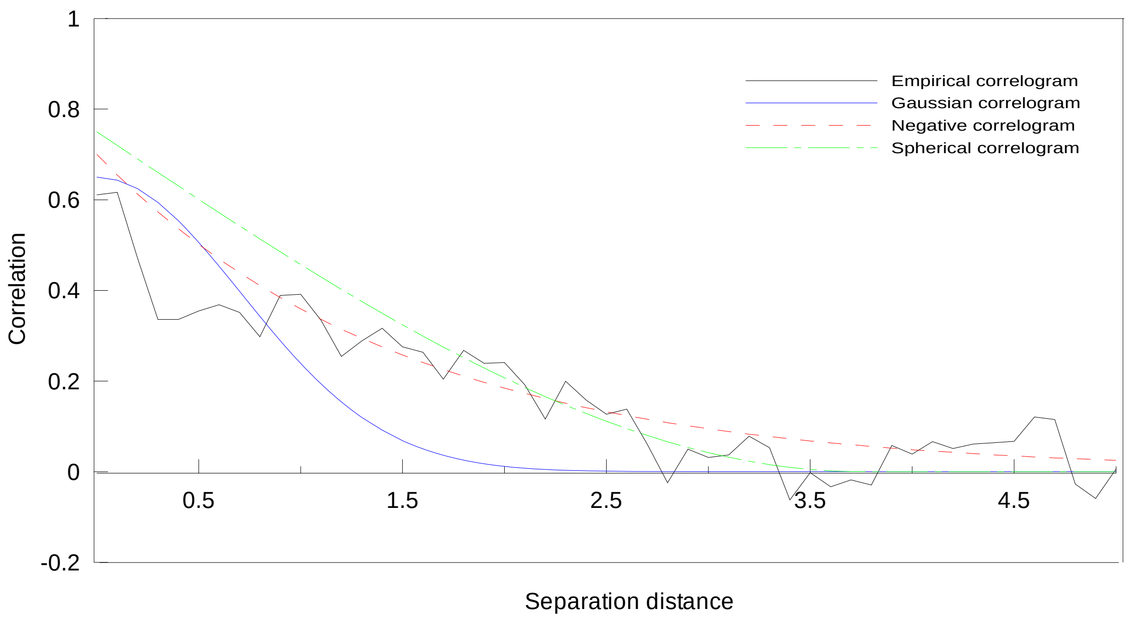


Table 4.11 Estimation and prediction results from OLS

Year	Jul 92-Jun 93	Jul 93-Jun 94	Jul 94-Jun 95	Jul 95-Jun 96	Jul 96-Jun 97	Jul 97-Jun 98	Jul 98-Jun 99	Jul 99-Jun 00	Jul 00-Jun 01	Jul 01-Jun 02	Jul 02-Jun 03
In-sample Size	1273	2839	2286	2460	961	704	2953	3234	1619	1256	1361
Ex-sample Size	500	500	500	500	500	500	500	500	500	500	500
Estimation											
CON	12.2620** (0.0731)	12.3950** (0.0506)	12.2070** (0.0638)	12.5580** (0.051)	12.6690** (0.0919)	12.9130** (0.098)	12.3650** (0.0525)	12.4690** (0.054)	12.6670** (0.0648)	12.5830** (0.067)	12.5580** (0.0638)
LEVEL	0.0024** (0.0007)	0.0044** (0.0005)	0.0032** (0.0007)	0.0054** (0.0006)	0.0035** (0.0012)	0.0031** (0.0013)	0.0058** (0.0006)	0.0057** (0.0006)	0.0043** (0.001)	0.0051** (0.001)	0.0076** (0.0009)
AGE	-0.0112** (0.0018)	-0.0092** (0.0011)	-0.0032** (0.0012)	-0.0065** (0.0011)	-0.0084** (0.0017)	-0.0125** (0.0018)	-0.0135** (0.0009)	-0.0098** (0.0009)	-0.0137** (0.0013)	-0.0112** (0.0013)	-0.0091** (0.0013)
FLRAREA	0.0037** (0.0001)	0.0033** (0)	0.0036** (0)	0.0045** (0)	0.0039** (0)	0.0048** (0.0001)	0.0046** (0)	0.0045** (0)	0.0048** (0.0001)	0.0043** (0)	0.0047** (0.0001)
FREEHOLD	0.2072** (0.0125)	0.2313** (0.0093)	0.3373** (0.012)	0.0294** (0.0071)	0.3395** (0.0178)	0.2864** (0.021)	0.2969** (0.0098)	0.3288** (0.0097)	0.3358** (0.0143)	0.4118** (0.0154)	0.3800** (0.0146)
DIS_PRI1	-0.0070 (0.0085)	0.0241** (0.0059)	0.0353** (0.0078)	0.0212** (0.0084)	0.0484** (0.0112)	0.0230** (0.0136)	0.0166** (0.0064)	0.0430** (0.0065)	0.0264** (0.0096)	0.0243** (0.011)	0.0243** (0.0105)
DIS_PRI2	0.0551** (0.0105)	0.0079 (0.0071)	0.0183** (0.0093)	-0.0523** (0.0049)	0.0161 (0.0131)	0.0036 (0.0165)	0.0012 (0.0074)	-0.0066 (0.008)	0.0208** (0.0109)	0.0006 (0.0124)	0.0039 (0.0117)
DIS_SEC	-0.0286** (0.0058)	-0.0270** (0.0043)	-0.0648** (0.0055)	-0.0294** (0.0024)	-0.0447** (0.0081)	-0.0147** (0.0079)	-0.0296** (0.0042)	-0.0241** (0.0039)	-0.0297** (0.0056)	-0.0252** (0.0063)	-0.0349** (0.0059)
DIS_JC	-0.0283** (0.0027)	-0.0263** (0.0022)	-0.0263** (0.0028)	-0.0374** (0.0054)	-0.0348** (0.0041)	-0.0289** (0.0047)	-0.0275** (0.0024)	-0.0284** (0.0026)	-0.0264** (0.0036)	-0.0244** (0.0038)	-0.0112** (0.0036)
MRT	0.0095 (0.0062)	-0.0272** (0.0046)	-0.0356** (0.0061)	-0.0001 (0.0032)	-0.0536** (0.009)	-0.0478** (0.011)	-0.0335** (0.0052)	-0.0237** (0.0052)	-0.0354** (0.0079)	-0.0566** (0.0078)	-0.0190** (0.0067)

