2002

Three essays on real estate research

Ming-Long Lee
Louisiana State University and Agricultural and Mechanical College

THREE ESSAYS ON REAL ESTATE RESEARCH

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Interdepartmental Program in Business Administration (Finance)

by
Ming-Long Lee
B.B.A., Soochow University, 1990
Master of Business in Property, University of South Australia, 1995
August, 2002

ACKNOWLEDGEMENTS

I am deeply indebted to my committee chairman, Dr. R. Kelley Pace, for his dedication, effort, and guidance through my doctoral program and in completing this dissertation. I would also like to thank the members of my dissertation committee, Drs. James Garand, R. Carter Hill, Ji-Chai Lin, Gary C. Sanger, and V. Carlos Slawson, Jr., for their insights and comments into this research.

Special thanks are to Dr. Chip Ryan for his excellent teaching in corporate finance. In his class, the third essay in this dissertation was developed. I would like to dedicate this work to my father, to my mother, to my wife, Shou-Lu Lee, to my daughter, Jenny C. Lee, to my brother, and to my sisters. Their love and financial help are vital for me to accomplish this work.

TABLE OF CONTENTS

ABSTRACT

This dissertation consists of essays relating to the three important real estate research topics: spatial statistics, mortgages, and real estate investment trusts (REITs). In the first essay, "Spatial Distribution of Retail Sales," we apply retail gravity models to examine the spatial distribution of retail sales for a retail chain in the Houston market. Unlike previous empirical studies, our study models both the spatial dependencies among both consumers and retailers. Our results show both the spatial dependencies have significant impacts on the estimates of parameters in retail gravity models. Contrary to the suggestions of Guitschi (1981) as well as Eppli and Shilling (1996), our results show the importance of the distance parameter in retail gravity models may be understated about 68%. Thus, previous studies may overestimate the deterministic extent of trade areas and, thus understate the importance of good locations.

The second essay, "Local Housing Prices and Mortgage Refinancing in US Cities," has implications the valuation of mortgages, in particular mortgage-backed securities (MBS). This essay provides additional evidence that house prices significantly impact aggregate refinancing and thus directly impact mortgage termination. Previous studies typically focus on the effect of negative appreciation on refinancing. In contrast to previous studies, this essay provides empirical evidence that a positive house price appreciation may motivate borrowers to refinance for capital structure-based or consumption reasons.

The third essay, "Monitoring and Dividend Policy for REITs under Asymmetric Information," examines the interaction between monitoring and two competing explanations for REIT dividend policy under asymmetric information. Specifically, the REIT empirical literature offers two competing theories for the level of dividend payouts under asymmetric information. Wang, Erickson, and Gau (1993) confirm the agency-cost theories. Bradley, Capozza, and

Seguin (1998) support the signaling explanations dominating agency cost explanations. In this essay, we demonstrate that by introducing proxies for taxable income and monitoring, we provide evidence that supports agency cost explanations for ineffectively monitored REITs. Furthermore, in contrast to Bradley, Capozza, and Seguin (1998), we show agency cost explanations dominate signaling explanations for these REITs.

CHAPTER 1

INTRODUCTION

Over the past 14 years, *The Journal of Real Estate Finance and Economics* (*JREFE*) has devoted special issues to important research topics of real estate.[1] In particular *JREFE* devoted its volume 23 number 2 published in year 2001 to mortgage modeling, its volume 20 number 2 published in year 2000 to real estate investment trusts (REITs), and its volume 17 number 1 published in year 1998 to spatial statistics and real estate. Currently, *JREFE* has a second forthcoming special issue on spatial statistics and real estate.

Other real estate journals also reflect the importance of these three research topics. Particularly *The Journal of Housing Research* (*JHR*) published a special issue (volume 6 issue 1) on mortgage-backed securities (MBS) pricing in 1995. *The Journal of Real Estate Research* (*JRER*) devoted its volume 6 number 3 to REITs in 1998.[2] In year 2002 *The Journal of Real Estate Portfolio Management* (*JREPM*) also published a special issue (volume 7 number 1) on REITs.

This dissertation consists of essays relating to the three important real estate research topics: spatial statistics, mortgages, and REITs. The first essay is an empirical application of spatial statistics to retail location. The second essay is about housing prices and mortgage refinancing, and the third essay is about dividend policies of real estate investment trusts.

The first essay, "Spatial Distribution of Retail Sales," applies retail gravity models to examine the spatial distribution of retail sales for a retail chain in the Houston market. Although many studies have empirically examined the concept of retail gravitation, the literature has

---

[1] JREFE published only 12 special issues in its total 78 issues from 1988 (volume 1, issue 1) to 2001 (volume 25, issue 1) (*JREFE*, 2002).

[2] REITs and mortgages were ranked the 6th and 10th popular research topics in the *JRER* from 1986 to 1996 (Jud, 1996).

inconsistent evidence on the importance of distance between consumers and retailers (Mejia and Benjamin, 2002). In particular, Guitschi (1981) as well as Eppli and Shilling (1996) suggest that the distance parameter may be significantly overstated in most previous retail gravity studies.

Different from previous empirical studies, this essay models both the spatial dependencies among consumers and stores. The essay shows both the spatial dependencies have significant impacts on the estimates of parameters in retail gravity models. Contrary to Guitschi's (1981) as well as Eppli and Shilling's (1996) suggestions, the results of the first essay show the importance of the distance parameter in most retail gravity models may be understated. In particular, this essay applies spatial statistics and shows the Ordinary Least Square (OLS) underestimates the distance parameter about 68% compared with the spatial simultaneous autoregressive error (SAR in errors).

The title of the second essay is "Local Housing Prices and Mortgage Refinancing in US Cities." This essay has implications the valuation of mortgages, in particular mortgage-backed securities (MBS). The proper valuation of MBS rests on the ability to model mortgage termination risk. Mortgage termination risk consists of default risk, refinancing risk, and mobility-related prepayment risk. Recent MBS valuation models have recognized that house prices are another important determinant of mortgage termination in addition to interest rates. Furthermore, refinancing appears to be the strongest means for house prices to affect mortgage termination (Mattery and Wallace, 1998). Nevertheless few studies investigate the relation between house price changes and aggregate refinancing activities. Two exceptions are Mattery and Wallace (1998) and Bennett, Keane, and Mosser (1999).

Mattery and Wallace (1998) provide evidence that a decrease in house prices significantly decreases county-level aggregate refinancing activities in California. On the other

hand, Bennett, Keane, and Mosser (1999) find no significant relation between house prices with the seasonally adjusted Mortgage Banker Association Refinance Index. The second essay provides evidence that an increase in house prices has a significant positive effect on city-level aggregate refinancing activities. Furthermore, the essay provides empirical evidence that house price appreciation may motivate borrowers to refinance for capital structure-based reasons.

The third essay, "Monitoring and Dividend Policy for REITs under Asymmetric Information," examines the interaction between monitoring and two competing explanations for REIT dividend policy under asymmetric information. Specifically, both signaling and agency-cost explanations have received empirical support in the REIT empirical literature. Wang, Erickson, and Gau (1993) provide evidence supporting the agency-cost explanation on REITs, while Bradely, Capozza, and Seguin (1998) provide evidence supporting the signaling explanation. Furthermore, Bradely, Capozza, and Seguin (1998) also show that signaling explanations dominate agency cost explanations for REITs.

By introducing proxies for taxable income and monitoring, this essay provides evidence that supports agency cost explanations for ineffectively monitored REITs. Furthermore, in contrast to Bradley, Capozza, and Seguin (1998), we show agency cost explanations dominate signaling explanations for these REITs. In other words, agency-cost explanations dominate when effective non-dividend monitoring does not exist. This result is consistent with Easterbrook's (1984) rationale of substitution among agency-cost control devices.

1.1 References

Bennett, Paul, Frank Keane, and Patricia C. Mosser. (1999). "Mortgage Refinancing and the Concentration of Mortgage Coupons," *Current Issues in Economics and Finance* 5, 1-6.

Bradley, Michael, Dennis R. Capozza, and Paul J. Seguin. (1998). "Dividend Policy and Cash-flow Uncertainty," *Real Estate Economics* 26, 555-580.

Dubin, Robin, R. Kelley Pace, and Thomas G. Thibodeau. (1999). “Spatial Autoregression Techniques for Real Estate Data,” *Journal of Real Estate Literature* 7, 79-95.

Eppli, Mark J., and James D. Shilling. (1996). “How Critical Is a Good Location to a Regional Shopping Center?” *The Journal of Real Estate Research* 12, 459-468.

Gautschi, David A. (1981). “Specification of Patronage Models for Retail Center Choice,” *Journal of Marketing Research* 18, 162-174.

Jud, G. Donald. (1996). “The Journal of Real Estate Research: A Ten-Year Review,” *The Journal of Real Estate Research* 12, 249-313.

Mattery, Joe, and Nancy Wallace. (1998). “Housing Prices and the (In)stability of Mortgage Prepayment Models: Evidence from California,” FRBSF Working Papers No. 98-05, 1-49.

Pace, R. Kelley, and Ronalid Barry. (1997). “Sparse Spatial Autoregressions,” *Statistics and Probability Letters* 33, 291-297.
The Journal of Real Estate Finance and Economics. (2002). http://www.jrefe.org.

Wang, Ko, and John Erickson, and George W. Gau. (1993). “Dividend Polices and Dividend Effects for Real Estate Investment Trusts,” *Journal of the American Real Estate and Urban Economics Association* 21, 185-201.

CHAPTER 2

SPATIAL DISTRIBUTION OF RETAIL SALES

2.1 Introduction

When considering opening a new store, retail chain store executives take a holistic view of market performance across their entire store network. In particular, they wish to avoid opening a profitable store at the expense of existing stores. To avoid such a loss, the executives need the spatial distribution information of customers and competitors to accurately define trade areas for site selection. In addition, managers of individual stores could use the spatial distribution of customers and competitors to promote sales. From an overall market perspective, the technology of forecasting sales can affect the location premia of retail properties.

We apply retail gravitation notions to examine empirically the spatial distribution of retail sales.[1] Seventy years ago Reilly (1931; cited in Huff, 1965) published his seminal proposition, known as "the law of retail gravitation." Retail gravity models draw an analogy with Newton's gravitational law to account for human behaviors related to shopping activities. In retail gravity models, various store features such as size attract customers, just as larger astronomical bodies have greater gravitational force. Distance between the customers and the store diminishes this attraction, just as gravity diminishes with distance. The simple functional form suggested by the analogy with gravity provided the prime contribution of Reilly's (1931; cited in Brown, 1992) work.

Many studies have empirically examined this concept of retail gravitation.[2] Nevertheless the literature often differs on the importance of the distance between retailers and consumers

[1] Hardin and Wolverton (2001) provide evidence that retail gravitation affects rental rates.
[2] Examples of these studies are Gautschi (1981), Okoruwa, Nourse, and Terza (1988 and 1994; ONT hereafter), and Eppli and Shilling (1996).

(Mejia and Benjamin, 2002). In particular, Guitschi (1981) as well as Eppli and Shilling (1996) suggest that the distance parameter may be significantly overstated in previous retail gravity research.

Studies have recognized that the spatial distribution of consumers and retailers influence the calibration outcomes of retail gravity models (Brown 1992). However, the existing studies of retail gravity assume no spatial dependencies among consumers and retailers, as well as independence among errors after controlling for distance. Therefore, the estimated parameters in the existing studies may reflect not only the nature of the spatial interaction between retailers and consumers but also their spatial distribution (Brown, 1992).

In this study, we apply spatial statistics to estimate a retail gravity model. [3] In particular, we model the spatial dependencies with a spatial simultaneous autoregressive error (SAR in errors) model among both consumers and retailers in a gravity model. The spatial dependency among retailers reflects important retailer and consumer behaviors in the choice of retail shopping trips, such as the clustering of heterogeneous retailers and the agglomeration of homogeneous retailers at site. The spatial dependence among consumers reflects the spatial autocorrelation of consumer shopping behavior arising from unobservable variables such as information sharing among consumers. To estimate the dependencies, we employ the spatial simultaneous autoregressive error model (Ord, 1975) to capture the spatial dependencies among consumers and stores.

Our results confirm the importance of modeling both the spatial dependencies in a retail-gravity model. When the spatial dependence among consumers (retailers) is explicitly taken into account, the estimated parameters of variables pertaining to consumers (retailers) change their magnitude considerably, and several reverse their signs. Compared with SAR in errors, OLS

[3] Porojan (2001) applied spatial statistics to estimate gravity models of international trade flows.

significantly underestimates the magnitude of the distance parameter. Contrary to Guitschi (1981) as well as Eppli and Shilling (1996), the results show the importance of the distance parameter may be understated about 68% in previous studies. Our results imply previous studies may overestimate the deterministic extent of trade areas for retail stores, and thus understate the importance of good locations.

The rest of this paper is organized as follows: Section 2 discusses retail gravitation and the spatial dependencies; section 3 describes retail sales data and empirical methodology; section 4 presents the empirical results; and section 5 concludes with the key results.

2.2 Retail Gravitation and Spatial Dependencies

Social scientists have drawn an analogy between the spatial interaction of people and Newton's law of gravity in physics. Approximately seventy years ago, Reilly (1931; cited in Brown, 1992) formally applied the Newton's gravity concept to retail geography. After that, many models of shopping behavior have been developed based on the concept of retail gravitation.[4] Nevertheless, most of these models, in their general forms, relate the interaction (shopping trips or expenditures) between retail store b and consumer c, I_{bc}, to the characteristics of store b, m_b, and the characteristics of consumer c, m_c, and the separation measurement between b and c, d_{bc}, in the manner of Equation (1):

$$I_{bc} = \lambda m_b^{\beta_b} m_c^{\beta_c} d_{bc}^{\beta_d} \tag{1}$$

where λ is a constant, and β_b, β_c, and β_d are parameters to be estimated.[5]

Many earlier studies include only store size and distance in their gravity models.[6] Examples are Huff's (1962) Lakshmanan and Hansen's (1965) models. Using survey data of

[4] See Brown (1992) for a list of studies of retail gravity models.
[5] Gravity models in this form can be applied to all sorts of spatial interaction behavior such as retail shopping, and population migration (Fotheringham and Webber, 1980).

shopping trips, Guitschi (1981) calibrated Huff's (1962) variation of Equation (1). He suggested that previous studies omitting other retail center variables might overstate the distance parameter in retail gravity models. Stanley and Sewall (1976) calibrated Huff's variation on single stores in a retail chain. Same to Kolter (1984, cited in Stanely and Sewall), they did not find that store size contributed significantly to estimates of store patronage. They, as well as Kolter (1984, cited in Stanely and Sewall), conclude that Huff's model is of limited value in estimating sales potential for single stores. Obtaining actual sales data, Eppli and Shiling (1996) calibrate Lakshmanan and Hansen's (1965) variant with an interactive approach and OLS. They find store location (proximity to the competition) is of little importance and conclude the distance parameter for retail gravity models may be significantly overstated.[7]

More recent research incorporates more characteristics of stores and consumers in retail gravity models. For example, Okoruwa, Nourse, and Terza (1988 and 1994) include retail center variables such as age and type, as well as economic and demographic characteristics of shoppers in estimating shopping trip frequencies obtained from a survey. Okoruwa, Nourse, and Terza (1988 and 1994) calibrate Equation (1) with a Possion regression. Contradicting the typical hypothesis of retail gravity models, they find that retail center size exerts a negative influence on patronization rates.

Some studies have recognized that the spatial distribution of origins and destinations influence the calibration outcomes of gravity-type models that study the interaction of individuals between two places (Curry, 1972; Brown 1992). The estimated parameters of gravity-type models reflect not only the nature of the spatial interaction between origins and destinations, but also their spatial distribution (Curry, Griffith, and Sheppard, 1975 and 1976;

---

[6] See Okoruwa, Nourse, and Terza (1988) for a list of such studies.
[7] In most retail gravity models, location is consumers' distance to retailers.

Fotheringham and Webber, 1980; Brown, 1992). In particular, Curry (1972) shows both the spatial dependencies among origins and destinations can influence the estimated parameters of gravity-type models. When the spatial dependencies are confounded with the estimated parameters, their values are not comparable from place to place (CGS, 1975). In other words the calibration results from a particular area are not applicable to anywhere else because the parameters do not measure their own "true" value.

In the context of retail shopping, the spatial dependence among destinations is the spatial dependence among stores. This spatial dependence reflects the clustering of retailers, accessibility, visibility of a retail site, and retail demand externalities within a shopping center. Clustering among heterogeneous retailers reflects multi-purpose shopping behavior of consumers to reduce total travel costs, and clustering of homogeneous retailers reflects comparison-shopping behavior (Eppli and Benjamin, 1994).[8] Studies have established the importance of these retailer and consumer behaviors in the choice of retail shopping trips (Eppli and Benjamin, 1994). Nevertheless, most of previous empirical studies did not incorporate these behaviors in retail gravity models.[9]

On the other hand, the spatial dependence among consumers equates to the spatial dependence among origins in the context of retail shopping. This dependence reflects the unobservable shopping patterns of a neighborhood caused by clustering of similar consumer populations and similar shopping environments. Clustering of similar consumer populations reflects that individuals with a similar preference tend to live together geographically. The workplaces of the consumer populations who live together also tend to cluster together.

[8] Most of gravity models assume that consumers shop from fixed points (e.g., their places of residence) and buy just one type of good or service per shopping trip (Carter, 1993).
[9] One exception is Nevin and Houston (1980, cited in Hardin and Wolverton, 2001) who include multipurpose shopping opportunities in their gravity model.

Therefore, they share shopping information easily and face similar shopping environments such as transportation network available to retailers. However, previous studies did not incorporate the unobservable shopping patterns in retail gravity models.

2.3 Retail Sales Data and Empirical Methodology

This section contains two parts. Part 1 describes the retail sales data and census data employed in this study. Part 2 presents the empirical methodology used to calibrate the retail gravity equation.

2.3.1 Retail Sales Data and Census Data

A retail consultant provided individual store and consumer data of a retail chain in the Houston market. The consumer data are for each household who shopped at a particular store. The data are the total dollar amounts each household spent at each individual store for the year 2000, and the block group where each household resides. We aggregate the data to the block-group level and calculate retail sales in a block group for each store ($Sales_{cs}$). The individual store data are total store sales in year 2000 ($storesales$) and in year 1999 ($lagged\ storesales$), store size in square feet ($storesize$), type of shopping center where each store resides ($strip$, pad, or $mall$), age of the shopping center ($centerage$), and longitude ($lons$) and latitude ($lats$) of each store.[10]

We supplement the retail sales data with 1990 census data and 1998 census estimates. In particular, we obtain census data relevant to the total potential expenditure for a block-group. The data are median medical supplies expenditure ($medicalsupp$), median household income ($medhsinc$), median house value ($medhsval$), median house age ($houseage$), total population ($totpop$), land area ($arealand$), median age ($medage$), white population ($popwhite$), and

[10] A strip (linear) shopping center consists of a line of stores with a pedestrian walk along the storefronts. A pad (cluster) center is a group of freestanding retail sites linked together by pedestrian walkways.

female population (*females*). Median house value, median house age, and land area are for 1990, while the other data are for 1998. We also obtain the longitude (*lon*) and latitude (*lat*) of each block-group to calculate retailer distance to consumers. Specifically, we compute the distance that consumers travel from their block group to the stores ($distance_{cs}$). We also obtain average travel time to work (*l_trvtim*) for 1990 from 1990 census data to take into account shopping originating from places of employment.

Geographically, the Houston market covers Baytown, Friendswood, Houston, Humble, Lake Jackson, Sugar Land, and Texas City. The retail chain has 14 stores in the Houston market. Twelve stores are located in shopping malls, one in a strip shopping center, and one in a pad-type shopping center.