CBD	-0.0041 (0.0041)	0.0104** (0.003)	0.0182** (0.0038)	0.0412** (0.0033)	-0.0007 (0.0059)	-0.0096 (0.006)	0.0007 (0.0031)	0.0001 (0.0032)	-0.0056 (0.0038)	-0.0071** (0.0036)	-0.0119** (0.0035)
DIS_RC	0.0244** (0.0047)	0.0369** (0.0034)	0.0619** (0.0042)	0.2957** (0.0102)	0.0496** (0.0063)	0.0179** (0.0061)	0.0295** (0.0033)	0.0304** (0.0037)	0.0286** (0.004)	0.0293** (0.0039)	0.0191** (0.0034)
BBQ	0.0744** (0.0114)	0.0004 (0.008)	-0.0299** (0.0103)	-0.0072 (0.0096)	-0.0024 (0.0166)	-0.0085 (0.0195)	0.0414v (0.0096)	0.0034 (0.0103)	0.0322** (0.0161)	0.0119 (0.0167)	0.0363** (0.016)
CARPARK	0.0239** (0.0119)	0.0050 (0.0083)	-0.0050 (0.0109)	0.0093 (0.0106)	0.0193 (0.0179)	0.0249 (0.0205)	0.0456** (0.011)	0.0652** (0.0117)	-0.0558** (0.0162)	-0.0131 (0.0178)	-0.0024 (0.0172)
GYM	0.0519** (0.0127)	0.0955** (0.0091)	0.1397** (0.0115)	0.0821** (0.0104)	0.0910** (0.0177)	0.0687** (0.0204)	0.0328** (0.0093)	0.0498** (0.0096)	0.0617** (0.0154)	0.0634** (0.0164)	0.0199 (0.0155)
JACUZZI	0.1124** (0.0268)	0.0699** (0.0157)	0.1444** (0.0197)	0.0771** (0.0159)	0.1185** (0.0264)	-0.0194 (0.0301)	0.0311** (0.0118)	0.0367** (0.0124)	0.0345** (0.0191)	0.0975** (0.0195)	0.0076 (0.0179)
FITNESS	-0.0150 (0.0145)	0.0403** (0.0103)	0.1077** (0.0126)	0.0566** (0.0116)	0.0319 (0.0189)	-0.0061 (0.0236)	-0.0229** (0.0103)	-0.0598** (0.0111)	-0.0135 (0.0166)	-0.0328** (0.0177)	-0.0191 (0.0165)
MINIMART	0.0088 (0.0156)	0.0224** (0.0109)	-0.0137 (0.0151)	-0.0281** (0.014)	-0.0432 (0.024)	-0.0691** (0.0269)	-0.0514** (0.0137)	-0.0787** (0.015)	0.0694** (0.0212)	-0.0254 (0.0233)	0.0249 (0.0226)
MPH	0.0009 (0.0144)	0.0014 (0.0097)	-0.0206 (0.0117)	0.0248** (0.0106)	0.0233 (0.0172)	0.0372** (0.0196)	0.0299** (0.0091)	0.0205** (0.0091)	0.0391** (0.0136)	0.0349** (0.0146)	0.0271** (0.0144)
PLAYGROU	-0.0404** (0.0123)	-0.0249** (0.009)	-0.0131 (0.0121)	-0.0247** (0.0112)	-0.0297 (0.0198)	-0.0015 (0.0219)	-0.0256** (0.0111)	-0.0067 (0.0119)	0.0568** (0.0189)	-0.0050 (0.0185)	-0.0314 (0.0197)
SAUNA	0.0478** (0.0131)	0.0374** (0.0096)	0.0152 (0.0116)	0.0320** (0.0106)	0.0268 (0.0176)	0.0044 (0.0199)	-0.0079 (0.0092)	0.0203** (0.0097)	-0.0397** (0.0138)	-0.0088 (0.0152)	-0.0031 (0.0142)
SQUASH	-0.0448** (0.0167)	-0.0175 (0.012)	-0.0040 (0.0156)	-0.0303** (0.0143)	-0.0282 (0.0243)	0.0000 (0.0277)	0.0272** (0.0111)	0.0294** (0.0112)	0.0476** (0.0158)	-0.0089 (0.0174)	-0.0398** (0.0172)
SWIMMING	0.0408** (0.0217)	0.0505** (0.0161)	0.0188 (0.0196)	0.0467** (0.018)	0.0598** (0.0331)	0.0426 (0.0415)	0.0825** (0.0227)	0.1446** (0.0213)	0.1941** (0.0323)	0.1191** (0.0329)	0.1090** (0.0318)

TENNIS	-0.0329** (0.0131)	-0.0020 (0.0101)	0.0173 (0.0137)	-0.0088 (0.0125)	0.0414** (0.0221)	-0.0322 (0.0265)	0.0449** (0.0119)	0.0920** (0.0132)	-0.0573** (0.0187)	0.1055** (0.0191)	0.0348** (0.019)
WADING	0.0500** (0.013)	0.0270** (0.0093)	0.0558** (0.0126)	0.0547** (0.0115)	0.0193 (0.0193)	0.0398** (0.0211)	0.0142 (0.0105)	0.0557** (0.0111)	0.0379** (0.0164)	-0.0542** (0.017)	0.0696** (0.0173)
SECURITY	0.0923** (0.0285)	0.0828** (0.0203)	0.1387** (0.0246)	0.1535** (0.0226)	0.1501** (0.0394)	0.1239** (0.047)	0.0786** (0.0249)	0.0331 (0.0211)	0.0516 (0.0305)	0.0335 (0.0293)	0.0065 (0.0285)
OTHERS	0.0043 (0.0145)	-0.0189** (0.0102)	0.0045 (0.0136)	-0.0078 (0.0121)	0.0405** (0.0211)	0.0032 (0.0242)	0.0425** (0.0115)	-0.0492** (0.0116)	-0.0718** (0.0187)	0.0101 (0.0204)	-0.0312 (0.0197)
TOTALUNI	0.0001 (0.3E-5)	0.0000 (0.3E-5)	0.0000 (0.2E-5)	0.0000 (0.3E-5)	0.0000 (0.2E-5)	0.0001** (0.4E-5)	0.0000 (0.3E-5)	0.0000 (0.4E-5)	-0.0001** (0.3E-5)	0.0000 (0.2E-5)	0.0000 (0.3E-5)
T2	0.0127 (0.0165)	0.0270** (0.009)	0.0606** (0.0115)	0.0106 (0.0113)	-0.0414** (0.0199)	-0.0701** (0.0211)	-0.0386** (0.0155)	0.0482** (0.0104)	0.0170 (0.0144)	-0.0338 (0.0258)	0.0065 (0.0151)
T3	0.0876** (0.016)	0.1082** (0.0094)	0.0544** (0.0135)	0.0609** (0.0108)	0.0060 (0.0189)	-0.1582** (0.0241)	0.0119 (0.0144)	0.1074** (0.0102)	-0.0689** (0.0157)	-0.0215 (0.0161)	-0.0602** (0.017)
T4	0.1431** (0.015)	0.1877** (0.0084)	0.1351** (0.0113)	0.1206** (0.0106)	-0.0256 (0.018)	-0.2501** (0.019)	0.1332** (0.0134)	0.0408** (0.0107)	-0.0980** (0.0151)	-0.0160 (0.0162)	-0.0529** (0.0165)
$\bar{R}^2$	0.8555	0.8377	0.8490	0.8546	0.8303	0.8205	0.8065	0.7836	0.8039	0.8150	0.7968
Prediction											
MSE	0.0242	0.0291	0.0387	0.0481	0.0383	0.0478	0.0382	0.0467	0.0530	0.0460	0.0456
The'U	0.0009	0.0011	0.0014	0.0017	0.0014	0.0017	0.0014	0.0017	0.0019	0.0017	0.0017
Mean ABE	0.1142	0.1228	0.1473	0.1512	0.1441	0.1453	0.1336	0.1542	0.1719	0.1524	0.1566
Median ABE	0.0916	0.0979	0.1154	0.1115	0.1105	0.0925	0.0952	0.1134	0.1431	0.1082	0.1112

Table 4.12 Estimation results from 2STAR Model

Year	Jul 92-Jun 93	Jul 93-Jun 94	Jul 94-Jun 95	Jul 95-Jun 96	Jul 96-Jun 97	Jul 97-Jun 98	Jul 98-Jun 99	Jul 99-Jun 00	Jul 00-Jun 01	Jul 01-Jun 02	Jul 02-Jun 03