Table 2.1 presents the descriptive statistics of the data used in this study. On average a store had $1,846 of retail sales in a block group. Retail sales in a block group vary widely from only $5 to $113,323. This variation indicates retail sales do not uniformly originate over space. Figures 2.1-1 to 2.1-4 map the spatial distribution of retail sales for one of the 14 stores. Clearly, retail sales cluster together. This clustering indicates a potential spatial dependence among consumers. The distance among retailers and consumers is measured using the Euclidean matrix. Over fifty percent of the consumers traveled less than 10.49 miles to the store where they shopped from their residence. Some consumers lived only 0.15 mile away from the store they shop. Some consumers lived over 712 miles away in Texas.[11]

For the 14 stores in our study, total store sales were stable over year 1999 and year 2000. A store on average generated about the $1.3 million total sales a year. The stores generated sales between about $0.5 million and $2 million. Figure 2.2 maps total store sales for the 14 stores

[11] The longest straight-line distance of Texas is 801 miles in a general north-south direction and 773 miles in a general east-west direction (*Texas Almanac*, 2001).

Table 2.1: Descriptive Statistics

Variables	Label	Mean	Std Dev.	Median	Minimum	Maximum
Retail sales ($)	$sales_{bc}$	1,846.06	3,827.73	524.00	5.00	113,323.00
Distance (miles)	$distance_{bc}$	22.29	45.99	10.49	0.15	712.26
Store sales ($)	$storesales$	1,339,666.71	380,706.15	1,315,324.50	505,601.00	2,030,448.00
Lagged store sales ($)	$lagged\ storesales$	1,324,174.57	383,531.07	1,268,676.50	515,538.00	1,853,128.00
Shopping center age (years)	$centerage$	10.36	3.12	10.88	4.87	15.75
Store size (square feet)	$storesize$	4,578.79	929.01	4,511.00	3,157.00	6,612.00
Medical supplies expenditure ($1000)	$medicalsupp$	61.67	42.60	51.00	3.00	782.00
Median household income ($)	$medhsinc$	45,706.62	24,808.02	39,328.00	11,976.00	150,000.00
Median house value ($)	$medhsval$	71,301.38	57,253.59	57,100.00	14,999.00	500,001.00
Median house age (years)	$houseage$	31.98	11.94	29.00	10.00	61.00
Travel time to work (minutes)	l_trvtim	24.53	6.28	24.40	6.90	54.10
Population (persons)	$totpop$	1,837.00	1,007.17	1,627.00	70.00	11,125.00
Land area (0.001 square kilometers)	$arealand$	10,428.72	33,201.48	1,125.00	60.00	446,589.00
Median age (year)	$medage$	33.57	6.50	32.70	11.20	75.00
White population (persons)	$popwhite$	1,407.99	876.56	1,282.00	5.00	10,135.00
Female population (persons)	$females$	927.87	507.12	826.00	43.00	6,019.00

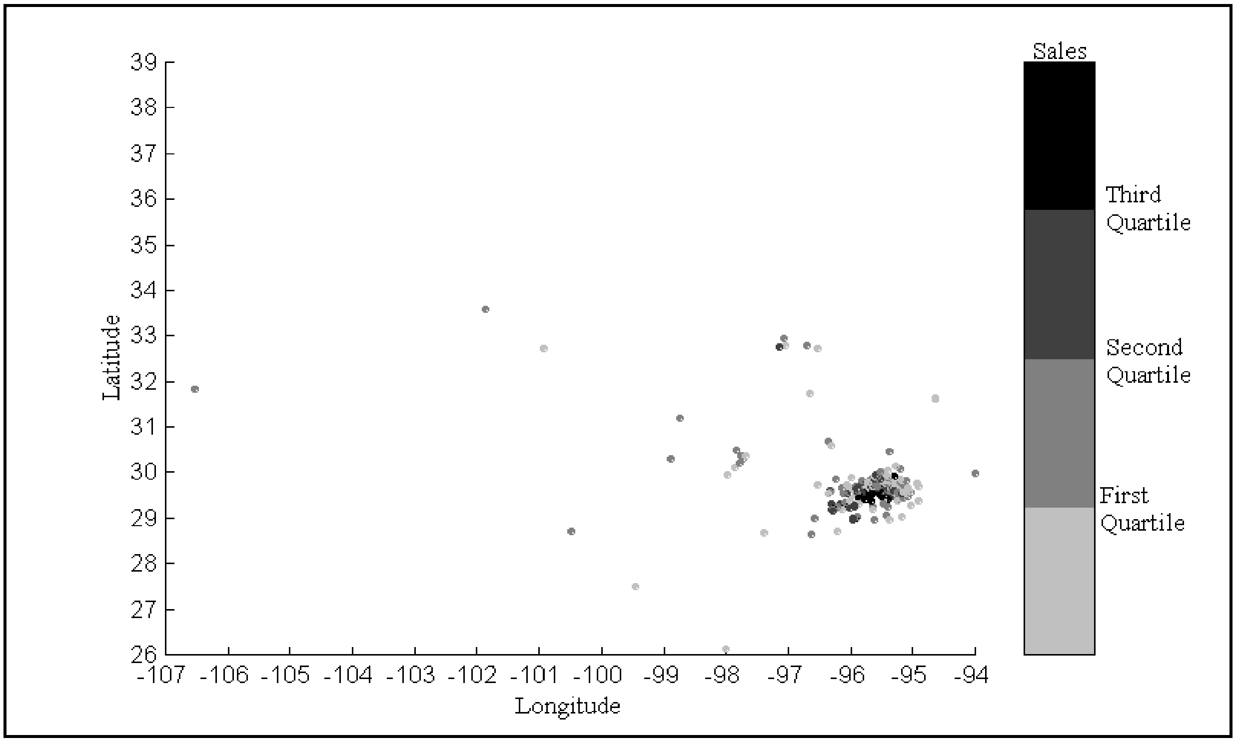

Figure 2.1-1 Spatial Distribution of Retail Sales

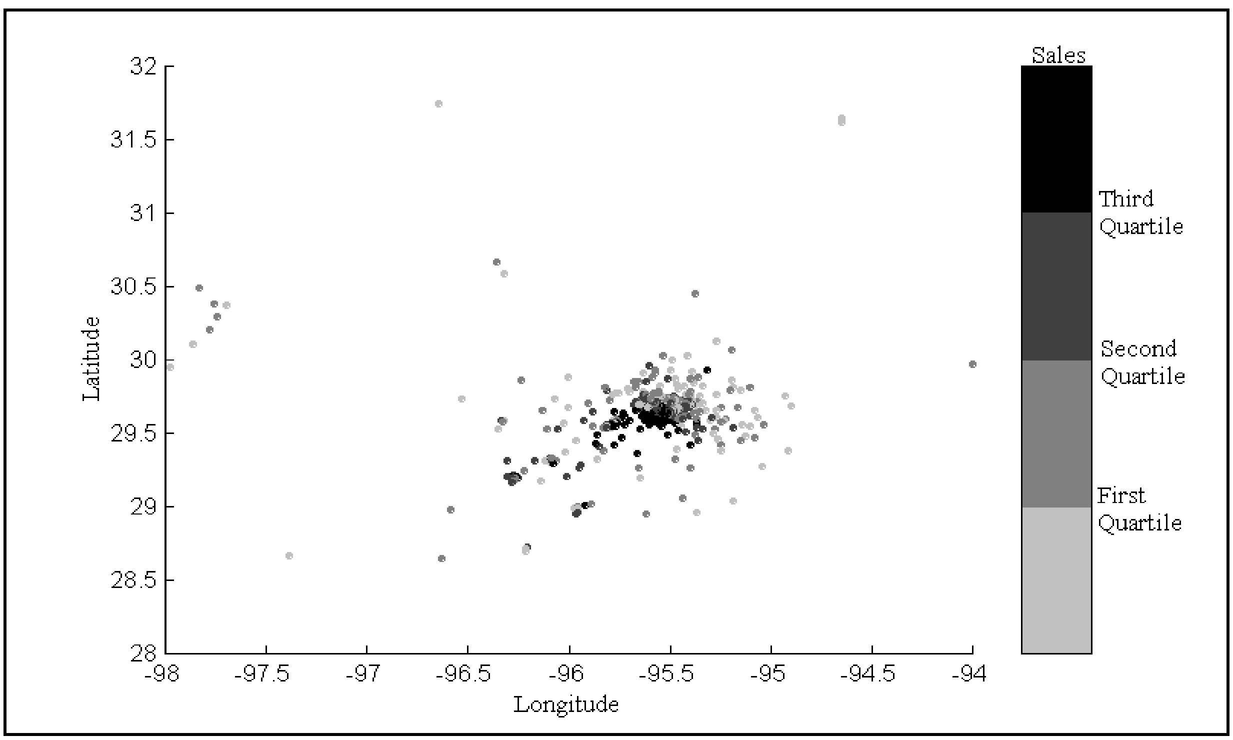

Figure 2.1-2 Spatial Distribution of Retail Sales

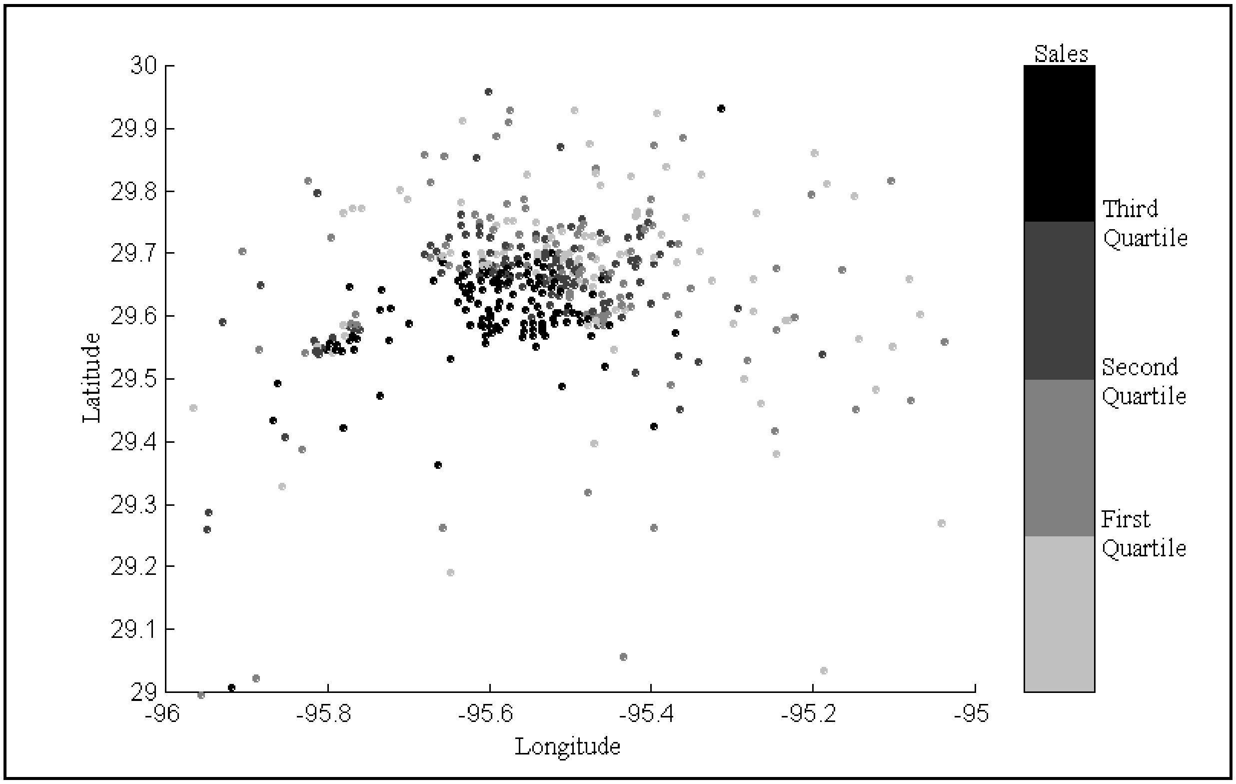

Figure 2.1-3 Spatial Distribution of Retail Sales

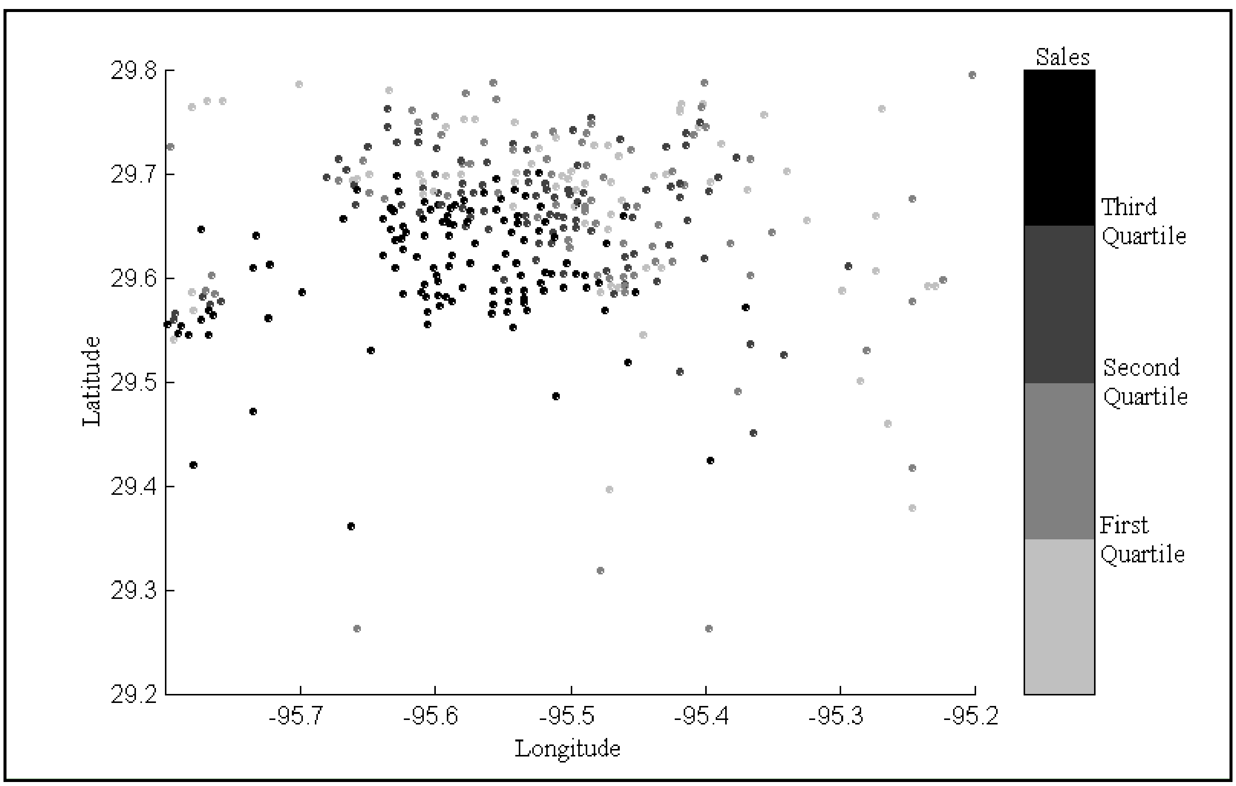

Figure 2.1-4 Spatial Distribution of Retail Sales

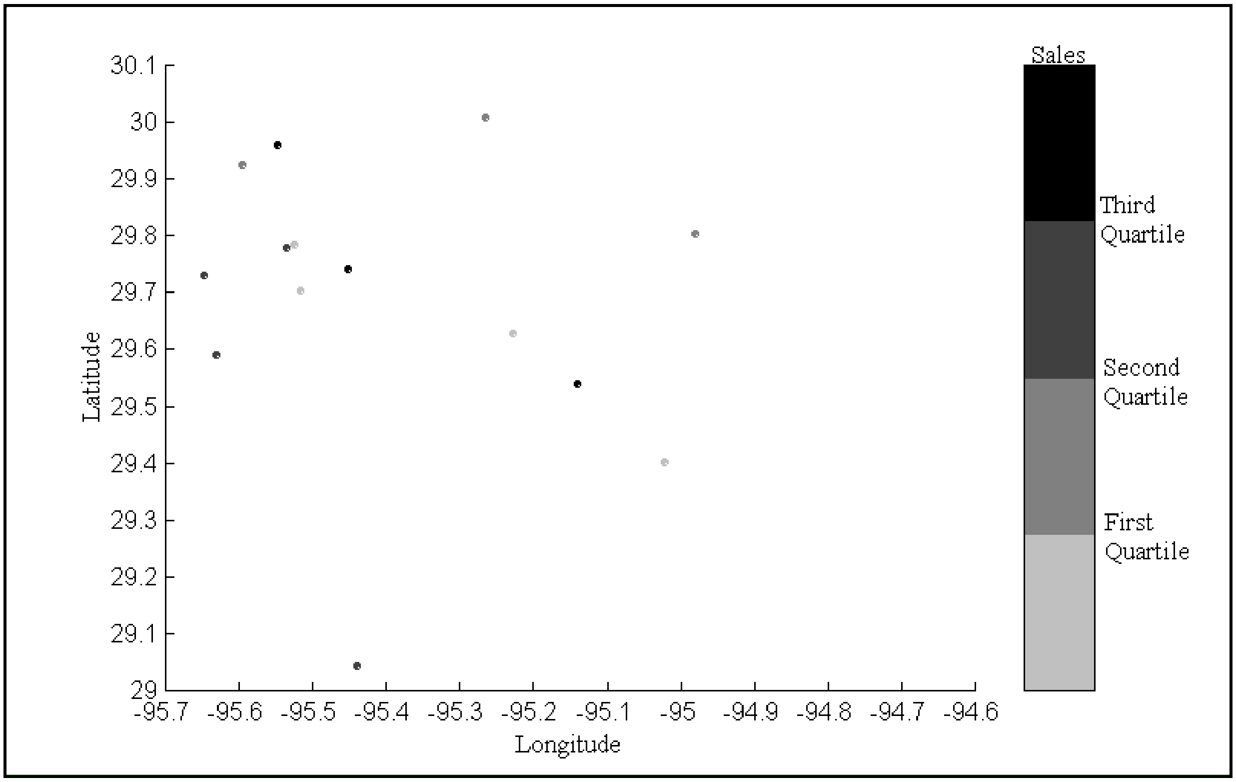

Figure 2.2 Spatial Distribution of Store Sales

over space. Store sales clearly spatially cluster together. This clustering indicates a potential spatial dependence among store locations. The average store size was 4,579 square feet. The smallest store had 3,157 square feet, and the largest store had 6,612 square feet. The average shopping center age was 10 years. The newest center opened 5 years ago, and the oldest center opened 15 years ago.

There are 2,977 block-groups in Texas whose residents shopped at the stores in the study during 2000. Residents in a block group spent on average $62,000 on medical supplies in 1998. The median household income was $33,688 a year. The median house value was $71,000. Residents on average spent 25 minutes traveling from their home to work. On average a block group has 1,387 residents and 10 square kilometers in area. The median age of a resident was of 34 years old. A block group on average has 993 white residents and 699 female residents.

2.3.2 Empirical Methodology

This overall section presents the empirical methodology. Part 1 briefly describes a SAR in errors model. Part 2 provides the detail on the construction of the spatial weight matrix for this study. Part 3 discusses the maximum likelihood computations. Part 4 shows the empirical model for the retail gravity model.