In-sample Size	1273	2839	2286	2460	961	704	2953	3234	1619	1256	1361
Ex-sample Size	500	500	500	500	500	500	500	500	500	500	500
Variables											
Constant	12.4519 (0.1606)	11.7000 (0.1710)	12.0970 (0.1829)	11.5030 (0.1821)	12.8450 (0.1828)	12.9300 (0.2340)	11.3110 (0.1814)	11.4610 (0.1887)	12.1550 (0.2026)	12.8180 (0.1983)	12.0880 (0.1945)
Level	0.0045** (0.0046)	0.0054** (0.0006)	0.0062** (0.0007)	0.0071** (0.0007)	0.0072** (0.0012)	0.0067** (0.0016)	0.0070** (0.0007)	0.0056** (0.0007)	0.0038** (0.0010)	0.0069** (0.0011)	0.0045** (0.0010)
Age	-0.0083** (0.0048)	-0.0181** (0.0022)	-0.0080** (0.0020)	-0.0109** (0.0018)	-0.0168** (0.0019)	-0.0125** (0.0024)	-0.0164** (0.0013)	-0.0132** (0.0012)	-0.0156** (0.0014)	-0.0166** (0.0013)	-0.0141** (0.0013)
Flarea	0.0039** (0.0050)	0.0031** (0.0001)	0.0034** (0.0001)	0.0039** (0.0001)	0.0037** (0.0001)	0.0040** (0.0001)	0.0040** (0.0001)	0.0040** (0.0001)	0.0043** (0.0001)	0.0041** (0.0001)	0.0043** (0.0001)
Freehold	0.2748** (0.0152)	0.2889** (0.0206)	0.3394** (0.0223)	0.3165** (0.0206)	0.2510** (0.0214)	0.2493** (0.0291)	0.2745** (0.0170)	0.3224** (0.0168)	0.3168** (0.0191)	0.2977** (0.0187)	0.3131** (0.0180)
W1×Level	0.0033** (0.0054)	0.0020** (0.0008)	-0.0008 (0.0010)	0.0008 (0.0010)	0.0007 (0.0017)	-0.0032 (0.0026)	-0.0011 (0.0010)	-0.0005 (0.0009)	-0.0029* (0.0014)	-0.0011 (0.0017)	0.0026 (0.0014)
W1×Age	-0.0085** (0.0056)	-0.0003 (0.0025)	-0.0068** (0.0024)	-0.0093** (0.0020)	-0.0031 (0.0026)	-0.0080** (0.0038)	-0.0055** (0.0015)	-0.0093** (0.0014)	-0.0026 (0.0019)	-0.0043** (0.0018)	-0.0053** (0.0017)
W1×Flra.	0.0007** (0.0058)	0.0003** (0.0001)	0.0006** (0.0001)	0.0004** (0.0001)	0.0005* (0.0002)	0.0004 (0.0003)	0.0004** (0.0001)	0.0005** (0.0001)	0.0004** (0.0002)	0.0004** (0.0002)	0.0002 (0.0002)
W1×Freeh.	-0.0059 (0.0160)	-0.0292 (0.0224)	-0.0307 (0.0249)	-0.0034 (0.0230)	0.0370 (0.0289)	-0.0155 (0.0415)	0.0043 (0.0194)	-0.0054 (0.0190)	-0.0252 (0.0245)	0.1119** (0.0256)	0.0221 (0.0242)
W2×Level	-0.0061**	-0.0053*	-0.0009	-0.0036**	-0.0028	0.0036	-0.0056**	0.0001	0.0016	-0.0016	0.0002

	(0.0062)	(0.0011)	(0.0014)	(0.0013)	(0.0020)	(0.0031)	(0.0012)	(0.0012)	(0.0016)	(0.0020)	(0.0018)
W2×Age	0.0063**	0.0094**	0.0100**	0.0119**	0.0130**	0.0125**	0.0137**	0.0154**	0.0109**	0.0129**	0.0176**
	(0.0034)	(0.0020)	(0.0020)	(0.0017)	(0.0024)	(0.0034)	(0.0013)	(0.0012)	(0.0017)	(0.0017)	(0.0015)
W2×Flra.	-0.0017**	-0.0013**	-0.0022**	-0.0025**	-0.0026**	-0.0016**	-0.0018**	-0.0017**	-0.0018**	-0.0013**	-0.0028**
	(0.0066)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0004)	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)
W2×Freeh.	-0.2026**	-0.1666**	-0.2458*	-0.2433*	-0.2028**	-0.1148**	-0.1501**	-0.1661**	-0.1524**	-0.1663**	-0.2084**
	(0.0068)	(0.0152)	(0.0185)	(0.0175)	(0.0275)	(0.0381)	(0.0162)	(0.0154)	(0.0231)	(0.0242)	(0.0238)
T×Level	0.0069**	0.0091**	0.0031	0.0000	0.0124**	-0.0146**	0.0001**	0.0111**	-0.0058	0.0062	-0.0046
	(0.0070)	(0.0026)	(0.0027)	(0.0029)	(0.0045)	(0.0068)	(0.0027)	(0.0028)	(0.0039)	(0.0040)	(0.0039)
T×Age	-0.0171**	-0.0365**	0.0008	-0.0301**	-0.0017	-0.0101	-0.0045	0.0116**	-0.0107**	-0.0116**	-0.0205**
	(0.0072)	(0.0070)	(0.0062)	(0.0057)	(0.0060)	(0.0083)	(0.0042)	(0.0040)	(0.0056)	(0.0044)	(0.0048)
T×Flra.	0.0033**	0.0016**	0.0022**	0.0035**	0.0032**	0.0017**	0.0025**	0.0013**	0.0008**	0.0021**	0.0036**
	(0.0074)	(0.0002)	(0.0002)	(0.0003)	(0.0004)	(0.0007)	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0004)
T×Freeh.	0.0075	0.0032	-0.0167	-0.1517*	0.1624*	-0.1256	0.0660	0.4605**	0.2283**	0.1468**	0.2173**
	(0.0076)	(0.0643)	(0.0734)	(0.0710)	(0.0876)	(0.1073)	(0.0575)	(0.0530)	(0.0646)	(0.0670)	(0.0721)
S×T×Level	0.0277**	-0.0070	0.0287**	-0.0313**	0.0060	-0.0063	-0.0438**	0.0298**	-0.0011	-0.0033	-0.0077
	(0.0078)	(0.0052)	(0.0053)	(0.0055)	(0.0060)	(0.0143)	(0.0051)	(0.0052)	(0.0075)	(0.0086)	(0.0077)
S×T×Age	0.0279**	-0.0774**	0.0197**	0.0268**	-0.0005	0.0240	-0.0056	0.0022	0.0364**	0.0108**	0.0067
	(0.0080)	(0.0109)	(0.0094)	(0.0078)	(0.0108)	(0.0180)	(0.0073)	(0.0052)	(0.0056)	(0.0058)	(0.0077)
S×T×Flra.	0.0023**	0.0021**	0.0018**	-0.0022**	0.0004	0.0025**	-0.0004	-0.0019**	-0.0025*	-0.0016**	0.0004
	(0.0082)	(0.0004)	(0.0004)	(0.0005)	(0.0008)	(0.0013)	(0.0005)	(0.0005)	(0.0009)	(0.0008)	(0.0009)
S×T×Freeh.	0.3183**	0.1931**	0.3117**	-0.2626**	0.0232	0.3124**	-0.3136**	0.6301**	0.1133	0.3112**	0.3017**
	(0.0084)	(0.0681)	(0.0751)	(0.0888)	(0.1063)	(0.1478)	(0.0822)	(0.0706)	(0.0940)	(0.1019)	(0.1002)
T×S×Level	-0.0285**	-0.0056	0.0264**	-0.0034	0.0013	-0.0232**	-0.0261**	0.0045	0.0197**	-0.0021	0.0206**
	(0.0086)	(0.0038)	(0.0047)	(0.0045)	(0.0075)	(0.0110)	(0.0041)	(0.0041)	(0.0057)	(0.0068)	(0.0060)