2.3.2-1 A SAR in Errors Model

We describe a SAR in errors model, mainly following the notation of Pace and Gilley (1997). When errors exhibit spatial autocorrelation, a SAR in errors model uses the difference between the dependent variable and the model prediction from nearby observations to correct the usual prediction of the dependent variable:

$$Y = X\beta + \alpha D(Y - X\beta) + \varepsilon \qquad (2)$$

where D is an $n \times n$ spatial weighting matrix and ε is an $n \times 1$ vector of error terms. To prevent an observation from predicting itself, D has zeros on its diagonal. To facilitate interpretation, each row of D sums to 1. To ensure the stability of the entire error process, the spatial autocorrelation parameter, $\alpha,$ is restricted to lie within the interval $[0,1)$. The errors, $\varepsilon,$ are independently and normally distributed. These assumptions are summarized in the following:

$$\begin{aligned}
&(a) \quad \underset{(n\times n)}{D}\underset{(n\times n)}{[1]} = \underset{(n\times n)}{[1]} \\
&(b) \quad diag(D) = \underset{(n\times n)}{[0]} \\
&(c) \quad 0 \le \alpha < 1 \\
&(d) \quad \varepsilon \sim N(0, \sigma^2 I)
\end{aligned} \tag{3}$$

In the SAR in error model, a $\alpha > 0$ indicates a positive spatial dependence. This implies that errors of same sign are geographically clustered together. On the other hand, a $\alpha < 0$ implies a negative spatial dependence, and this implies that the errors of the opposite sign are clustered together geographically. When $\alpha = 0$, the SAR in errors model reduces to an OLS model. In a retail context, similar consumer populations and shopping environments should exhibit a positive spatial dependence among consumers. Individual stores should share similar retailing environments with other stores in the same retail chain located nearby. This sharing should present a positive spatial dependence among stores in a retail chain. Therefore, for convenience, we use the restriction $0 \le \alpha < 1$.[12]

2.3.2-2 Specification of the Spatial Weight Matrix

To model the spatial dependence among stores and consumers, we specify a spatial weighting matrix $D = wC + (1-w)S,$ where C and S are weighting matrices for consumers and

[12] This assumption is for convenience. If $\alpha < 0,$ the estimates should lie on the boundary $\alpha = 0.$ We did not observe such a boundary solution.

stores respectively, and $0 \le w \le 1$. When $w = 1$, D reduces to C. When $w = 0$, D reduces to S. Empirically, we search the optimal w over [0.00, 0.01,…1.00].

We form C using the approach of nearest neighbors with geometrically decaying weights. Under this scheme, the weight given to the block group depends on their proximity for each block group relative to all other block groups and the rate of geometric decay.

Let d_{ij} represent the distance between the block groups of every pair of observations i and j, and let d_i^{mth} stand for the distance from the block group of observation i to its mth nearest block group. Let $w_{ij}^c = 1$ if $0 < d_{ij} \le d_i^{mth}$ and $w_{ij}^c = 0$ otherwise. Let ρ represent 1 minus the rate of geometric decay of weights such that the hth closest neighbor of the block group of observation i has a weight of ρ^h where $0 \le \rho \le 1$. Then in a row-stochastic C,

$$C_{ij} = \rho^h w_{ij}^c \Big/ \sum_{\substack{j=1 \\ i \ne j}}^{n} \rho^h w_{ij}^c \text{ and } C_{ii} = 0.$$

In our computation, we construct C by searching optimal m for 37 values over [0,1, 2…36] and optimal ρ for 101 values over [0.00, 0.01…1.00].

We form S with the Delaunay triangulation approach. The Delaunay triangulation is the geometric dual of the Voronoi diagram that is geometric expression of connections among contiguous stores (Calciu and Salerno, 1997).[13] Each store at a Delaunay triangle serves similar consumer populations who tend to live together. With this approach, $w_{ij}^s = 1$ when observations i and j dot not belong to the same store and are at the same Delaunay triangle; $w_{ij}^s = 0$ otherwise. To have a row-stochastic S, we standardize each row so that

[13] The Voronoi diagram has the property that for each store every point in the region around that store is closer to that site than to any of the other stores.

$$S_{ij} = w_{ij}^s \Big/ \sum_{\substack{j=1 \\ i \neq j}}^{n} w_{ij}^s \text{ and } S_{ii} = 0.$$

Because S is not sparse enough, a computer with 500 MB of RAM cannot do other computations after constructing S physically.

To overcome the computer memory requirement, we recognize S before being standardized is a multiplication of smaller matrixes. Specifically $U = AGA'$ when observations are sorted by store, where U denote unstandardized S, G is a $n_s \times n_s$ weight matrix of Delaunay triangulation for n_s stores, A is a 0, 1 matrix whose dimension is $n \times n_s$. $A_{ij} = 1$ when observation i belongs to store j, and $A_{ij} = 0$ otherwise. Let J be a $n \times 1$ column vector whose ith element equals the sum of U over columns for its ith row. Also let H stands for the $n \times n_s$ matrix whose ith row equals the ith row of AG divided by the ith element of J. Then $S = HA'$. With this relationship, we can perform operations with S without physically constructing the weighting matrix for stores and thus reduce required computer memory. As a result, in our actual computation, $D = wC + (1-w)HA'$ instead of $D = wC + (1-w)S$.

Here is a numerical example showing how to express S with H and A. Assume we have following matrixes for 4 retail sale observations for 3 stores:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } G = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}.$$

In this example, $H = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 0 \end{bmatrix}$. Then S can be expressed as:

$$S = HA' = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \end{bmatrix}.$$

2.3.2-3 Maximum Likelihood Computations

The SAR in error model in (2) and (3) has the following log-likelihood function,

$$L(\alpha,\beta,\sigma^2) = \frac{1}{2}\ln|B| - \frac{1}{2}\left[n\ln(2\pi\sigma^2) + \sigma^{-2}(Y - X\beta)'B(Y - X\beta)\right] \tag{4}$$

where B equals $(I-\alpha D)'(I-\alpha D)$. After being rearranged, the profile log-likelihood can be written as

$$L(\alpha,\beta,\sigma^2) \propto \ln|I-\alpha D| - \frac{1}{2}\ln\left(SSE\left(\alpha\right)\right). \tag{5}$$

where $SSE = (Y - X\beta)'B(Y - X\beta)$ (Pace, Barry, Slawson, and Sirmans, 2002). We maximize the log-likelihood by computing Equation (5) for 100 values of α over $\left[0.00, 0.01, ...0.99\right]$.

To overcome the computer memory requirement, we use Barry and Pace's (2002) Matlab function of Monte Carlo Log-determinant Estimator to compute estimates of $\ln|I-\alpha D|$, following Barry and Pace (1999). In particular, $-\ln|I-\alpha D|$ can be expanded in a power series as follows:

$$-\ln|I-\alpha D| = \sum_{r=1}^{q}\frac{\text{tr}\left(D^r\right)\alpha^r}{r} + \sum_{r=q+1}^{\infty}\frac{\text{tr}\left(D^r\right)\alpha^r}{r}.$$

Using the above expansion together with $\mathrm{E}\dfrac{\mathbf{x}'Z\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \dfrac{tr\left(Z\right)}{n}$ for any real $n\times n$ matrix Z and $\mathbf{x} \sim N_n\left(0, I\right)$, Barry and Pace (1999) show that $\ln|I-\alpha D|$ can be approximated by $\bar{V}$, the expected value of V_i:

$$V_i = -n\sum_{r=1}^{q}\frac{\mathbf{x}_i' D^r \mathbf{x}_i}{\mathbf{x}_i'\mathbf{x}_i}\frac{\alpha^r}{r}, \ i = 1, 2, \ldots p,$$

where $\mathbf{x}_i \sim N_n\left(0, I\right)$, $\mathbf{x}_i$ independent of $\mathbf{x}_j$ if $i \neq j$. Recognizing $n \geq \left|\mathrm{tr}\left(D^r\right)\right|$, they also show the interval $\left(\bar{V} - F, \bar{V} + F\right)$ as the asymptotic 95% confidence interval for $\ln\left|I - \alpha D\right|$, where $F = \dfrac{n\alpha^{q+1}}{(q+1)(1-\alpha)} + 1.96\sqrt{\dfrac{s^2(V_1, V_2, \ldots V_p)}{p}}$. Empirically, we set $p = 30$ and $q = 98$ in computing $\bar{V}$ and use the lower bound of the 95% confidence interval for $\ln\left|I - \alpha D\right|$ as conservative estimates in maximizing the log-likelihood specified in Equation (5).

2.3.2-4 The Empirical Model

We rewrite the retail gravity models, Equation (1), in log form and empirically model retail sales as:

$$\ln(Sales_{bc}) = X\beta + \upsilon_{bc} \qquad (6)$$

where

$X = \left[X_{bc} \vdots X_b \vdots X_c\right]$,

$X_{bc} = \left[1, \ \ln(distance_{bc})\right]$,

X_b = a vector of variables pertaining to store b,

X_c = a vector of variables pertaining to shoppers' area c,

and υ_{bc} = a error term.

X_b contains 5 variables of store characteristics. The 5 variables are $\ln(lagged\ storesales)$, $\ln(centerage)$, $\ln(storesize)$ as well as two store-type dummies, $strip$ and pad. X_c contains 9 variables in log form relevant to total potential expenditure and average

travel time to work in log form for block group c. Specifically, the 10 variables are $\ln(medicalsupp)$, $\ln(medhsinc)$, $\ln(medhsval)$, $\ln(houseage)$, $\ln(totpop)$, $\ln(arealand)$, $\ln(medage)$, $\ln(popwhite)$, $\ln(females)$, and $\ln(l\_trvtim)$.

To model the spatial dependence among stores and consumers, we fit Equation (6) with a SAR in error model using Pace and Barry's (2002) spatial statistics toolbox 1.1. Specifically we assume $\upsilon_{bc} \sim N(0,[(1-\alpha D)'(1-\alpha D)]^{-1})$.

The hypothesized signs for the variables in Equation (6) are summarized below.

(a) Positive signs (+): $\ln(lagged\ storesales)$, $\ln(storesize)$, $\ln(medicalsupp)$, $\ln(medhsinc)$, $\ln(medhsval)$, $\ln(totpop)$, $\ln(arealand)$, $\ln(medage)$, $\ln(popwhite)$, and $\ln(females)$.

(b) Negative signs (-): $\ln(distance_{bc})$, $\ln(centerage)$, $\ln(houseage)$, $\ln(l\_trvtim)$, $strip$, and pad.

2.4 Empirical Results

To understand the importance of the spatial dependencies, we calibrate four models that consider different components of the spatial dependencies among consumers and stores. The first model ignores the spatial dependence. The second model considers the spatial dependence among stores. The third model considers the spatial dependence among consumers. The fourth model considers both the spatial dependencies among stores and consumers. Table 2.2 presents the calibration results of Equation (6).

The first model calibrates the gravity model with OLS. As hypothesized, the distance variable has a significant coefficient of –0.818, with signed root deviance (SRD) –57.346. This coefficient is the constant distance elasticity of retail sales, which measures the percentage change in retail sales in respect of one percentage change in distance. In particular, the

Table 2.2: Retail Gravity Models

	Model 1		Model 2	
Independent variables	β_{OLS}	SRD	β_{SAR}	SRD
$\ln(distance_{bc})$	-0.818	-57.346***	-0.815	-57.204***
$\ln(lagged\ storesales)$	-0.007	-0.154	0.054	0.761
$\ln(centerage)$	0.565	9.638***	0.514	7.529***
$\ln(storesize)$	-1.059	-9.378***	-0.888	-5.355***
$strip$	-0.310	-8.947***	-0.305	-8.888***
pad	0.006	0.126	-0.016	-0.258
$\ln(medicalsupp)$	-0.030	-0.495	-0.029	-0.470
$\ln(medhsinc)$	-0.113	-3.123***	-0.109	-3.015***
$\ln(medhsval)$	0.144	3.581***	0.153	3.744***
$\ln(houseage)$	-0.358	-7.201***	-0.346	-6.926***
$\ln(totpop)$	-0.360	-2.586***	-0.252	-1.685*
$\ln(arealand)$	0.189	17.861***	0.186	17.476***
$\ln(medage)$	0.236	2.425**	0.229	2.349**
$\ln(popwhite)$	0.070	4.330***	0.064	3.902***
$\ln(females)$	0.509	3.910***	0.408	2.868***
$\ln(l\_trvtim)$	-0.683	-11.122***	-0.683	-11.117***
Intercept	12.700	8.802***	13.683	6.016***
w			0	
α			0.040	1.337
m			34	
ρ			1	
p			30	
q			98	
n	7,983		7,983	
k	17		20	
Log likelihood	-36,165.975		-36,165.081	
R^2	0.386		0.386	
$adjR^2$	0.385		0.384	
SSE	8,610.588		8,608.654	
Median $\|e\|$	0.712		0.715	

Note:

1. w is the weight on the spatial weight matrix for consumers.
2. α is the spatial autoregression parameter.
3. m is the number of nearest blocks used in constructing the spatial weight matrix for consumers.
4. ρ represents 1 minus the rate of geometric decay of weights on neighboring blocks in the spatial weight matrix for consumers.
5. p is the number of trials performed to obtain the Monte Carlo estimates for the log-determinant term in the profile log-likelihood.
6. q is the highest order of power series expansion for the log-determinant term in the profile log-likelihood when obtaining the Monte Carlo estimates.
7. SRD stands for signed root deviance that equals the square root of likelihood statistics with a sign of its coefficient.
8. * significant at 10% level; ** significant at 5% level; *** significant at 1% level

(Table 2.2 continued)

	Model 3		Model 4	
Independent variables	β_{SAR}	SRD	β_{SAR}	SRD
$\ln(distance_{bc})$	-1.372	-68.304***	-1.377	-69.721***
$\ln(lagged\ storesales)$	0.190	4.644***	0.300	7.012***
$\ln(centerage)$	0.262	4.958***	-0.079	-1.394
$\ln(storesize)$	-0.557	-5.290***	0.490	4.021***
$strip$	-0.251	-8.424***	-0.148	-5.065***
pad	0.096	2.347**	-0.112	-2.736***
$\ln(medicalsupp)$	0.431	6.049***	0.386	5.537***
$\ln(medhsinc)$	-0.038	-0.895	-0.001	-0.024
$\ln(medhsval)$	0.186	4.101***	0.198	4.398***
$\ln(houseage)$	-0.019	-0.292	-0.051	-0.839
$\ln(totpop)$	-0.170	-1.184	0.003	0.021
$\ln(arealand)$	0.031	2.259**	0.049	3.390***
$\ln(medage)$	0.010	0.098	0.034	0.352
$\ln(popwhite)$	0.078	3.282***	0.064	2.755***
$\ln(females)$	0.163	1.228	0.044	0.333
$\ln(l\_trvtim)$	-0.104	-1.332	-0.077	-0.997
Intercept	7.782	5.296***	2.518	1.573
w	1		0.840	
α	0.870	42.409***	0.990	44.777***
m	34		34	
ρ	1		1	
p	30		30	
q	98		98	
n	7,983		7,983	
k	20		21	
Log likelihood	-35,264.586		-35,163.504	
R^2	0.540		0.548	
$adjR^2$	0.539		0.547	
SSE	6,448.063		6,331.684	
Median $\lvert e \rvert$	0.607		0.596	

Note:

1. w is the weight on the spatial weight matrix for consumers.
2. α is the spatial autoregression parameter.
3. m is the number of nearest blocks used in constructing the spatial weight matrix for consumers.
4. ρ represents 1 minus the rate of geometric decay of weights on neighboring blocks in the spatial weight matrix for consumers.
5. p is the number of trials performed to obtain the Monte Carlo estimates for the log-determinant term in the profile log-likelihood.
6. q is the highest order of power series expansion for the log-determinant term in the profile log-likelihood when obtaining the Monte Carlo estimates.
7. SRD stands for signed root deviance that equals the square root of likelihood statistics with a sign of its coefficient.
8. * significant at 10% level; ** significant at 5% level; *** significant at 1% level

coefficient here predicts a 0.818% decrease in a store's sales in a block group when the distance between the store and the block group increases 1%.

The dummy for a strip shopping center has a significant and negative coefficient, as hypothesized. This coefficient is consistent with Sirmans and Guidry (1993) and Oppewal and Timmermans (1999, cited in Mejia and Benjamin, 2002). A mall has more aesthetically appealing design and usually provides more protection from stormy weather to shoppers than other types of shopping centers do (Sirmans and Guildry, 1993). Oppewal and Timmermans (1999, cited in Mejia and Benjamin, 2002) find that design influences consumer perception of shopping centers and, thus affects retail sales. However, several other variables have significant coefficients with signs opposite to those hypothesized by retail gravity. For variables pertaining to consumers, median household income and total population of a block group have significant and negative coefficients. Their signs are inconsistent with our hypotheses. For variables pertaining to retail stores, store size and shopping center age have significant coefficients with signs inconsistent with our hypotheses. Shopping center age has a positive coefficient while hypothesized to have a negative coefficient. In addition, store size has a negative coefficient while hypothesized to have a positive coefficient.

The second model models only the spatial dependence among stores. The distance variable decreases slightly the magnitude of its coefficient by only 0.37% in absolute value. The coefficient changes from –0.818 to –0.8165and remains significant with SRD –57.204. Although lagged store sales changes its coefficient from negative to hypothesized positive, the coefficient is insignificant. The dummy for a pad shopping center changes to have a hypothesized negative coefficient but remains insignificant. Store size still has a significant and negative coefficient. However, its SRD decreases near 43% in absolute value from –9.378 to –5.355. The spatial

dependence among stores does not change either the signs or the significances of the coefficients of variables pertaining to consumers. Nevertheless, the singed root deviances for median household income and total population of a block group decrease 3.46% and 34.84% in absolute value respectively.

The third model calibrates only the spatial dependence among consumers into the gravity model. The distance variable dramatically increases the magnitude of its coefficient by 68.34% in absolute value to –1.372 compared to –0.815 in the second model. The SRD for the distance variable also considerably increases more than 19% in absolute value from –57.204 to –68.304. In addition, this spatial dependence has a great influence on variables pertaining to consumers. The coefficients of median household income and total population of a block group become insignificant, but still have signs opposite to our hypotheses. As a result, now there are no consumer variables having significant coefficients with signs opposite to our hypotheses. Median house age, median age of consumers, female population, and average travel time to work change to have insignificant coefficients. Nevertheless, median medical supplies expenditure changes to have a positive and significant coefficient as hypothesized.

The spatial dependence among consumers also has influences on store variables. Lagged store sales, a proxy for store management, changes its coefficient from insignificant to significant positive, as hypothesized. This coefficient is consistent with Black's (1966; cited in Eilon, Tilley and Fowkes, 1969) argument. More sales enable a store to carry a greater variety of products, improve its services to customers, and compete with other stores. In addition, lagged store sales may pick up the effect of store age on better management or other unobservables. Start-up problems may adversely affect sales for stores in new locations (Hise, Kelly, Gable, and McDonald, 1983). Stores in business longer should have overcome the start-up problems and

have larger store sales. All these effects can increase the attractive power of a store. Store size still has a significant and negative coefficient. Nonetheless, its magnitude decreases 37.27% in absolute value from –0.888 to –0.557, and its SRD decreases 1.21% in absolute value from –5.355 to –5.290. However, the dummy for a pad shopping center changes to have a significant and positive coefficient. This coefficient is not consistent with Sirmans and Guidry (1993) and Oppewal and Timmermans (1999, cited in Mejia and Benjamin, 2002).

The fourth model models both the spatial dependencies among stores and consumers. Compared to the third model, the coefficient for distance slightly increases another 0.36% in its absolute magnitude from –1.372 to –1.377. The SRD for the distance variable increases another 2.07% in absolute value from –68.304 to –69.721. The other most important variable, store size, in the retail gravity model now have a significant coefficient with hypothesized sign. Its coefficient changes to significant and positive in Model 4 from significant and negative in Models 1, 2, and 3. The coefficient changes from –0.557 in Model 3 to 0.490 in Model 4. The change is a near 188% increase in magnitude. The SRD for store size changes from –5.290 to 4.021. The dummy for a pad shopping center changes to have a significant and negative coefficient, as hypothesized. In addition, center age changes to have a negative coefficient. Although the coefficient is not significant, it has a hypothesized sign. This sign agrees with Sirmans and Guidry's (1993) and Gatzlaff, Sirmans, and Diskin's (1994) arguments. They argue that older shopping centers generally suffer functional or physical deficiencies and have an inappropriate tenant mix due to changing markets, and thus have less attractive power. The coefficient for lagged store sales increases about 58% in magnitude from 0.190 to 0.300. Its SRD also increase considerably from 4.644 to 7.012. This change constitutes near 51% in magnitude. The coefficients of total population of a block group become positive, as hypothesized, but

remain insignificant. Median household income still has an insignificant and negative coefficient. Nevertheless, the coefficient decreases more than 97% in absolute value from –0.038 to -0.001. The SRD for the median household income also decreases more than 97% in absolute value from –0.895 to –0.024.

Besides the coefficients, model selection criteria also show models perform better when incorporating both the spatial dependencies than otherwise. Adjusted R-squared increases from 0.385 for the first model to 0.547 for the fourth model. The fourth model displays considerably lower error than the first model. In fact the SSE from the first model of 8,610.588 is 35.99% higher than the SSE from the fourth model of 6,331.684. Median absolute errors decrease 16.29% from 0.712 for the first model to 0.596 for the fourth model. The log-likelihoods also show the fourth model outperforms the other three models. In fact, the likelihood ratio statistics between Model 1 and Model 4 is about 2,005 that is very highly significant.

2.5 Conclusions

Gravity-type models have been applied to the retail context extensively. Guitschi (1981) and Eppli and Shilling (1996) suggested that the distance parameter for retail gravity models may be significantly overstated in the existing studies. Using actual sales for a retail chain in the Houston market and modeling spatial dependence among customers and stores, we find that previous retail gravity research may have understated the magnitude of the distance parameter by as much as 68%. This result implies previous studies may overestimate the deterministic extent of trade areas reflected in the distance parameter and understate the importance of good locations.

Unlike previous studies, we incorporate both the spatial dependencies among stores and consumers in a retail gravity model, using a SAR in error model. Our results show ignoring the

spatial dependence among consumers underestimates the distance parameter more than 68%. This percentage indicates the spatial dependence among consumers is far more important than the spatial dependence among stores in terms of estimating the parameter of the retailer distance to consumers.

Besides the spatial dependence among consumers, we also incorporate the spatial dependence among stores. Our results show the spatial dependence among store needs to be incorporated into empirical retail gravity studies, along with the spatial dependence among consumers. The results are consistent with the retail literature documenting the importance of comparison-shopping and multi-purpose shopping behaviors of consumers. Our result also implies existing studies have failed to recognize the random component of trade areas reflected on the spatial autocorrelation parameter and the spatial weight matrix in a SAR in errors model.

2.6 References

Barry, Ronald Paul, and R. Kelley Pace. (1999). "Monte Carlo Estimates of the Log Determinant of Large Sparse Matrices," *Linear Algebra and Applications* 289, 41-54.

Barry, Ronald Paul, and R. Kelley Pace. (2002) Matlab function of Monte Carlo Log-determinant Estimator. Available: http://www.spatialstatistics.com.

Brown, S. (1992). "The Wheel of Retail Gravitation?" *Environment and Planning A* 24, 1409-1429.

Calciu M., and F. Salerno. (1997). "A New Approach to Spatial Management of Retail Networks, based on German School's Central Place Theory. Application to Bank Location," dans H.Mühlbacher /J-P Flipo (Eds.), *Advances in Services Marketing.* Gabler, Wiesbaden.

Carter, Charles C. (1993). "Assumptions underlying the retail gravity model," *The Appraisal Journal* 4, 509-518.

Curry, Leslie. (1972). "A Spatial Analysis of Gravity Flows," *Regional Studies* 6, 131-147.

Curry, Leslie, Daniel A. Griffith, and Eric S. Sheppard. (1975). "Those Gravity Parameters Again," *Regional Studies* 9, 289-296.

Curry, Leslie, Daniel A. Griffith, and Eric S. Sheppard. (1976). "A Final Comment on Mis-Specification and Autocorrelation in Those Gravity Parameters," *Regional Studies* 10, 337-339.

Eilon, S., R.P.R. Tilley, and T.R. Fowkes. (1969). "Analysis of a Gravity Demand Model," *Regional Studies* 3, 115-122.

Eppli, Mark J., and John D. Benjamin. (1994). "The Evolution of Shopping Center Research: A Review and Analysis," *The Journal of Real Estate Research* 9, 5-32.

Eppli, Mark J., and James D. Shilling. (1996). "How Critical Is a Good Location to a Regional Shopping Center?" *The Journal of Real Estate Research* 12, 459-468.

Fotheringham, A. Stewart and M. J. Webber. (1980). "Spatial Structure and the Parameters of Spatial Interaction Models," *Geographic Analysis* 12, 33-46.

Gatzlaff, D. H., G. S. Sirmans, and B. A. Diskin. (1994). "The Effect of Anchor Tenant Loss on Shopping Center Rents," *Journal of Real Estate Research* 9, 99-110.

Gautschi, David A. (1981). "Specification of Patronage Models for Retail Center Choice," *Journal of Marketing Research* 18, 162-174.

Hardin, William G. III and Marvin L. Wolverton. (2001). "Neighborhood Center Image and Rents," *Journal of Real Estate Finance and Economics* 23, 31-46.

Hise, Richard T., J. Patrick Kelly, Myron Gable, and James B. McDonald. (1983). "Factors Affecting the Performance of Individual Chain Store Units: an Empirical Analysis," *Journal of Retailing* 59, 22-39.

Mejia, Luis C. and John D. Benjamin. (2002). "What Do We Know About the Determinants of Shopping Center Sales? Spatial vs. Non-Spatial Factors," *Journal of Real Estate Literature* 10, 3-26.

Ord, J. K. (1975). "Estimation Methods for Models of Spatial Interaction," *Journal f the American Statistical Association* 70, 120-126.

Pace, R. Kelley, Ronald Barry, V. Carlos Slawson, Jr., and C. F. Sirmans. (2002). "Simultaneous Spatial and Functional Form Transformations," in Luc Anselin and Raymond Florax (Ed.), *Advances in Spatial Econometrics,* forthcoming, Springer-Verlag, Heidelberg. Available: http://www.spatialstatistics.com.

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

Pace, R. Kelley and Ronald Barry. (2002). Spatial Statistics Toolbox 1.1. Available: http://www.spatial-statistics.com.

Porojan, A. (2001). "Trade Flows and Spatial Effects: The Gravity Model Revisited," *Open Economic Review* 12, 265-280.

Sirmans, C. F. and Krisandra A. Guidry. (1993). "The Determinants of Shopping Center Rents," *Journal of Real Estate Research* 8, 107-115.

Stanley, Thomas J. and Murphy A. Sewall. (1976). "Image Inputs to a Probabilistic model: Predicting Retail Potential," *Journal of Marketing* 40, 48-53.

Texas Almanac. (2001). 2002-2003 edition, The Dallas Morning News. Available: http://www.texasalmanac.com/texasenviro_2000.htm

CHAPTER 3

LOCAL HOUSING PRICES AND MORTGAGE REFINANCING IN U.S. CITIES

3.1 Introduction

The U.S. has experienced three booms of refinancing activity in the past fifteen years.[1] Such refinancing activity affects mortgage bankers, investors, and regulators. Clearly, these refinancing waves challenge mortgage bankers, since refinancing constitutes a substantial portion of their business. Refinancing booms influence their decisions in allocating their human and technological assets. Mortgage servicers lose business once a mortgage they currently serve is refinanced, and hence unexpected refinancing reduces profits. Naturally, this refinancing risk affects the lending institution that sells and holds mortgages. And as a result, a regulator who desires to develop measures of capital adequacy among these institutions should also consider the effects of refinancing upon regulated firms. Mortgage investors are also concerned with refinancing. Refinancing terminates mortgages, or the mortgages underlying mortgage-backed securities (MBS). Thus, proper valuation of mortgages as well as MBS hinges on the ability to model refinancing and other types of mortgage termination risk.

Recent option-based mortgage and MBS valuation models have recognized that house prices are another important determinant of mortgage termination in addition to interest rates (e.g., Hillard, Kau, and Slawson (1998)). Most existing models, however, consider the direct impact of house price change on default behavior (Downing Stanton and Wallce, 2001; DSW hereafter).[2] In other words, a derease in house prices first increases the likelihood of default. The likelihood of

---

[1] The three refinancing booms occured in 1986-87, 1992-93, and 1998-1999.

[2] One exception is Mattey and Wallace (2001) who provide evidence that house price declines directly impact mortgage termination through collateral constraints.

default in turn decreases the possibility of prepayment. Nevertheless the two factors result in an increase in total terminations (DSW, 2001). DSW (2001) argue that the effect of a drop in house prices could lead to a significant decrease in total terminations, if house prices have a first order effect not only on default but also on refinancing or mobility related mortgage prepayment. Since refinancing is the dominant component of prepayments, we examine the influence of house price change on mortgage refinancing.[3]

Many studies model refinancing risk separately or together with other termination risks. Some of these studies have examined the relation between housing prices and prepayment decisions of individual borrowers. Examples of such studies are Cunningham and Capone (1990), Dickinson and Heuson (1993), Peristiani, Bennett, Monsen, Peach and Raiff (1997 a and b; PBMPR hereafter), Caplin, Freeman, and Tracy (1997), as well as Bennett, Peach, and Peristiani (2001; BPP hereafter). Except Dickinson and Heuson (1993), the studies provide evidence only on the collateral constraint brought by house price changes. On the other hand, Dickinson and Heuson (1993) have evidence that borrowers react to the relaxation of the house value constraint brought on by appreciation to the fullest extent possible given their income.

A few studies investigate the relation between housing price changes and aggregate refinancing activities that are of concern to the valuation of MBS. Two such studies are Mattey and Wallace (1998), and Bennett, Keane and Mosser (1999; BKM hereafter). Mattey and Wallace (1998) focus on the collateral constraint effect and provide evidence that a decrease in local housing prices significantly decreases county-level aggregate refinancing activities. Furthermore, they confine their study to California where housing prices declined in their study period from

[3] Patruno (1994) argue that refinancing is the dominant risk in prepayment risk. Mattey and Wallace (1998) provide evidence that the refinancing channel appears to be the strongest channels for housing prices to affect mortgage terminations.

1992 to 1996. In fact, they call for research to investigate the robustness of their finding. BKM (1999), on the other hand, investigate the relation with the seasonally adjusted Mortgage Bankers Association Refinance Index and find no significant relation between housing prices and aggregate refinancing activities at the national level for the period from 1990 to 1998.

In this paper, we provide evidence that local housing prices have a significant effect on aggregate refinancing activities. We examine MSA or PMSA refinancing activities from 1990 to 2000. Furthermore, we extend the previous studies by providing evidence that positive housing price appreciation promotes cash-out refinancing and negative housing price depreciation discourages cash-out refinancing in addition to the collateral constraint effect suggested by Mattey and Wallace (1998) and other previous studies. The evidence is consistent with the ideas that positive house value appreciation provides a motivation for refinancing to change the borrower's capital structure or to fund his consumption, and house value depreciation discourages the borrower to do cash-out refinance. Our evidence supports DSW's (2001) idea that option-based MBS pricing models should be amended to accommodate the direct impact of house prices on refinancing in addition to default behavior.

The rest of this paper is organized as follows: section 3.2 discuss a borrower's incentives and abilities to refinance; section 3.3 discusses the relation between local housing prices and aggregate refinancing activity; section 3.4 describes the data sources and empirical methodology; section 3.5 section presents the empirical results; and section 3.6 concludes with the key results.

3.2 Incentives and Abilities to Refinance

The observed refinancing behavior reflects an individual mortgage borrower's incentives and ability to refinance. The incentive emphasized most in the literature is to exercise the call option imbedded in the standard residential mortgage contracts. In either a fixed-rate mortgage (FRM) or

an adjustable rate mortgage (ARM), the mortgage borrower can be viewed as the seller of a callable mortgage bond to the lender. The borrower can call the mortgage bond at any time prior to its maturity, similar to American-type call options. In theory, a borrower will only exercise the option when it is "in the money".

A common approach measuring whether the refinancing option in a FRM is "in the money" focuses on the difference between the present value of the reductions in monthly interest payments in the future ($PV(PMT_e) - PV(PMT_n)$) and the present value of the transaction costs of refinancing (TC) (Follain and Tzang, 1988). $PV(PMT_e)$ equals the present value of the payments at the existing mortgage rate, and $PV(PMT_n)$ equals the present value of the payment discounted at the new mortgage rate. The transaction costs include legal fees and other origination fees for a new loan. Algebraically the option is in the money if $PV(PMT_e) - PV(PMT_n) > TC.$ The present value of the reductions in interest payments depends on the borrower's expected holding period of the mortgage and the interest spread between the prevailing market mortgage rate for the new loan and the contract rate of the existing loan. Therefore, the larger is the interest spread between the prevailing mortgage rate and the contract rate on the existing loan, the more likely will the borrower refinance. In addition, the shorter is his expected holding period, the less likely will the mortgagor refinance.

Follain, Scott and Yang (1992) reach a similar conclusion. They focus on information asymmetry between the mortgage borrower and the lender. The information asymmetry exists because the borrower knows more about his expected holding period than the lender. When initiating a mortgage, the borrower values the mortgage based on his expected holding period; on the other hand, the lender assigns a value that depends on an estimate of the average holding period

of all borrowers. As a result, a borrower with a below market average holding period will value the call option imbedded in the new loan less than the lender does; on the other hand, a borrower with an above market average holding period will value the option more than the lender does. As such, a borrower with a shorter expected holding period finds refinancing more expensive and thus is less likely to refinance than a borrower with a longer expected holding period. In general, more mobile is a borrower, a shorter period will he expect to hold the mortgage. Thus, a borrower's mobility is likely to inversely relate with his likelihood to refinance.

The holder of an "in-the-money" American option often waits rather than exercise immediately (Hull, 2000). The option is said to have time value. Like an American option, the refinancing option often has time value because future mortgage rates are uncertain. Because the future path of mortgage rats is not known with certainty, there is a potential benefit from postponing refinancing until rates fall still further (Chen and Ling, 1989). Immediate exercise of the refinancing option results in the loss in value from a possible future exercise of the option. The time value of the refinancing option has implications on observed refinancing phenomena. The higher is the expected value or the volatility of future mortgage rates, the less likely is the mortgage borrower to refinance immediately when the refinancing option is in the money (Dickinson and Heuson, 1993). Gilberto and Thibodeau (1989) have supporting evidence for this relation.

A pure floating ARM would seldom be "in the money" because contract rates would closely track market rates (Ambrose and LaCour-Little, 2001). However, an actual ARM contract is not a pure floating rate instrument because of interest rate caps and teasers. A typical ARM has two types of interest rate caps. Periodic caps limit the change in the contract interest rate to a pre-specified percentage points above or below the prior period's contract rate and the lifetime cap establish a maximum and minimum contract rate within a pre-specified percentage points of the

initial contract rate. When interest rates decline so that caps restrict the borrower's access to lower market rates, the refinancing option in an ARM becomes "in the money" (Cunningham and Capone, 1990). Because interest rates are reset periodically, an ARM borrower is less sensitive to both historical and expected interest rate changes than a FRM borrower in terms of refinancing his mortgage (Cunningham and Capone, 1990). Nevertheless, a borrower with expected holding period below the market average is likely to self-select an ARM contract (Yang and Maris, 1993). The shorter holding period offsets the benefit of refinancing from a lower interest rate. In fact, Brueckner (1992) argue that ARM borrowers as a group are unlikely to refinance for financial reasons given their relatively short holding periods. However, the refinancing option embedded in the existing loan may become in the money because a teaser offered by a new ARM loan. By self-selection, a teased ARM borrower may refinance faster than a non-teased ARM borrower. Evidence regarding this assumption is mixing. Green and Shilling (1997) show that teased ARMs are not prepaid faster; however, Ambrose and LaCour-Little (2001) find they are prepaid faster than non-teased ARMs. Ambrose and LaCour-Little (2001) suggest difference of regionally related transaction costs in the two studies' samples may explain their difference in results.[4]

Most existing studies implicitly assume the refinancing borrower replaces his/her existing loan with a new loan of the same type. For instance, a FRM borrower refinances with a new FRM, and an ARM borrower refinances with a new ARM contract. However, the mortgage borrower is not confined to refinance with a same type of loan. Therefore, the intrinsic value of the refinancing option depends on only the interest rates for mortgages of the same type but also the rates for mortgages of the other type. Specifically an ARM borrower with a lower current mortgage rate may be motivated to refinance with a FRM with a higher current mortgage rate when he expects

---

4 Specifically mortgage recording taxes in New York and Florida may effectively increase the transaction costs of

mortgage rates have reached a trough (McConnell and Singh, 1994; cited in LaCour-Little, 2001). Conversely, a FRM borrower may be motivated to refinance with an ARM when he plans to sell the house in the near future (QuickenLoans.com, 2002).

In addition to the incentive to exercise the refinancing option, a mortgage borrower may have incentives to do a cash-out refinance to adjust his housing consumption or to alter his existing capital structure. *Fungibility* is the key concept underlying the cash-out refinancing incentives here. Drawing on the equity in the borrower's property to get cash is fungible from the standpoint of the borrower's overall investment and consumption decision. Since money carries no labels and a mortgage loan usually costs less than other loans (especially after-tax), the borrower can refinance his mortgage to adjust the composition of his personal financial portfolio by taking equity out of his home to pay off debts, invest in other assets, or finance consumption.[5, 6] Refinancing motivated by housing consumption adjustments is driven by a borrower's demographic or socioeconomic characteristics that influence his demand for housing (Gilberto and Thibodeau, 1989). Supporting the assumption, Gilberto and Thibodeau (1989) provide empirical evidence showing that refinancing rates increase with increases in household income and household size, and vary by age of household head. Mortgage.com companies have heavily advertised the advantages of cash-out refinancing to consolidate other debts, in particular credit card debt. Examining the major components of household debt over the period from 1986 to 1992,

---

prepayment and, thus slow the prepayment speed.

5 According to Canner, Luckett and Dukin (1990), a consumer survey sponsored by the Federal Reserve Board during mid-1989 showed that nearly 60 percent of those who refinance also borrow additional funds. Follain, Lekkas and Lehman (1999) document that more than half of all mortgage-swapping deals involved a cash-out refinance in 1998. BPP (2001) find over 43% of refinancers in their study take equity out of their properties.

6 Specifically an individual is allowed to take out up to $100,000 from his or her principal residence in addition to the original debt used to buy the home, and deduct the interest charged before it is repaid (see IRS Publication 936, Home Mortgage Interest Deduction).

Eugeni (1993) have evidence suggesting that consumers have been substituting home equity borrowing for other types of debt.

Just because a mortgage borrower has incentives to refinance, it does not mean that he/she will be able to refinance. There is one important hurdle for the borrower to cross: he/she must be able to qualify for a new mortgage loan. In addition to evaluating the borrower's credit report and the title, the lender generally uses FNMA/FHLMC guidelines to evaluate a borrower's risk in his underwriting process.[7] The lender is likely to reject the borrower if he/she is considered too risky according to the guidelines.

We briefly describe the guidelines here (see Floyd and Allen (1999) for more detail). The guidelines fall into three categories: loan-to-value (LTV) ratio, income ratios, and down payment sources. The LTV ratio is determined by dividing the requested loan amount by the value of the applicant's property at time of requesting the loan. As a general guideline, a loan application with an LTV ratio below 80 percent is routinely accepted. An application with the ratio above 80 percent will generally be asked to reduce the ratio to 80 percent or to apply for mortgage insurance. Two income ratios are considered: the mortgage debt ratio (MDR) and the total debt ratio (TDR). The MDR is defined as the percentage of a borrower's gross monthly income required to meet monthly housing expenses. As a guideline for conventional loans, the MDR must not exceed 28 percent.[8] The TDR is defined as the percentage of a borrower's gross monthly income required to meet monthly contractual expenses. In general, the TDR must not exceed 36 percent for conventional loans. The third category of guideline refers to sources for the borrower's down

[7] Peristiani, Bennett, Monsen, Peach and Raiff (1997a and b) have empirical evidence showing that a borrower's personal creditworthiness affects his refinancing possibility.