T×S×Age	0.0187 (0.0088)	-0.0098 (0.0091)	-0.0007 (0.0092)	0.0247 (0.0075)	-0.0262 (0.0094)	0.0303 (0.0123)	-0.0062 (0.0052)	-0.0195 (0.0050)	0.0331 (0.0067)	0.0151 (0.0061)	0.0088 (0.0059)
T×S×Flra.	0.0019** (0.0090)	0.0023** (0.0003)	0.0034** (0.0004)	0.0010** (0.0005)	0.0019** (0.0007)	0.0018 (0.0010)	0.0010** (0.0004)	0.0019** (0.0004)	0.0004 (0.0007)	-0.0002 (0.0007)	-0.0003 (0.0008)
T×S×Freeh.	0.2091** (0.0092)	0.1708** (0.0731)	0.3754** (0.0858)	0.1889** (0.0840)	0.1834** (0.1100)	0.2431 (0.1364)	-0.0489 (0.0651)	-0.0429 (0.0605)	0.1143 (0.0806)	0.3041** (0.0888)	0.3022** (0.0876)
$W1\times Y$	0.0043** (0.0094)	0.0017 (0.0023)	0.0025 (0.0025)	0.0037 (0.0022)	-0.0064** (0.0029)	-0.0020 (0.0043)	0.0033** (0.0018)	0.0032** (0.0017)	0.0009 (0.0023)	-0.0041 (0.0024)	-0.0017** (0.0024)
$W2\times Y$	0.5137** (0.0096)	0.6210** (0.0239)	0.7891** (0.0231)	0.7671** (0.0207)	0.8077** (0.0302)	0.5151** (0.0508)	0.6443** (0.0201)	0.6541** (0.0189)	0.6158** (0.0311)	0.5975** (0.0330)	0.7332** (0.0296)
$T\times Y$	-0.5724** (0.0098)	-0.5293** (0.0282)	-0.7057** (0.0283)	-0.5935** (0.0265)	-0.8007** (0.0349)	-0.4572** (0.0534)	-0.4671** (0.0252)	-0.5723** (0.0251)	-0.5261** (0.0353)	-0.6233** (0.0375)	-0.6957** (0.0346)
S×T ×Y	0.0499** (0.0100)	0.0452** (0.0153)	-0.0853** (0.0150)	0.0459** (0.0146)	-0.0238 (0.0184)	-0.0745** (0.0253)	0.0444** (0.0142)	-0.0257** (0.0115)	-0.0221 (0.0177)	-0.0167 (0.0163)	-0.0323** (0.0177)
$T\times S\times Y$	-0.0101** (0.0102)	-0.0434** (0.0208)	-0.0673** (0.0184)	-0.0815** (0.0202)	0.0031 (0.0194)	-0.0561** (0.0277)	-0.0557** (0.0183)	-0.0109 (0.0170)	-0.0439** (0.0208)	0.0130 (0.0203)	-0.0051 (0.0212)
$\overline{R}^2$	0.8401	0.8204	0.8663	0.8495	0.8625	0.7568	0.8157	0.8135	0.8232	0.8381	0.8306