[8] Conventional loans are those that not insured by the Federal Housing Administration (FHA), guaranteed by the Veterans Administration (VA), or insured by Farmer Home Administration (FmHA).

payment for a home purchase. In general, funds for the down payment should be provided by the borrower rather than outside sources.

The lending environment may also affect a borrower's incentives and ability to refinance (PBMPR, 1997 a and b; BKM, 1999, BPP, 2001). Structural changes on both the supply side and the demand side of the mortgage market have reduced transaction costs and frictions associated with obtaining a mortgage loan (BPP, 2001). The lending industry has become much more competitive and aggressive in regards to soliciting refinancing. Mortgage lenders have begun to contact potential borrowers to encourage them to refinance. Information process technology has shortened the period from application to approval and then from approval to closing. The increased competition in the primary mortgage market together with information process technology has lowered both financial and nonfinancial transaction costs to refinance. Having experienced successive waves of intensive refinancing activity, mortgagors have increased the general level of awareness of the potential benefits of refinancing and have a propensity to refinance with a smaller interest rate spread than before (BPP, 2001).

3.3 Local Housing Prices and Aggregate Refinancing Activities

Aggregate refinancing activities are outcomes of individual refinancing behavior. Local housing prices are the aggregate representations of individual house prices in local markets. Therefore, local housing prices may be related to aggregate refinancing activities because individual house prices influence individual homeowners' incentives and abilities to refinance his/her adjustable-rate or fixed-rate mortgage.

As discussed in Section 3.2, an application for a new loan with the LTV ratio below 80% is more likely, or less expensive, to be approved. The LTV ratio guideline is designed to protect the lender from default risk: the higher the LTV ratio, the less equity has the borrower on his house. A

borrower who has a little equity is more likely to default, if the value of the property should fall below the loan amount. In a refinancing case, the LTV ratio is defined as the requested new loan amount divided by the borrower's house price at time of requesting the loan. Given a fixed loan amount, the higher a house is priced, the more likely the LTV ratio is below 80%. The borrower is therefore more likely to qualify for refinancing. Aggregate house prices and the LTV ratios across individuals. The higher is a current local housing price, the more proportion of individual borrowers is likely to have the LTV ratios below 80 percent. Thus we expect current local housing prices to be positively related with aggregate refinancing rates.

Home equity not only provides protection for the lender, but also provides the ability for the borrower to borrow additional funds to adjust his/her housing consumption, or to alter his/her existing capital structure, when refinancing his/her mortgage. Home equity changes over time after the origination of a loan. The borrower actually builds up his/her equity when he makes his/her monthly mortgage payments in an amortizing loan.[9] Each monthly payment consists of an interest portion and a principal portion. The principal portion is the amount of equity built by the borrower in each payment. Home equity also changes when the borrower's house price changes. Home equity is the value of the house price minus the unpaid mortgage balance. Therefore, house appreciation increases home equity, and house depreciation decreases home equity. Thus, changes in the house price positively affect the likelihood of the borrower to apply a cash-out refinance.

Nevertheless, changes of a borrower's house price may not have a symmetric influence in magnitude. House price appreciation may have a weaker influence on the likelihood of a borrower to refinance than house price depreciation because home equity increases over time whenever the borrower makes monthly mortgage payments. When enough time passes, the built-up home equity

alone can provide enough benefit for the borrower to refinance in order to borrow additional funds. In other words, the borrower would still do cash-out refinancing even if his house price does not change. On the other hand, house price depreciation would erode the built-up equity and discourage the borrower to refinance at the moment when monthly payments build just enough equity.

Furthermore, house price appreciation and depreciation may interact with mortgage rate movements. The borrower is much more likely to cash out the untapped equity brought on by house appreciation when his existing mortgage carries an above-market interest rate than when the existing mortgage carries a below-market rate. We have this proposition because the borrower is likely to apply a subordinate mortgage to cash out his home equity when his existing mortgage carries a below-market interest rate. On the other, house price depreciation may be more easily to discourage the borrower to do a cash-out refinance when his existing loan carries a below-market mortgage rate than when carries an above-market rate. We expect the borrower to have this behavior because the borrower has to give up a loan with a positive net present value when his loan carries a below-market rate. Thus, the opportunity cost per dollar of the additional borrowed funds is larger when the existing loan carries a below-market interest rate than when the loan carries an above-market rate.

Reflecting at the aggregate level, positive changes of local housing prices are expected to increase cash-out refinances, and negative changes of local housing prices are expected to decrease refinances. In additional, positive changes of local housing prices are expected to have greater influences on refinancing when interest rates have declined than when rates have increased or unchanged. On the other hand, negative changes of local housing prices are expected to have

[9] Amortizing loans are the dominant loan type in the US housing industry after 1930 because of the promotion of the

greater influence on refinancing when mortgage rates have increased or unchanged compared with when mortgage rates have declined.

3.4 Data Sources and Methodology

This section consists of two parts. Part 1 discusses the data employed and describes refinance rates over U.S. cities from 1990 to 2000. Part 2 presents the empirical methodology used to investigate the relations between local housing prices and refinance rates discussed in Section 3.3.

3.4.1 Data Sources

To examine the relationship between local housing prices and aggregate refinancing activities from the period of 1990 to 2000, we collected data from five sources. The sources are the Home Mortgage Disclosure Act (HMDA) Aggregate Reports, the American Housing Survey (AHS) for selected metropolitan areas, Monthly Interest Rate Survey (MIRS) of the Federal Housing Finance Board and the Federal Home Loan Bank Board, the Office of Federal Housing Enterprise Oversight (OFHEO) house price index, and the 1990 census. From these sources, we obtained relevant data for 14 Metropolitan Statistical Areas (MSAs) or Primary Metropolitan Statistical Areas (PMSAs).[10] In total we obtained 154 observations.

The HMDA Aggregate Reports provide total numbers of originated refinancing loans for 1-to-4 family dwellings for the 14 metropolitan areas each year.[11, 12] We estimate mortgage

---

Federal Housing Administration (Floyd and Allen, 1999).

[10] The 14 metropolitan areas are Atlanta, GA, Chicago, IL, Columbus, OH, Denver, CO, Detroit, MI, Houston, TX, Indianapolis, IN, Kansas City, MO-KS, Minneapolis-St. Paul, MN-WI, Philadelphia, PA-NJ, Pittsburgh, PA, Rochester, NY, St. Louis, MO-IL, and San Diego, CA.

[11] The Federal Institutions Examinations Council (FFIEC) produces the HMDA Aggregate Reports using information provided on the loan/application registers. In general, a depository lender that has assets above a certain level and has a home or branch office or a nondepository lender that has lending activity in a MSA or PMSA has to maintain a loan/application register to enter data about each application received and each loan originated or purchased. See "A Guide to HMDA Reporting: Getting it Right" (Federal Financial Institutions Examination Council, 1998) for more details.

[12] The refinanced loans can be conventional, FHA-insured, VA-guaranteed, or FmHA-insured loans for home purchases or home improvements. The refinanced loans can be either adjustable-rate or fixed-rate loans.

inventories for 1-to-4 family residential properties for the 14 metropolitan areas from the AHS. The AHS provides owner-occupied numbers with mortgages only for 1-unit structures. The survey also provides owner-occupied numbers of 1-unit detached structures, 1-unit attached structures, and 2-to-4 unit structures. To estimate mortgage inventories for 1-to-4 family properties, we assume the proportion of 2-to-4 unit structures with mortgages is the same to the proportion of 1-unit structures with mortgages. The AHS is conducted 5 times for the period from 1984 to 2004: 1984-1987, 1988-1991, 1992-1994, 1995-1998, and 1999-2004. Not every metropolitan area is surveyed each time. The estimated numbers of the mortgage inventories from the conducted surveys are cubic splined to estimate the annual numbers.

Table 3.1 presents the refinancing rates computed with the estimated mortgage inventories and the numbers of refinancing loans from the HMDA Aggregate Reports for the 14 metropolitan areas from 1990 through 2000. The computed refinancing rates clearly pick up the 1992-1993 and 1998-1999 refinancing booms. As shown in Table 3.1, there are substantial variations in refinancing rates not only over the period 1990 to 2000 at each metropolitan area, but also among the 14 metropolitan areas each year, in particular 1992-1993 and 1998-1999. The two peaks and variations of refinancing activity also clearly stand out in Figure 3.1.

We use the OFHEO quarterly repeat sales house price index for the housing price time series over the 14 metropolitan areas. The OFHEO estimates the index for single-family residential properties using the data on conventional conforming mortgage transactions obtained from the Freddie Mac and Fannie Mae. To obtain the dollar values of local housing prices every year, we first take the four-quarter average for each year then multiply the annual numbers with the median owner-occupied house values for the metropolitan areas from the 1990 Census. As shown in Figure 3.2, there are substantial variations among local housing prices each year and over time.

Table 3.1: Refinance Rates for the 14 Metropolitan Areas

Refinance Rates	Atlanta, GA	Chicago, IL	Columbus, OH	Denver, CO	Detroit, MI	Houston, TX	Indianapolis, IN	Kansas City, MO-KS
1990	1.25%	1.39%	1.42%	0.76%	1.92%	0.59%	2.01%	1.34%
1991	3.29%	3.90%	4.09%	2.75%	5.43%	1.20%	4.18%	2.90%
1992	9.75%	10.77%	12.16%	8.83%	15.08%	5.36%	11.67%	8.92%
1993	17.80%	15.42%	16.51%	21.96%	20.95%	10.73%	15.42%	15.43%
1994	5.94%	6.90%	6.17%	7.25%	6.69%	4.15%	6.57%	5.79%
1995	4.13%	4.84%	4.09%	5.13%	6.04%	1.52%	4.61%	3.88%
1996	6.58%	8.06%	8.90%	10.34%	9.70%	2.75%	9.63%	7.35%
1997	7.23%	9.17%	8.38%	11.27%	11.63%	1.94%	9.48%	8.41%
1998	17.91%	21.53%	17.65%	28.74%	27.96%	8.32%	19.16%	18.54%
1999	10.83%	13.37%	11.12%	16.45%	17.10%	6.09%	11.05%	12.48%
2000	6.20%	7.65%	5.76%	9.46%	10.92%	3.35%	6.05%	7.40%
Standard Deviation	5.45%	5.72%	5.19%	8.33%	7.65%	3.19%	5.12%	5.28%

(Table 3.1 continued)

Refinance Rates	Minneapolis -St. Paul, MN-WI	Philadelphia, PA-NJ	Pittsburgh, PA	Rochester, NY	St. Louis, MO-IL	San Diego, CA	Standard deviation
1990	1.36%	0.89%	0.75%	1.88%	2.69%	5.62%	1.27%
1991	4.49%	2.06%	1.72%	2.76%	4.86%	9.74%	2.10%
1992	12.17%	6.78%	6.36%	7.73%	14.99%	21.55%	4.29%
1993	25.35%	10.62%	9.38%	12.37%	21.47%	29.08%	5.83%
1994	6.50%	4.15%	4.81%	4.48%	7.83%	11.32%	1.85%
1995	3.54%	2.30%	3.37%	2.42%	5.17%	6.46%	1.41%
1996	9.29%	3.83%	6.17%	4.62%	8.38%	8.29%	2.34%
1997	7.93%	3.91%	6.44%	4.42%	9.68%	9.71%	2.81%
1998	21.89%	8.82%	13.00%	12.56%	22.49%	27.42%	6.63%
1999	11.78%	6.54%	9.84%	8.00%	13.17%	17.30%	3.57%
2000	7.31%	3.71%	6.57%	3.76%	7.83%	8.87%	2.21%
Standard Deviation	7.45%	2.99%	3.60%	3.78%	6.57%	8.40%	6.27%

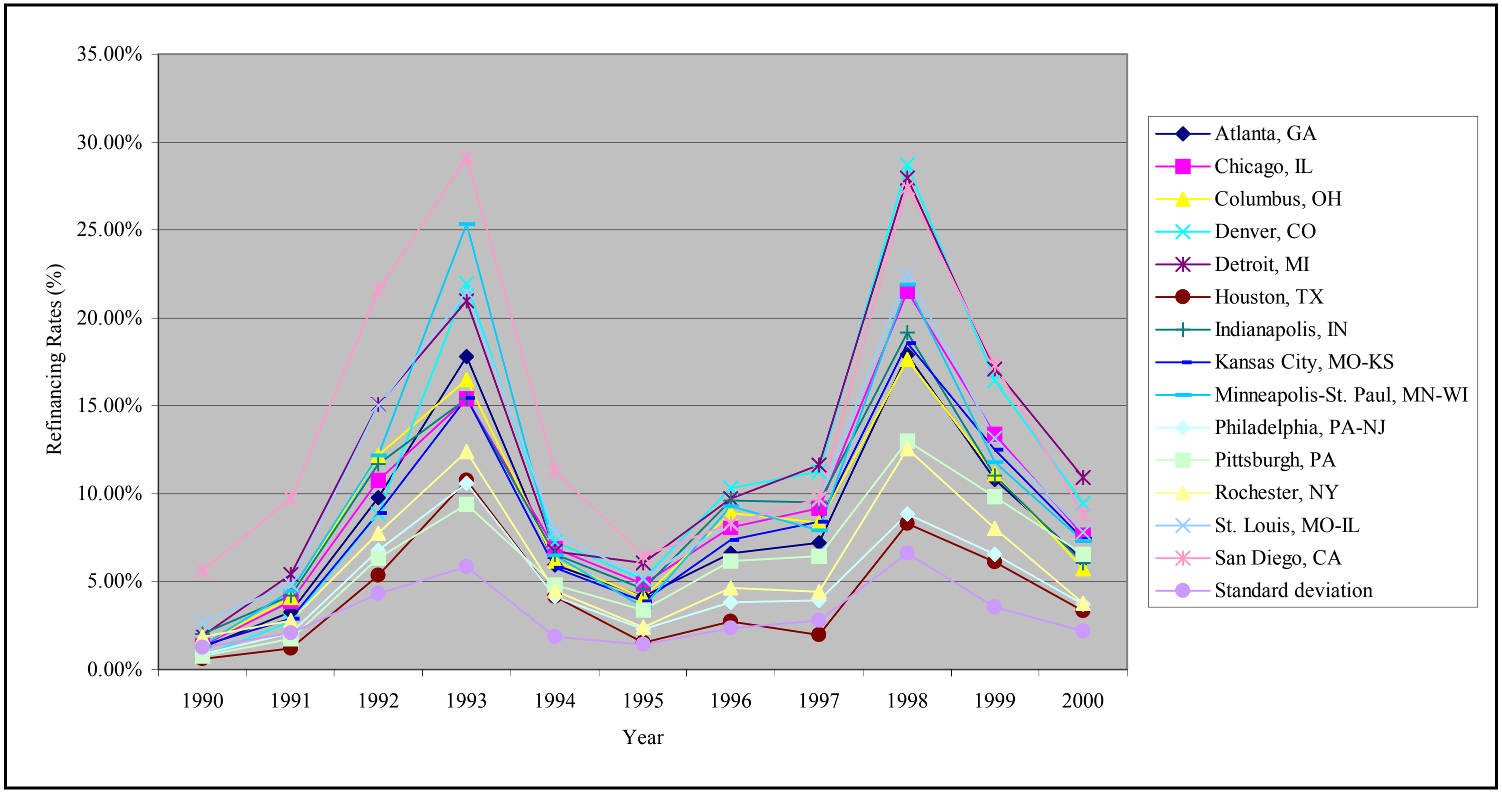

Figure 3.1 Local Refinance Rates

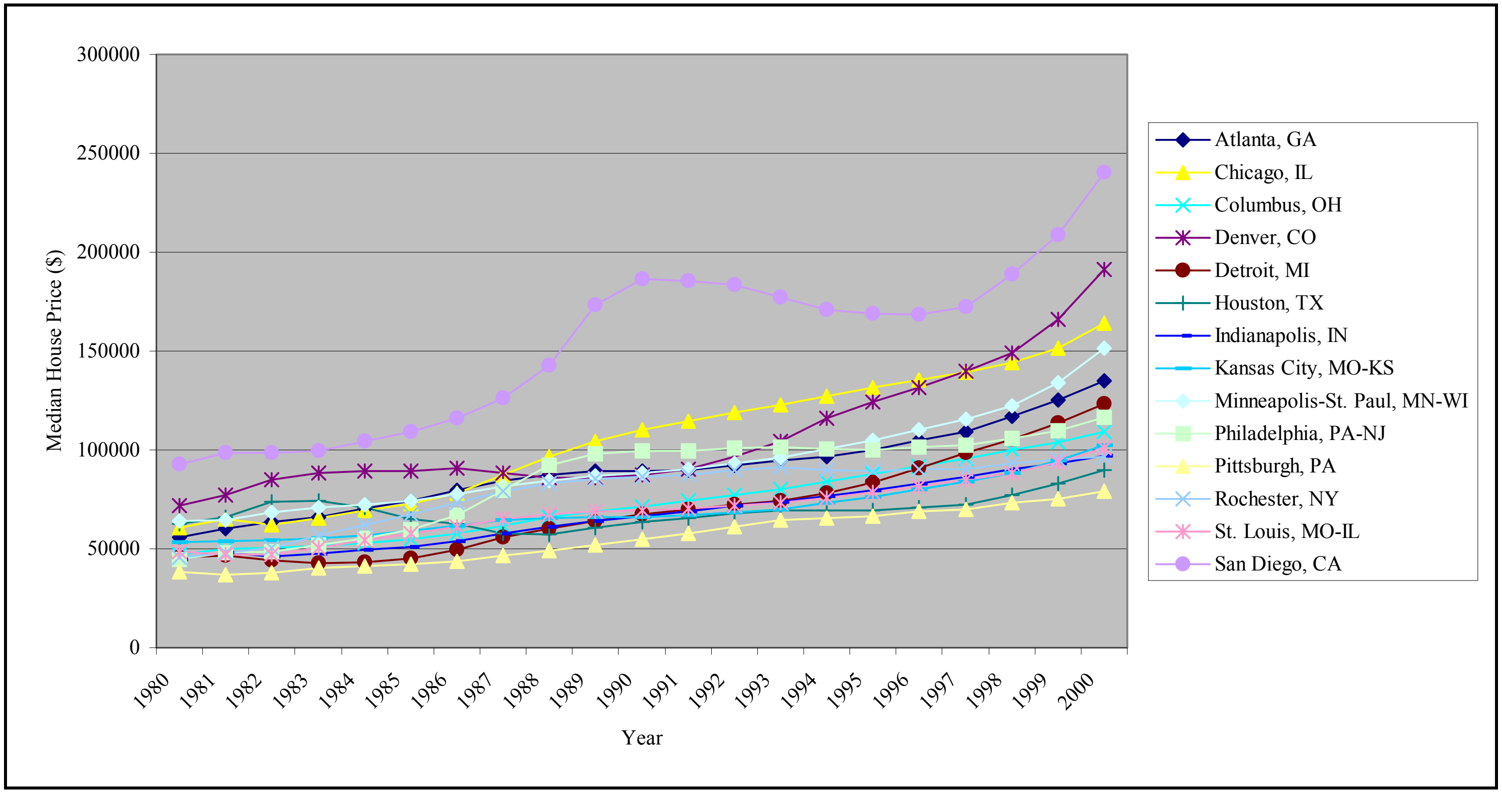

Figure 3.2 Local Housing Prices

The MIRS is conducted by the Federal Housing Finance Board since 1989 and was conducted by the Federal Home Loan Bank Board on a sample of major mortgage lenders that includes savings and loan associations, savings banks, commercial banks, and mortgage companies. Such lenders have accounted for more than 90 percent of all conventional home mortgage loan originations. We use the annual contract interest rates of the MIRS to trace the mortgage rate movements for the 14 metropolitan areas. The MIRS provides annual historical contract interest rates on single-family conventional loans for selected metropolitan areas dated back to 1978. The contract rates are averages, weighted by lenders' mortgage holdings, of the initial interest rates reported on the mortgages for both adjustable-rate and fixed-rate loans. As shown in Figure 3.3, local mortgage interests are very close and move very closely over time. The pattern suggests that interest rate variation is unlikely to be the only force causing the variation of refinancing activity among the metropolitan areas.

3.4.2 Methodology

Obtaining the data above, we operationalize the conceptual relations between local housing prices and aggregate mortgage refinancing activities. As discussed earlier, current local housing prices may affect refinancing due to the collateral constraint effect, and housing price changes may influence cash-out refinancing due to home equity changes. Housing price changes may have an asymmetric influence in magnitude due to house equity built up by monthly mortgage payments. Furthermore, housing price depreciation and housing price appreciation may interact with interest rate movements. In particular, housing price depreciation affects refinancing more when interest rates have increased or unchanged than when interest rates have declined, due to different opportunity costs of giving up existing loans. On the other hand, housing price appreciation affects

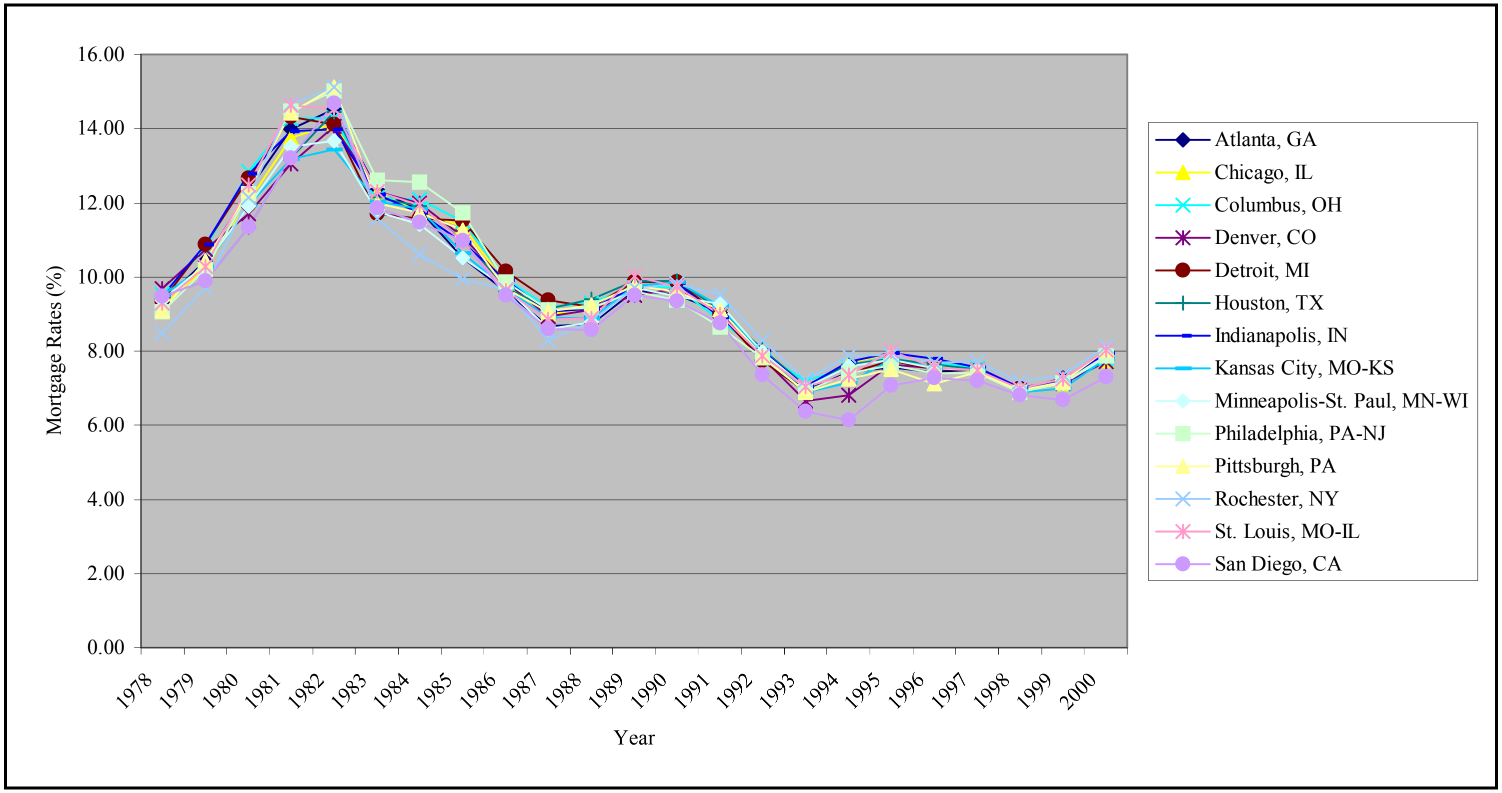

Figure 3.3 Local Mortgage Rates

refinancing more when interest rates have declined, because a borrower may apply a subordinate mortgage when interest rates have increased or unchanged.

To allow the possibilities of asymmetric influence, we calibrate the relations empirically with Equation (1).[13]

$$
\begin{aligned}
\ln(\textit{refinancing loans}) = \beta_1 + \beta_2(\textit{conditional housing price depreciation}) \\
+ \beta_3(\textit{housing price depreciation}) \\
+ \beta_4(\textit{conditional housing price appreciation}) \\
+ \beta_5(\textit{housing price appreciation}) \\
+ \beta_6 \ln(\textit{current housing prices}) + \beta_7 \ln(\textit{mortgaged households}) \\
+ \beta_{8-17}(\textit{year dummies}) + \beta_{18-30}(\textit{city dummies}) + \varepsilon
\end{aligned}
\tag{1}
$$

Where refinancing loans means the number of refinancing loans for 1-to-4 family dwellings, mortgaged households means the number of 1-to-4 family dwellings with outstanding mortgages, house price depreciation is the absolute value of negative housing price changes, housing price appreciation is positive housing price changes, conditional housing price deprecation is defined as housing price depreciation conditional on interest rates having increased or unchanged and, conditional housing price appreciation is defined as housing price appreciation conditional on interest rate having declined. [14,15]

The 10 individual year dummies are used to capture interest rate movements and the lending environment. The 13 city dummies are used to proxy for household characteristics and

---

13 RESET tests and R-squared suggest the double-log model instead of linear or semi-log models.

14 Housing price changes are cumulative changes of median housing prices over past five years. We choose five years as the interval for measuring housing price changes because of two reasons. First, recent three refinancing booms occur every five years. This pattern suggests that five years probably is the holding period of a mortgage for most borrowers. Second, we collect the median origination years of primary mortgages from various volumes of AHS since 1984. Then we calculate the average holding periods form the median origination years for the 14 metropolitan areas. The average holding period is 4.63 years. This also suggests that five years probably is the holding period of a mortgage for most borrowers.

15 We implement the calibration of Equation (1) in this paper using LeSage's (1999) MATLAB functions in Applied Econometrics using MATLAB.

location related unobservable variables. The year dummies and the city dummies also help capture the effects of compositions of loan types and coupon concentration of in a metropolitan area. [16]

3.5 Empirical Results

Before the empirical results for Equation (1), the properties of important variables are presented. Tables 3.2-1 and 3.2-2 present their summary statistics. On average there are 42,021 loans refinanced in a metropolitan area in a year during 1990 to 2000. However there exists great variation. In Chicago, IL, 258,469 loans were refinanced during 1998. In Pittsburgh, PA, only 2,088 loans were refinanced during 1990. The standard deviation is 40,385 loans, nearly as many as the mean. Housing prices, on average, increase \$3,158 over a 5-year period that most borrowers hold their mortgages before refinancing. Unlike the observations studied by Mattey and Wallace (1998) that most declined in their housing prices, only 18 observations experienced housing price depreciation among our 154 year-city observations. Housing price depreciation reached a maximum of \$6,532. Among the 18 observations, 7 observations experienced declines in mortgage rates as well. The rest of the observations experienced increases in mortgage rates. The other 136 observations experienced housing price appreciation. Housing price appreciation reached a maximum of \$30,279. There are 48 out of the 136 observations that had increases in interest rates, and 88 out of the 136 that experienced declines in interest rates.

Tables 3.3-1 to 3.3-3 cross-tabulates refinance rates for local housing price movements and interest rate movements. Consistent with the discussion in Section 3.2, Table 3.3-1 clearly indicates the need to control factors other than interest rate movements and local housing price movements. Specifically, refinance rates are higher on average when interest rates have increased

---

[16] The refinanced loans used in our study can be either FRMs or ARMs, or conventional loans, FHA-insured, VA-guaranteed, or FmHA-insured loans.

Table 3.2-1: Descriptive Statistics

Variables-Description	n	Mean	Median	Maximum	Minimum	Standard Deviation
Refinancing loans = numbers of loans refinanced	154	42,021.084	30,396.500	258,469,000	2,088.000	40,384.551
Housing price changes = differences between current median house value and median house value 5 years ago in thousand dollars	154	3.158	2.772	30.279	-6.532	3.872
Housing price depreciation = absolute value of negative housing price changes	154	0.220	0.000	6.532	0,000	0.900
Conditional housing price depreciation = housing price depreciation conditional on interest rates have increased or unchanged over past 5 years	154	0.0587	0.000	2.485	0.000	0.291
Housing price appreciation = positive housing price changes	154	3.377	2.772	30.279	0.000	3.562
Conditional housing price appreciation = housing price appreciation conditional on interest rates have declined over past 5 years	154	2.057	1.226	16.524	0.000	2.784
Current house prices = current median house value in thousand dollars	154	100.894	92.080	240.476	55.000	33.806
Mortgaged households = numbers of owner-occupied households with mortgages in thousand households	154	480.793	370.470	1203.920	156.615	292.172

Table 3.2-2: Descriptive Statistics for Housing Price Movements Excluding Zero Value Observations

Variables-Description	n	Mean	Median	Maximum	Minimum	Standard Deviation
Housing price depreciation= absolute value of negative housing price changes in thousand dollars	18	1.879	1.133	6.532	0.076	1.996
Conditional housing price depreciation = housing price depreciation conditional on interest rates have increased or unchanged over past 5 years	11	0.822	0.635	2.485	0.076	0.780
Housing price appreciation = positive housing price changes in thousand dollars	136	3.824	3.0035	30.279	0.006	3.558
Conditional housing price appreciation = housing price appreciation conditional on interest rates have declined over past 5 years	88	3.601	2.831	16.524	0.006	2.830

Table 3.3-1: Contingency Table for Refinance Rates

	Housing price movement		
Interest rate movement	Housing price depreciation	Housing price appreciation	Total
Interest rates increased	10.01% (11)	14.30% (48)	13.50% (59)
Interest rates declined	8.06% (7)	5.73% (88)	5.90% (95)
Total	9.25% (18)	8.75% (136)	8.81% (154)

Note: The numbers of observations are in the parentheses.

Table 3.3-2: Contingency Tables for Refinance Rates for the 14 Metropolitan Areas

		Housing price movement		
	Interest rate movement	Depreciation	Appreciation	Total
Atlanta, GA	Increased	-	-	-
	Declined	4.13%	8.68%	8.26%
	Total	4.13%	8.68%	8.26%
Chicago, IL	Increased	-	-	-
	Declined	14.79%	8.16%	9.36%
	Total	14.79%	8.16%	9.36%
Columbus, OH	Increased	11.12%	11.64%	11.51%
	Declined	-	7.17%	7.17%
	Total	11.12%	8.51%	8.75%
Denver, CO	Increased	0.76%	10.58%	9.49%
	Declined	-	18.78%	18.78%
	Total	0.76%	12.22%	11.18%
Detroit, MI	Increased	5.43%	12.80%	12.13%
	Declined	-	-	-
	Total	5.43%	12.80%	12.13%
Houston, TX	Increased	8.32%	2.19%	3.21%
	Declined	-	5.34%	5.34%
	Total	8.32%	3.77%	4.18%
Indianapolis, IN	Increased	-	4.18%	4.18%
	Declined	-	9.56%	9.56%
	Total	-	9.08%	9.08%

(Table 3.3-2 continued)

	Interest rate movement	Housing price movement: Depreciation	Appreciation	Total
Kansas City, MO-KS	Increased	-	-	-
	Declined	7.35%	8.51%	8.40%
	Total	7.35%	8.51%	8.40%
Minneapolis-St. Paul, MN-WI	Increased	-	-	-
	Declined	11.78%	9.98%	10.14%
	Total	11.78%	9.98%	10.14%
Philadelphia, PA-NJ	Increased	-	-	-
	Declined	-	4.87%	4.87%
	Total	-	4.87%	4.87%
Pittsburgh, PA	Increased	-	7.33%	7.33%
	Declined	1.72%	0.75%	1.24%
	Total	1.72%	6.67%	6.22%
Rochester, NY	Increased	4.46%	6.73%	6.05%
	Declined	4.48%	-	4.48%
	Total	4.46%	6.73%	5.91%
St. Louis, MO-IL	Increased	12.11%	7.75%	9.39%
	Declined	-	14.49%	14.49%
	Total	12.11%	10.28%	10.78%
San Diego, CA	Increased	8.87%	-	8.87%
	Declined	-	14.65%	14.65%
	Total	8.87%	14.65%	14.13%

Table 3.3-3: Contingency Tables for Refinance Rates from 1990 to 2000

		Housing price movement		
Year	Interest rate movement	Depreciation	Appreciation	Total
1990	Increased	-	-	-
	Declined	0.59%	1.79%	1.70%
	Total	0.59%	1.79%	1.70%
1991	Increased	-	-	-
	Declined	1.98%	4.12%	3.81%
	Total	1.98%	4.12%	3.81%
1992	Increased	7.09%	10.26%	9.69%
	Declined	-	15.18%	15.18%
	Total	7.09%	11.49%	10.86%
1993	Increased	21.96%	16.96%	17.32%
	Declined	-	-	-
	Total	21.96%	16.96%	17.32%
1994	Increased	5.79%	5.83%	5.82%
	Declined	-	6.70%	6.70%
	Total	5.79%	6.37%	6.32%
1995	Increased	-	-	-
	Declined	6.46%	3.93%	4.11%
	Total	6.46%	3.93%	4.11%
1996	Increased	-	-	-
	Declined	8.29%	7.35%	7.42%
	Total	8.29%	7.35%	7.42%
1997	Increased	-	-	-
	Declined	9.71%	7.68%	7.83%
	Total	9.71%	7.68%	7.83%
1998	Increased	10.69%	19.74%	18.35%
	Declined	27.42%	-	27.42%
	Total	16.27%	19.74%	19.00%
1999	Increased	9.48%	12.72%	11.79%
	Declined	-	-	-
	Total	9.48%	12.72%	11.79%
2000	Increased	8.87%	-	8.87%
	Declined	-	6.61%	6.61%
	Total	8.87%	6.61%	6.77%

than when have decreased. One possible reason is that an ARM borrower is little sensitive to historical interest rate increases and refinances to take advantages of the teaser offered by a new loan (Ambrose and LaCour-Little, 2001) or to take home equity out from an overall investment or consumption viewpoint, whenever the borrower has the ability to do so. Table 3.3-2 consists of 14 contingency tables each for a metropolitan area. The tables support this speculation. In the 8 metropolitan areas that experienced both interest rate increases and decreases, 5 metropolitan areas had higher refinance rates on average when interest rates had decreased than had increased. When pooled together for all years and metropolitan areas, finance rates on average are slightly higher when local housing prices have depreciated than when have appreciated in Table 3-1. A possible cause is that coupon concentration of existing FRMs varies over time.[17] A FRM borrower would like to refinance for a new loan with a lower interest rate as long as housing price depreciation does not disqualify the borrower. Table 3-3 consists of 11 contingency tables each for a year from 1990 to 200. The tables support this speculation. In the 11 years, 6 years had higher refinance rates on average when local housing price had appreciated than when had depreciated.

Table 3.4 presents the empirical results of Equations (1). Not surprisingly, the mortgaged households number always has a significant and positive coefficient in all specifications. The coefficient for current housing prices is always significant and positive in each specification. The coefficient strongly supports the idea that the collateral constraint or the LTV ratio guideline has a great influence on a borrower's ability to refinance. This empirical result is consistent with the results of previous studies such as Cunningham and Capone (1990), and PBMPR (1997 a and b). The two studies show LTV ratio having a negative impact on mortgage refinancing. On the other hand, Caplin, Freeman, and Tracy (1997) and Mattey and Wallace (1998) use local housing price

[17] BKM (1999) show that coupon concentration of outstanding mortgages affects aggregate refinance rates over time.

Table 3.4: Housing Price Movement and Refinancing Activity

Dependent variable: Ln (refinancing loans)

	Specification 1		Specification 2		Specification 3		Specification 4	
Independent variables:	β	t	β	t	β	t	β	t
Conditional housing price depreciation	-0.0253	-0.8051						
Housing price depreciation	-0.0362	-1.7167*	-0.0370	-1.7127*	-0.0319	-1.6182		
Conditional housing price appreciation	0.0153	2.1872**	0.0150	2.1697**	0.0130	1.7760*	0.0168	2.5071**
Housing price appreciation	-0.0055	-1.1829	-0.0049	-1.0565				
Ln (Current house price)	1.4965	4.1650***	1.5123	4.3335***	1.4358	4.4812***	1.5140	4.8419***
Ln (mortgaged households)	0.8273	3.6873***	0.8313	3.6777***	0.8171	3.6240***	0.7827	3.4183***
n	154		154		154		154	
k	30		29		28		27	
R^2	0.9764		0.9763		0.9762		0.9757	
Adjusted R^2	0.9709		0.9710		0.9711		0.9708	
Model F-value	176.7361***		184.2058***		191.4600***		196.3711***	
F-value for RESET test	1.6027		1.5795		1.7537		1.6940	

Note:

1. Refinancing loans = numbers of refinanced loans.
2. Housing price depreciation, and appreciation, and current housing prices are in thousand dollars.
3. Conditional housing price depreciation is housing price depreciation conditional on interest rates have increased or unchanged.
4. Conditional housing price appreciation is housing price appreciation conditional on interest rates have declined.
5. Mortgaged households are in thousand households.
6. All specifications include a constant and 13 city dummies and 10 year dummies not shown here.
7. The RESET test introduces the square, cubic and fourth powers of predicted Ln(refinancing loans) as additional regressors.
7. t-values in parentheses are based on heteroscedasticity-consistent errors.
9. *** significant at 1% level. 8. ** significant at 5 % level. 9. * significant at 10% level.

movements alone to show the collateral constraint effect. Intuitively, housing movements pick up both the effect of collateral constraint and the effect on cash-out refinancing.

The conditional housing price appreciation always has a significant and positive coefficient in all specifications. This coefficient strongly supports that housing price appreciation boosts cash-out refinancing when interest rates have declined. The housing price appreciation does not have a significant coefficient. This coefficient is consistent with the idea that the borrower is likely to apply a subordinate mortgage to cash out his home equity when his existing mortgage carries a below-market interest rate.

The housing price depreciation has a significant and negative coefficient in 2 out of the 3 specifications. This result weakly supports that negative housing price changes decrease home equity and, thus discourage cash-out refinancing. Nevertheless, the conditional housing price depreciation does not have a significant and negative coefficient. This coefficient may reflect the importance of home equity in cash-out refinancing. Without sufficient home equity, the borrower is equally unlike to apply a cash-out refinance either when interest rates have declines or increased.

3.6 Conclusions

We provide evidence supporting that housing prices have a significant effect on aggregate refinancing activities. Consistent with Mattey and Wallace (1998), we have evidence showing the collateral constraint effect of housing prices on aggregate refinancing activity. In contrast to previous studies, we further contribute the refinancing literature by providing empirical evidence consistent with our proposed relations between housing price movements and aggregate cash-out refinancing activities.