Table 4.13 Estimation result from 2STAR model for the whole estimation sample

Variables	Beta	Std	p-value
Constant	1.8491	0.0992	0.0000
Level	0.0056	0.0003	0.0000
Age	-0.0202	0.0011	0.0000
Flrarea	0.0035	0.0000	0.0000
Freehold	0.3527	0.0134	0.0000
W1×Level	0.0007	0.0004	0.0568
W1×Age	0.0024	0.0011	0.0296
W1×Flrarea	0.0004	0.0000	0.0000
W1×Freehold	-0.0650	0.0139	0.0000
W2×Level	-0.0037	0.0005	0.0000
W2×Age	0.0120	0.0005	0.0000
W2×Flrarea	-0.0019	0.0001	0.0000
W2×Freehold	-0.1904	0.0069	0.0000
T×Level	-0.0030	0.0011	0.0051
T×Age	-0.0066	0.0022	0.0023
T×Flrarea	-0.0014	0.0001	0.0000
T×Freehold	-0.2627	0.0353	0.0000
S×T×Level	-0.0044	0.0022	0.0451
S×T×Age	-0.0051	0.0020	0.0110
S×T×Flrarea	0.0004	0.0001	0.0046
S×T×Freehold	0.2551	0.0292	0.0000
T×S×Level	-0.0070	0.0017	0.0001
T×S×Age	0.0150	0.0027	0.0000
T×S×Flrarea	0.0004	0.0002	0.0111
T×S×Freehold	0.2175	0.0398	0.0000
W1×*Y*	0.0032	0.0011	0.0023
W2×*Y*	0.7455	0.0071	0.0000
T×*Y*	0.3820	0.0116	0.0000
S×T ×Y	-0.0940	0.0115	0.0000
T×*S*×*Y*	-0.1938	0.0144	0.0000
$\bar{R}^2$	0.8366		
SSE	824.55		
Number of observations	20,946		
Number of variables	30		