We have evidence supports the idea that local housing price changes interacting with mortgage rate movements affect cash-out refinancing. Our evidence shows that positive

appreciation in housing price expedites the borrower to refinance in response to the associated increased borrowing capacity when mortgage rates have declined. On the other hand, depreciation in housing price may hold down the borrower to refinance.

Our results have important implications. Our results confirm that housing price dynamics play an important role in estimating refinancing risk and, therefore, are important in adequately pricing mortgages and MBS. In particular, we show that housing price movements, not only depreciation pointed out by DSW (2001) but also appreciation, should be included in modeling total termination risks of MBS.

3.7 References

Avery, Robert B., Raphael W. Bostic, Paul S. Calem, and Glenn B. Canner. (2000). "Credit Scoring: Statistical Issues and Evidence from Credit-Bureau Files," *Real Estate Economics* 28, 523-547.

Ambrose, Brent W., and Michael LaCour-Little. (2001). "Prepayment Risk in Adjustable Rate Mortgages Subject to Initial Year Discounts: Some New Evidence," *Real Estate Economics* 29, 305-327.

Archer, Wayne, David C. Ling, and Gary A. McGill. (1996). "The Effect of Income and Collateral Constraints on Residential Mortgage Terminations," *Regional Science and Urban Economics* 26, 235-61.

Bennett, Paul, Frank Keane, and Patricia C. Mosser. (1999). "Mortgage Refinancing and the Concentration of Mortgage Coupons," *Current Issues in Economics and Finance* 5, 1-6.

Bennett, P., R. Peach, and S. Peristiani. (2001). "Structural Change in the Mortgage Market and the Propensity to Refinance," *Journal of Money, Credit, and Banking* 33, 955-975.

Brueckner, Jan K. (1992). "Borrower Mobility, Self-Selection, and the Relative Prices of Fixed- and Adjustable-Rate Mortgages," *Journal of Financial Intermediation* 2, 401-421.

Canner, Glenn B., Charles A. Luckett, and Thomas A. Durkin. (1990). "Mortgage Refinancing," *Federal Reserve Bulletin* 76, 604-612.

Caplin, Andrew, Charles Freeman, and Joseph Tracy. (1997). "Collateral Damage: Refinancing Constraints and Regional Recessions," *Journal of Money, Credit, and Banking* 29, 496-516.

Chen, Andrew H., and David C. Ling. (1989). "Optimal Mortgage Refinancing with Stochastic Interest Rates," *AREUEA Journal* 17, 276-299.

Cunningham, Donald F., and Charles A. Capone, Jr. (1990). "The Relative Termination Experience of Adjustable to Fixed-Rate Mortgages," *The Journal of Finance* 45, 1687-1703.

Dickinson, Amy, and Andrea J. Heuson. (1993). "Explaining Refinancing Decisions using Microdata," *Journal of the American Real Estate and Urban Economics Association* 21, 293-311.

Downing, Chris, Richard Stanton, and Nancy Wallace. (2001). "An Empirical Test of a Two-Factor Mortgage Prepayment and Valuation Model: How Much Do House Price matter?" paper presented at *AFA 2001 New Orleans Meeting*.

Eugeni, Francesca. (1993). "Consumer Debt and Home Equity Borrowing," *Economic Perspectives: a Review from the Federal Reserve Ban of Chicago* 17, 2-13.

Federal Financial Institutions Examination Council. (1998). *A Guide to HMDA Reporting: Getting it Right.*

Follain, James R., Louis O. Scott, and T. L. Tyler Yang. (1992). "Microfoundations of a Mortgage Prepayment Function," *Journal of Real Estate Finance and Economics* 5, 197-217.

Follain, James R., and Dah-Nei Tzang. (1988). "Interest Reate Differential and Refinancing a Home Mortgage," *The Journal of Appraisal* April, 243-251.

Follain, James, Vassillis Lekkas, and H. Jane Lehman. (1999). "Refinancers Crave Cash on Top of Lower Monthly Prepayments," *Secondary Mortgage Markets* 16, 28-33.

Floyd, Charles F. and Marcus T. Allen (1999). *Real Estate Principles*, Sixth Edition, Dearborn Financial Publishing, Inc.

Gilberto, Michael, and Thomas G. Thibodeau. (1989). "Modeling Conventional Residential Mortgage Refinancings," *Journal of Real Estate Finance and Economics* 2, 285-299.

Green, Richard K., and James D. Shilling. (1997). "The Impact of Initial-Year Discounts on ARM Prepayments," *Real Estate Economics* 25, 373-385.

Hilliard J. E., I, B. Kau, and V. C. Slawson, Jr. (1998). "Valuing Prepayment and Default in a Fixed-rate Mortgage: A Bivariate Binomial Options Pricing Technique," *Real Estate Economics* 26, 431-468.

Hull, John C. (2000). *Options, Futures & Other Derivatives*, Fourth Edition, NJ: Prentice Hall.

IRS. (2002). *Publication 936, Home Mortgage Interest Deduction*. Available: http://www.irs.gov/pub/irs-pdf/p936.pdf.

Jud, G. Donald, and Daniel T. Winkler. (2002). "The Dynamics of Metropolitan Housing Prices," *The Journal of Real Estate Research* 23, 29-45.

LeSage, James P. (1999). Applied Econometrics Using MATLAB.

Mattey, Joe, and Nancy Wallace. (2001). "Housing Price Cycles and Prepayment Rates of U.S. Mortgage Pools," *The Journal of Real Estate Finance and Economics* 23, 161-184.

Mattey, Joe, and Nancy Wallace. (1998). "Housing Prices and the (In)stability of Mortgage Prepayment Models: Evidence from California," FRBSF Working Papers, No. 98-05, 1-49.

Patruno, Gregg N. (1994). "Mortgage Prepayments: A New Model for a New Era," *The Journal of Fixed Income* 4, 42-56

Peristiani, Stavros, Paul Bennett, Gordon Monsen, Richard Peach, and Jonathan Raiff. (1997a). "Credit, Equity, and Mortgage Refinancings," *Economic Policy Review* 3, 83-99.

Peristiani, Stavros, Paul Bennett, Gordon Monsen, Richard Peach, and Jonathan Raiff. (1997b). "Effects of Household Creditworthiness on Mortgage Refinancings," *The Journal of Fixed Income* 7, 7-21.

QickenLoans.com. 2002. Available: http://quickenloans.quicken.com/lpcontent/CnUtPage/ql/content_3col/refi_intro.en.html.

Richard, Scott F., and Richard Roll. (1989). "Prepayments on Fixed-rate Mortgage-backed Securities," *The Journal of Portfolio Management* 15, 73-82.

Siegel, Jeremy J. (1984). "The Mortgage Refinancing Decision," *Housing Finance Review* 3, 91-97.

VanderHoff, James. (1996). "Adjustable and Fixed Rate Mortgage Termination, Option Value, and Local Market Conditions: An Empirical Analysis," *Real Estate Economics* 24, 379-393.

Yang, T.L. Tyler, and Brian A. Maris. (1993). "Mortgage Refinancing with Asymmetric Information," *Journal of the American Real Estate and Urban Economics Association* 21, 481-510.

CHAPTER 4

MONITORING AND DIVIDEND POLICY OF REITS UNDER ASYMMETRIC INFORMATION

4.1. Introduction

Under the assumption that the capital market is perfect and investment decisions are independent, dividends are irrelevant to a firm's value because investors could create their own dividends by selling or borrowing against their portfolios. Despite this, we observe that a firm pays out dividends and its stock price changes upon its dividend announcements.

Considering that mangers have information unavailable to external market participants, finance researchers have proposed agency costs and signaling to explain dividend policy. The two competing explanations both receive empirical support, in particular, in the industry of real estate investment trusts (REITs). Wang, Erickson, and Gau (1993; WEG hereafter) support agency-cost theories, while Bradley, Capozza, and Seguin (1998; BCS hereafter) support signaling theories. However, neither of the two studies empirically takes into account the mandatory 95% (90% beginning in 2001) payout requirement of REIT taxable income.[1] In addition, neither one considers Easterbrook's (1984) monitoring rationale for paying dividends. By introducing proxies for taxable income and monitoring, we provide evidence that agency-cost explanations dominate signaling explanations for dividends policy of inefficiently monitored REITs. This evidence confirms Easterbrook's (1984) rationale and differs from the evidence provided by previous studies attempting distinguish signaling and agency cost explanations.

Filbeck and Mullineaux (1999 and 1993) as well as BCS (1998) have attempted to distinguish between the two competing dividend explanations: signaling and agency costs.

[1] REITs are required to pay out at 95% (90% beginning in 2001) of their taxable income in form of dividend to retain their REIT status (Internal Revenue Service, 2002).

Examining bank holding companies that are highly monitored by regulators, Filbeck and Mullineaux (1993) find that their results of stock price reaction on dividend announcements support signaling explanations, and Filbeck and Mullineaux (1999) find that agency costs are almost irrelevant to dividend paying of bank holding companies. Taking together their two studies, Filbeck and Mullineaux (1993 and 1999) provide evidence that signaling explanations dominate agency-cost explanations for dividends policy of effectively monitored firms.

Examining REITs, BCS (1998) find a negative relationship between cash flow volatility and dividend distribution. They conclude that their evidence supports signaling theories but is inconsistent with agency-cost theories. In other words, BCS's (1998) evidence shows that signaling explanations dominate agency-cost explanations for dividends policy of a pool of firms that may be or may not be effectively monitored. However, neither of Filbeck and Mullineaux (1993 and 1999) nor BCS (1998) examined the dominance between signaling explanations and agency-cost explanations for firms that are not effectively monitored.

We argue that the dominance between the two dividend policy explanations provided in the existing literature may be reversed for firms that are not effectively monitored based on two reasons. First, Easterbrook's (1984) rationale of substitution among agency-cost control devices suggests agency-cost explanations are valid only for firms that are not effectively monitored. This rationale implies that samples in both Filbeck and Mullineaux (1993 and 1999) and BCS (1998) are biased against agency-cost explanations and in favor of signaling explanations.

Second, BCS's (1998) signaling model implicitly assumes that managers are maximizing shareholders' wealth and the market knows it. Under this assumption, the market can infer a firm's private information from its manager' actions. Without this assumption, the market may not be able to infer information. Thus, the manager may not be able to communicate credible

signals to the market. For instance, mangers with compensation closely tied with current firm value have incentives to send false signals with a high level of dividends. Managers working in firms that are not effectively monitored are more likely to maximize their own wealth instead of shareholders wealth than managers in effectively monitored firms. Thus, firms that are not effectively monitored are less likely to signal credibly. This reasoning implies the dividend signaling model may be dominated by agency-cost explanations in explaining dividend policy for firms that are not effectively monitored.

Different from BCS (1998), we do not rely solely on cash flow volatility to distinguish the dominance between signaling explanations and agency-cost explanations. Following Noronha, Shome, and Morgan (1996, NSM hereafter), we stratify firms into "non-monitored firms" and "monitored firms" to take into account the Easterbrook's (1984) rationale. Adopting the encompassing principle, we then artificially nest the models for the two explanations to examine their dominance for non-monitored firms and monitored firms separately (Greene, 200).

The next section presents our reason for using measures other than cash flow volatility alone to distinguish signaling explanations and agency-cost explanations. The third section describes the sample and empirical methodology in detail. The fourth section presents the empirical results. The last section contains conclusions.

4.2 Cash Flow Volatility and Dividends

In this section, we present our reason for not using cash flow volatility alone to distinguish signaling explanations and agency-cost explanations. This section is comprised of two parts. The first part discusses the negative relationship between dividend distribution and cash flow volatility in the dividend-signaling framework. On the other hand, the second part discusses the ambiguous relation in the agency-cost framework. The discussion concludes that agency-cost

explanations of dividend do not predict a determinate direction of the relationship between dividend level and cash flow volatility. Therefore we differ from BCS (1998) and do not use the sign of the relationship alone to determine the dominance between the signaling explanations and the agency-cost explanations.

4.2.1 Cash Flow Volatility in the Dividend-Signaling Framework

At least three dividend signaling papers discuss the relation between dividend distribution and cash flow volatility: Eades (1982), Kale and Noe (1990), and BCS (1998).[2] All three dividend-signaling papers assume either explicitly or implicitly that firms are the managers are perfectly aligned with current shareholders. In other words, the managers are maximizing their firm values that are equivalent to their current shareholders' wealth.[3] In this case, we can divide agents into two groups: firms (managers and current shareholders) and the market (remaining agents). Firms have information unavailable to the market. Firms signal the asymmetric information to the market with dividend distribution. Nevertheless, the asymmetric information is different in the three papers. Specifically, dividends signal the expected future cash flow both in Eades' (1982) paper and BCS's (1998) paper, and the cash flow volatility in Kale and Noe's (1990) paper.

In Eades' (1982) and BCS's (1998) models, the asymmetric information is expected future cash flow. However, both firms and the market know the variance of the future cash flow. Firms signal the expected cash flow to the market to maximize the current firm value by distributing dividend. The market infers the expected cash flow from the promised dividend. Firms incur a

[2] Actually Eades (1982) stated the relation between dividend level and variance of liquidation value of a firm in their one-period signaling model. Since the liquidation value is the only cash flow at the end of one period, the relation can be viewed as one between dividend level and cash flow volatility. Kale and Noe (1990) shared the same interpretation here.

[3] Current shareholder's wealth is not necessarily equal to current market price of a firm under asymmetric information.

market-imposed penalty when realized future cash flow is short of the promised dividend. The market-imposed penalty is the signaling cost that increases with the degree of the shortfall between realized future cash flow and the promised dividend. Increases in the expected cash flow lower the expected cost of signaling by reducing the expected shortfall, and thus raise the level of dividend required for credible signaling. This implies a positive relationship between dividend level and expected cash flow. On the other hand, firms with known higher cash flow volatility are more likely to have a larger shortfall with a given level of dividend, and thus need a lower level of dividend to send a credible signal. This implies a negative relationship between dividend level and cash flow volatility.

Different from Eades' (1982) and BCS's (1998) models, both firms and the market know the expected future cash flow in Kale and Noe's (1990) model. Now the asymmetric information is volatility of expected future cash flow. So the information content that firms want to signal and the market wants to infer from dividend is cash flow volatility. Nevertheless, the intuitions behind the relationship between dividend and cash flow volatility, as well as expected future cash flows, are very similar to those in Eades' (1982) model. Firms have to obtain external financing to meet the shortfall between realized future cash flow and promised dividend. The external financing cost which increases with the degree of the shortfall is the signaling cost. Increases in cash flow volatility raise the expected external financing cost associated with a given level of dividend, which lowers the level of dividend necessary for credible signaling.[4] On the other hand, firms with known higher expected future cash flows are less likely to incur the signaling cost, and thus require a higher level of dividend to signal credibly.

[4] Actually Kale and Noe's (1990) separate cash flow volatility into systematic risk and unsystematic risk. They predict a negative relation between dividend level and unsystematic risk. But they do not a deterministic prediction about the sign of the relation between dividend level and systematic risk. Nevertheless the existing empirical evidence also suggests a negative relation between dividend level and systematic risk (see BCS, 1998).

To sum up, the three dividend signaling models all predict a negative (positive) relationship between dividend level and cash flow volatility (expected future cash flow). In addition to the theoretical prediction, Eades (1982) and BCS (1998) also provide empirical evidence supporting the negative (positive) relationship. Therefore, we conclude that the relationship between dividend and cash flow volatility is negative in the signaling framework.

4.2.2 Cash Flow Volatility in the Agency-Cost framework

At least two studies discuss the relation between cash flow volatility and dividend distribution in the agency-cost framework: Rozeff (1982) and BCS (1998). Nevertheless, the two studies have different predictions about the relationship. Rozeff (1982) predicts a negative relationship between dividend level and cash flow volatility in the agency-cost framework.[5] Notice Rozeff's (1982) prediction is the same as the prediction of the three signaling models discussed before. In other words, both the signaling models of Eades (1982), Kale and Noe (1990), as well as BCS (1998), and the agency-cost model of Rozeff (1982) predict a negative relationship between dividend level and cash flow volatility. On the other hand, BCS (1998) predict a positive relationship between cash flow volatility and dividend distribution in the agency-cost framework.

Rozeff (1982) argues that cash flow volatility increases a firm's dependence on external financing given fixed investment opportunities. External financing is costly compared to internal financing. Dividend payments reduce the available amount of internal financing when needed for investment. Therefore the opportunity cost of dividends for a firm with higher cash flow volatility is higher than a firm with lower volatility. The transaction cost effect induces a negative relationship between cash flow volatility and dividend payout. One can also view this

[5] See page 254 in Rozeff (1982).

argument from under-investment risk. An increase in cash flow volatility increases under-investment risk given fixed available financing. Dividend payments reduce the available financing under a firm's control. To avoid under-investment, a firm needs to increase its amount from external financing and thus incurs extra transaction costs.

An increase in cash flow volatility, however, also increases over-investment risk. Over-investment risk increases because investors attribute deviation in cash flows to the actions of corporate management or to the factors beyond management's control with less precision. Dividend payments reduce the funds under management's discretion. Hence dividend distributions can reduce over-investment risk. From this point of view, BCS (1998) argue that a positive relationship exists between cash flow volatility and dividend payout under agency-cost explanations.[6]

Agency-cost explanations, therefore, overall do yield a determinate prediction about the relationship between cash flow volatility and dividend distribution. As a result, examining the link between cash flow volatility and dividend distribution alone cannot clearly distinguish agency cost and signaling theories of dividends. Therefore we differ from BCS (1998) and do not only use the sign of the relationship to determine the dominance between the signaling explanations and the agency-cost explanations.

4.3 The Sample and Empirical Methodology

This section consists of two parts. The first part describes our data selection and the properties of the data. The second part explains the empirical methodology used in this study.

[6] See page 556 in BCS (1998).

4.3.1 The Sample

We obtained the initial list of REITs for this study from Research Insight by searching companies with SIC code 6798. We then collected relevant annual firm-specific data for REITs from the Research Insight and the Academic Universe for the period 1987 through 1998. Dividends, taxable income, market values of assets, leverage ratios, trading volumes, numbers of common shares, returns on total assets, and numbers of common stockholders are from the Research Insight.[7] Funds from operations (FFO) come computed annually following Graham and Knight (2000) from the Research Insight files as well.[8] Other annual data comes from the Academic Universe. Real estate investment information and managers and directors' stock holdings are from 10-K reports and proxy statements. Equity, mortgage, or hybrid REITs are identified from balance sheets in 10-K reports to shareholders. All hybrid REITs and mortgage REITs are dropped. REITs in merger or liquidating processes are also dropped. This leads to a final sample consisting of 332 observations of equity REITs.

Table 4.1 contains descriptive statistics for all REITs used in this study and shows that there is significant variation in REIT dividend policy. Specifically, there are 19 (5.72%) out of 332 observations paying no cash dividend. The REITs pay $31.86 million in dividend a year on average, with a standard deviation of $38.82 million. There are 252 (75.90%) out of 332 observations paying dividend more than the mandatory 95% payout requirement. Excess dividend above the 95% distribution requirement has a mean of $7.17 million and a standard deviation of $12.17 million.

---

[7] Pretax income, taxable income in concept, reported in the Research Insight is used as a proxy for taxable income reported by REITs to the Internal Revenue Service.

[8] Graham and Knight (2000) consider FFO as a cash flow measure for REITs and define FFO as net income plus depreciation, minority interest income, and extraordinary items, and excluding gain or loss on sales of property, plant and equipment.