Figure 5.1 Price index in Central Area

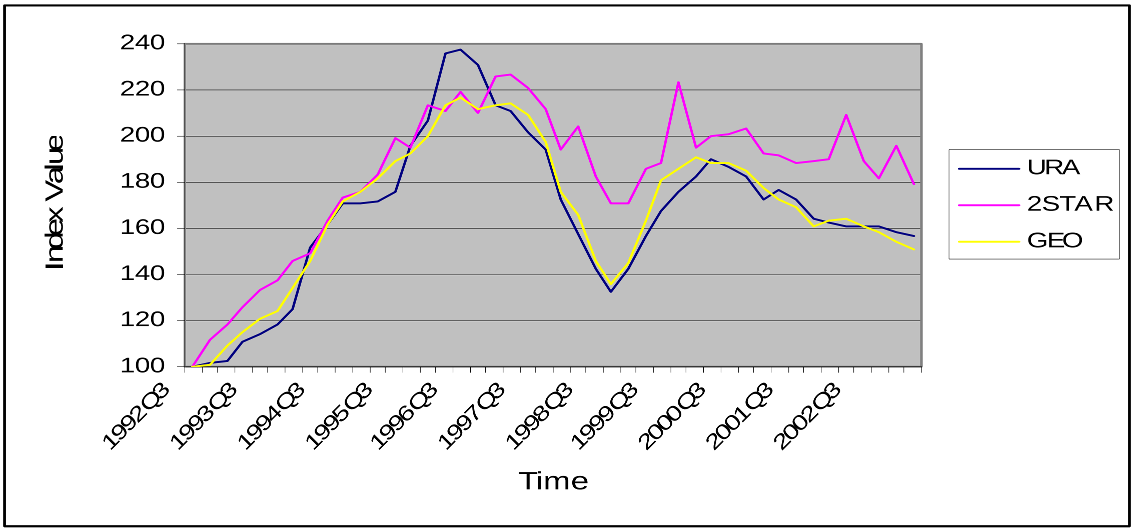


Figure 5.2 Price index in East Area

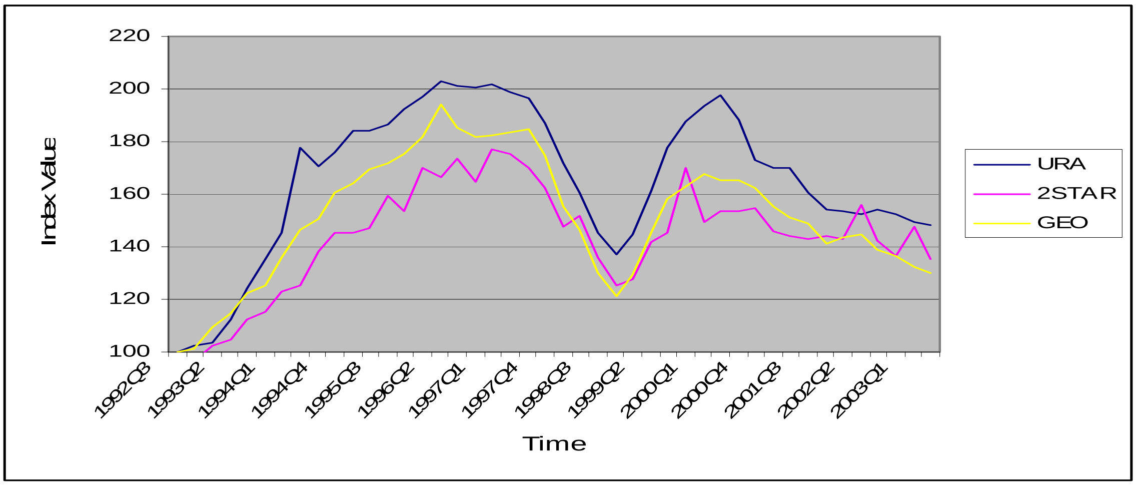


Figure 5.3 Price index in West Area

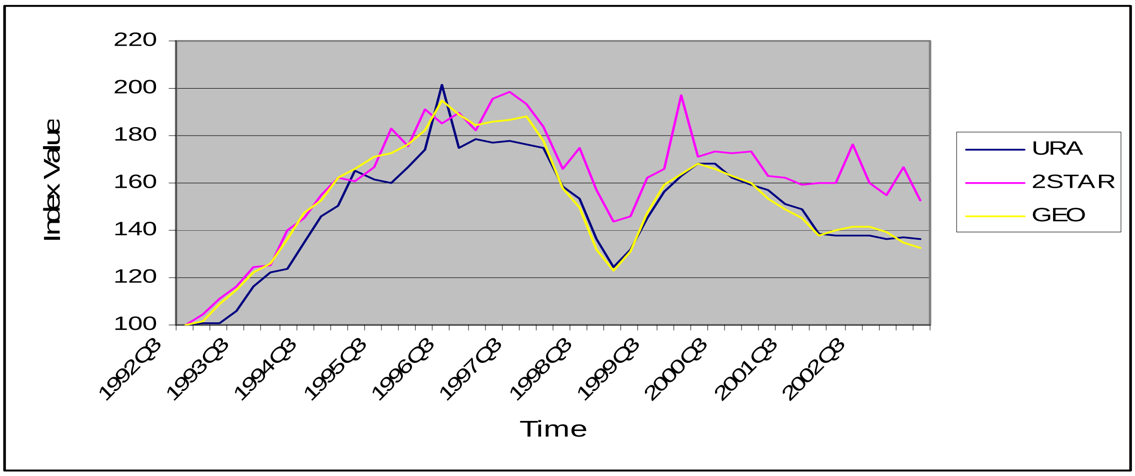


Figure 5.4 Price index in Northeast Area