Table 4.1: Summary Statistics

Variable	Label	n	Mean	Std Dev	Minimum	Maximum
Dividend ($ millions)	DV	332	31.8614	38.8208	0.0000	350.1830
Excess dividend ($ millions)	EXDV	332	7.1768	12.1746	0.0000	107.7167
Dividend per share ($)	DIVPS	332	4.5050	33.1397	0.0000	540.1200
Dividend yield (%)	DY	332	9.9544	43.7967	0.0000	801.7780
Dividend payout (%)	DP	332	126.6019	246.6806	-1935.1000	1909.3700
Dividends / funds from operations (%)	DV/FFO	332	77.9544	136.0732	-1165.7822	1769.7313
Fund from operations ($ millions)	FFO	332	47.4722	66.3656	-15.1600	676.3100
Excess fund from operations ($ millions)	EXFFO	332	18.4700	31.9288	-46.0964	301.1788
Net income ($ millions)	NI	332	25.9130	36.0694	-13.0000	349.0290
Taxable income ($ millions)	TI	332	30.2857	40.7256	-11.3020	394.8750
Market value ($ millions)	MV	332	449.2393	678.6196	1.0430	7875.6000
Leverage ratio (%)	LEVER	332	43.0130	18.8130	0.0000	92.1604
States with real estate investment	STATE	332	10.2801	8.1532	1.0000	43.0000
Trading volume/ number of common shares	INCENTIVE	332	56.2238	42.5398	4.5416	518.1000
Asset growth ratio (%)	GR	332	111.8029	118.2091	-35.3320	1456.0300
ROA (%)	ROA	332	3.3249	5.0780	-55.1510	15.9738
Insider ownership (%)	INS	332	13.7684	14.7651	0.0790	77.9000
Price to book ratio (%)	PTB	332	184.3660	156.7867	11.9000	1547.8500
Number of common shareholders (thousands)	STOCK	332	6.9602	16.2976	0.0190	161.6400
Blockholders' ownership (%)	BLOCK	332	14.4367	14.9879	0.0000	77.0100

For the REITs, dividend per share on average is $4.50 with a standard deviation of $33.14. Dividend yield is 9.95% of share price on average and has a standard deviation of 43.80%. Dividend payout measured in net income before extraordinary items is 126.60% on average. Dividend payout expressed as a proportion of funds from operations (FFO) has a mean of about 78% and has a standard deviation of 136%.

4.3.2 Empirical Methodology

Our empirical methodology is in the spirit of NSM (1996). Following NSM (1996), we stratify observations into "non-monitored firms," or "monitored firms" as shown in Figure 4.1. The non-monitored firms do not have non-dividend monitoring, and the monitored firms are monitored with non-divined devices. All REITs are stratified into two groups first: one group with both below mean price-to-book (PTB) ratios and below mean outside blockholders' ownership (BLOCK), and the other group with above mean PTB ratios or above mean BLOCK.

Because the transition from being non-monitored to being monitored may be gradual, we drop observations near the means to increase the distinction between non-monitored firms and monitored firms. Specifically, we drop observations with above 80th percentile of below mean PTB ratios or with above 80th percentile of below mean block-ownerships for the first group. This procedure leaves 91 observations with both below 80th percentile of below mean PTB ratios and below 80th percentile of below mean block ownerships for non-monitored firms.[9] We drop observations with below 80th percentile of PTB ratios or with below 80th percentile of block-ownerships for the second group. This procedure leaves 123 observations for monitored firms.[10]

[9] Among the 91 observations for non-monitored firms, there are 30 (32.97%) observations paying dividend no more than the mandatory 95% payout requirement.

[10] Among the 123 observations for non-monitored firms, there are 28 (22.76%) observations paying dividend no more than the mandatory 95% payout requirement.

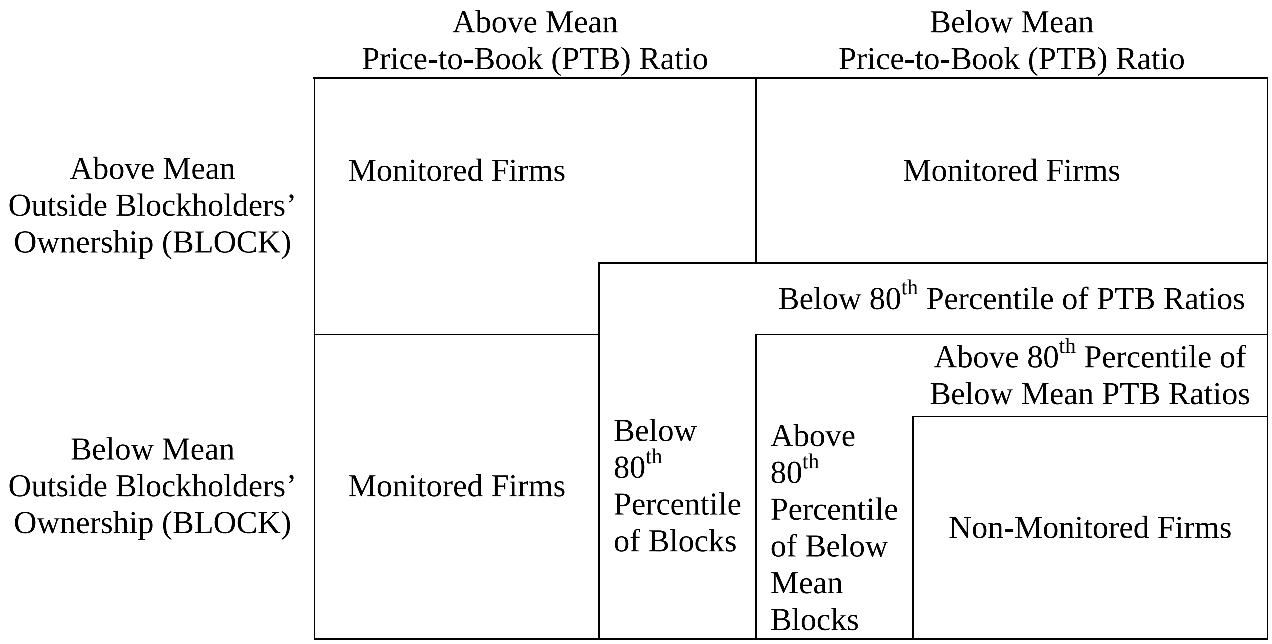

Figure 4.1: Classification of Non-Monitored Firms and Monitored Firms

As a result, we have 214 observations total in examining the dominance between signaling explanations and agency-cost explanations for REIT dividend policy.

Then, we examine dividend-signaling explanations, and agency-cost explanations: first for non-monitored firms and then for monitored firms. For non-monitored firms, we expect to find distinct supporting evidence for agency-cost explanations but not for signaling explanations. On the other hand, for monitored firms, we expect to find distinct evidence supporting signaling explanations but no distinct supporting agency-cost explanations. In other words, as shown in Figure 4.2, agency cost explanations are hypothesized dominating signaling explanations for the non-monitored firms and vice versa for the monitored firms. [11]

To examine empirically dividend-signaling explanations, we construct Equation (1) with the expected sign for each independent variable in the parentheses.

$$EXDV = F^{s}(EXFFO, Lagged\ EXFFO, MV, LEVER, STATE, INCENTIVE) \tag{1}$$

where

$EXDV$ = Annual cash dividend paid for common stocks in excess of the mandatory payout requirement (million dollars) during a fiscal year[12];

$EXFFO$ = FFO in excess of cash flow needed for the mandatory payout requirement for the current fiscal year (million dollars) $(+)$;

$lagged\ EXFFO$ = EXFFO for the previous fiscal year (million dollars) $(none)$;

[11] We do not consider all non-dividend monitoring devices in this study. Nevertheless, ignoring the other non-dividend monitoring devices bias against us to find the dominance of agency cost explanations over signaling explanations for our non-monitored firms. On the other hand, ignoring the other non-dividend monitoring devices should bias against us to find the dominance of signaling explanations over agency cost explanations for our monitored firms. As a result, our findings are stronger than they appear.

[12] The three signaling models suggest the total dividend as the dependent variable. Some studies use dividend yields as the dependent variable. This dividend yield approach is equivalent to scale total dividend with expected future cash flow in the denominator of the dependent variable or to include expected future cash flow as an independent variable (Eades, 1982).

	Non-Monitored Firms	Monitored Firms
Signaling Explanations	Expecting No Distinct Supporting Evidence	Expecting Distinct Supporting Evidence
Agency-Cost Explanations	Expecting Distinct Supporting Evidence	Expecting No Distinct Supporting Evidence
Expected Dominance	Dominant Explanations: Agency-Cost	Dominant Explanations: Signaling

Figure 4.2: Monitoring and Expected Dominance between Dividend Signaling and Agency-Cost Explanations

MV = Market value of assets (million dollars) at the end of the previous fiscal year $(+)$;

$LEVER$ = Leverage ratio (book debt-to-asset ratio) in percentage at the end of the previous fiscal year $(-)$;

$STATE$ = Geographic diversification measured by the number of states with real estate investment for an REIT at the end of the previous fiscal year $(+)$;

$INCENTIVE$ = Signaling incentive measured by the trading volume of a firm shares normalized for shares outstanding in the previous year (annual fiscal trading volume / annual common shares for basic earnings per share) $(+)$.

As concluded in Section 4.2.1, the three dividend-signaling explanations hypothesize that dividend levels are positively correlated with expected cash flows and negatively correlated with cash flow volatility. We include five proxies for expected cash flows and cash flow volatility in Equation (1). The first two independent variables, $EXFFO$ and $lagged\ EXFFO$, are used to account for expected future cash flow in excess of cash flow needed to retain a REIT status. Including the two variables separately is similar to including previous cash flow and the actual changes in cash flows in BCS (1998). In addition, including the two variables separately allows us to nest the signaling model with the agency cost model specified later on. A positive coefficient is hypothesized for $EXFFO$, and no sign is hypothesized for the coefficient for $lagged\ EXFFO$.

The next three variables, MV, $LEVER$, and $STATE$, are proxies for cash-flow volatility. The use of MV and $LEVER$ are followed from BCS (1998). MV is expected to be negatively correlated with cash flow volatility and thus is hypothesized to have a positive coefficient. This expectation arises because REITs with larger market values are associated with larger portfolios

that contain a larger number of discrete assets. If cash flows to these assets are not perfectly correlated, portfolio with larger number of discrete assets will experience lower volatility of cash flows from operations. $LEVER$ is expected to be positively correlated with cash flow volatility and thus is hypothesized to have a negative coefficient. This hypothesis is based on the positive effect of financial leverage on volatility of cash flows to shareholders, holding the cash flows from operation constant.

In addition to naïve diversification considered in MV, systematic diversification adopted by REITs reduces cash flow volatility as well. BCS (1998) construct geographic and property-type Herfindahl indices to account for systematic diversification. They find that only the geographic indices are significantly related to REIT dividends. Imitating the number of segments used in the finance diversification literature, we measure the geographic diversification with the number of states where an equity REIT has real estate investments. This index can vary from 1 for a geographically concentrated equity REIT to 50 for a well-diversified equity REIT.

In one-period models of Eades (1982) and BCS (1998), there is equivalently only one type of shareholder who has the same objective function of balancing signaling benefits and signaling costs, thus having the same incentive to signal. In the two-period model of Kale and Noe (1990), there are two types of shareholders: sellers and stayers. Sellers, who plan to sell out their shares, have an incentive to signal to have higher current market price for their shares. On the other hand, stayers, who plan to hold their shares for the longer run, have an incentive to avoid signaling costs. Facing the potential conflicts of interest between the current shareholders, mangers maximize a weighted average of the intrinsic value and the current market value of the firm.[13] When placing more weight on the current market value of the firm, managers will

[13] The intrinsic value of a firm equals the market value of the firm under full information.

distribute more dividend (Kale and Noe, 1990). To account for this effect, we include $INCENTIVE$ which is a proxy for the weight on the current market value in Equation (1), in addition to proxies for expected future cash flows and cash flow volatility.

To examine empirically dividend-signaling explanations, we construct Equation (2) with the expected sign for each independent variable in the parentheses.

$$EXDV = F^{a}\left(EXFFO, MV, LEVER, STATE, GR, ROA, STOCK, INS\right) \tag{2}$$

where

$EXFFO$ = FFO in excess of cash flow needed for the mandatory payout requirement for the current fiscal year (million dollars) $(+)$;

MV = Market value of assets (million dollars) at the end of the previous fiscal year $(none)$;

$LEVER$ = Leverage ratio (book debt-to-asset ratio) in percentage at the end of the previous fiscal year $(none)$;

$STATE$ = Geographic diversification measured by the number of states with real estate investment for an REIT at the end of the previous fiscal year $(none)$;

GR = Realized growth rate of total assets for the previous fiscal year $(-)$;

ROA = Return on total assets for the previous fiscal year $(-)$;

$STOCK$ = Outsiders' ownership dispersion measured by the number of common stockholders (thousands) at the end of previous fiscal year $(+)$[14];

[14] The 100-shareholder test for qualifying as a REIT applies to the total number of both common shareholders and preferred shareholders (Brandon, 1997).

INS = Insiders' ownership measured by the fraction of voting shares held by insiders at the end of previous fiscal year or the beginning of the current fiscal year $(-)$.

Unlike the three dividend-signaling models, mangers are not perfect agents of shareholders, because they pursue their own interests whenever they can in agency-cost models (Easterbrook, 1984). In the agency cost framework, dividend policy balances the agency costs of external equity and transaction costs of external financing (Rozeff, 1982). A positive relationship between dividend level and $EXFFO$ is hypothesized. This hypothesis arises because more cash flows give mangers more discretion to over-invest in non-positive net present value projects. As a result, shareholders demand more dividends than they would otherwise. Including the cash flow variable as an explanatory variable and dividend as the dependent variable is similar to putting dividend payout as the dependent variable in other studies. However, our approach allows us the nest the agency cost model with the signaling model specified before. As discussed in Section 4.2.2, cash flow volatility influences dividend policy in two opposing ways. Cash flow volatility increases over-investment risk but also under-investment risk (or expected costs of external financing). Therefore, unlike signaling explanations, no determinate signs are hypothesized for the next three variables, MV, $LEVER$, and $STATE$, that are proxies for cash-flow volatility.

Following Rozeff (1982) and WEG (1993), we include GR in the equation of agency-cost explanations. A negative relationship is hypothesized between dividend level and GR. Rozeff (1982) and WEG (1993) hypothesize this negative relationship because a firm experiencing or anticipating a rapid growth would tend to retain funds to minimize the frequency of raising new capital. We also include ROA in Equation (2) because WEG (1993) also hypothesize and show a negative relationship between ROA and dividend distribution. The negative relationship is

hypothesized because shareholders may feel less pressure to monitor the investment decisions of managers when the firm has a superior historical investment performance.

In addition, we also include $STOCK$ and INS, following Rozeff (1982). A positive relationship between $STOCK$ and dividend level and a negative relation between INS and dividend distribution are hypothesized. The positive relationship for $STOCK$ is hypothesized because outside shareholders with more concentration of ownership are more likely to influence insider behavior, thereby reducing agency costs and leading to a lower dividend distribution. The negative relationship for INS is hypothesized because as outside equity holders own a larger share of the equity, they will demand a higher dividend as part of the optimum-monitoring package.

Having specified the signaling model and the agency cost model, we artificially nest Equations (1) and (2) into the following equation (Greene, 2000):

$$EXDV = F\left(EXFFO, Lagged\ EXFFO, MV, LEVER, STATE, GR, ROA, STOCK, INS, INCENTIVE\right)$$

By examining the log-likelihoods and the coefficients of the variables in the three equations, we are able to distinguish the dominance between signaling explanations and agency cost explanations.

4.4 Empirical Results

Computed from the log-likelihoods of Specification 2 of Tables 4.2, 4.3, and 4.4, the chi-squared statistic for the Chow test is 71.08, which is significant at 1% level. This result indicates a structure difference between non-monitored and monitored firms and leads us to investigate the dominance between signaling and agency cost explanations for non-monitored and monitored firms separately. With this support, our discussion of empirical results consists of two parts. The

Table 4.2: Dividend Models for All Firms

Dependent variable: EXDV

Independent variable	Specification 1		2		
	Homoscedastic β	Marginal effects	Heteroscedastic β	α	Marginal effects
Constant	0.9145	0.5825	-0.6682	3.8590	-0.5372
	[0.7592]	[0.7602]	[0.6086]	[0.0001]***	[0.6096]
EXFFO (+)	0.1512	0.0963	0.3800	-0.0227	0.3236
	[0.0065]***	[0.0066]***	[0.0000]***	[0.0001]***	[0.0000]***
Lagged EXFFO	0.1899	0.1210	-0.0375	-0.0172	-0.0164
	[0.0049]***	[0.0051]***	[0.5240]	[0.0129]**	[0.7335]
MV	0.0033	0.0021	-0.0021	0.0026	-0.0038
	[0.3225]	[0.3240]	[0.5446]	[0.0000]***	[0.1426]
LEVER	-0.0630	-0.0401	0.0166	-0.0023	0.0152
	[0.1402]	[0.1417]	[0.2979]	[0.5083]	[0.2534]
STATE	0.1047	0.0667	-0.0132	-0.0100	-0.0027
	[0.3656]	[0.3663]	[0.8356]	[0.3313]	[0.9570]
GR (-)	0.0109	0.0070	0.0029	-0.0051	0.0064
	[0.0665]*	[0.0664]*	[0.0059]***	[0.0000]***	[0.0000]***
ROA (-)	0.2087	0.1330	0.0927	-0.0622	0.1240
	[0.2440]	[0.2411]	[0.1599]	[0.0156]**	[0.1240]
STOCK (+)	-0.0206	-0.0132	0.0786	-0.0004	0.0635
	[0.7966]	[0.7965]	[0.1188]	[0.9355]	[0.1218]
INS (-)	-0.1492	-0.0950	-0.0358	0.0045	-0.0323
	[0.0173]**	[0.0167]**	[0.1049]	[0.3902]	[0.0619]*
INCENTIVE (+)	-0.0077	-0.0049	0.0088	0.0181	-0.0074
	[0.7731]	[0.7732]	[0.5068]	[0.0000]***	[0.4974]
Log-Likelihood	-632.1748		-538.1187		
k	11		21		
n	214		214		

Note: 1. The variance of regression errors is assumed to be in the form $\sigma_i^2 = \exp\left(\alpha' \mathbf{x}_i\right)$ in Specification 2.

2. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 4.3: Dividend Models for Non-Monitored Firms

Dependent variable: EXDV

Independent variable	Specification 1		2			3			4		
	Homoscedastic	Marginal	Heteroscedastic		Marginal	Heteroscedastic		Marginal	Heteroscedastic		Marginal
	β	effects	β	α	effects	β	α	effects	β	α	effects
Constant	-3.8075	-2.2030	2.2088	1.2012	1.6920	0.6306	0.2796	0.5061	1.9161	1.9087	1.1778
	[0.3797]	[0.3671]	[0.3663]	[0.2493]	[0.3936]	[0.6504]	[0.0089]***	[0.6418]	[0.4152]	[0.1023]	[0.4370]
EXFFO (+)	0.1573	0.0910	0.3347	-0.0401	0.2588	0.4484	-0.0764	0.3968	0.5981	-0.0075	0.3677
	[0.2254]	[0.2225]	[0.0006]***	[0.0074]***	[0.0004]***	[0.0016]***	[0.0000]***	[0.0000]***	[0.0000]***	[0.5908]	[0.0000]***
Lagged EXFFO	0.0936	0.2225	-0.0739	-0.0378	-0.0544	0.0204	0.0487	-0.0071			
	[0.5983]	[0.5969]	[0.3661]	[0.1871]	[0.3762]	[0.8652]	[0.0084]***	[0.9392]			
MV	-0.0109	-0.0063	-0.0046	0.0043	-0.0038	0.0097	0.0050	0.0054	-0.0132	0.0044	-0.0018
	[0.2931]	[0.2895]	[0.6074]	[0.0022]***	[0.5304]	[0.2313]	[0.0000]***	[0.3957]	[0.0178]**	[0.0105]**	[0.6022]
LEVER	0.0612	0.0354	-0.0061	0.0244	-0.0061	-0.0139	0.0295	-0.0253	-0.0095	0.0507	0.0670
	[0.3679]	[0.3646]	[0.8639]	[0.0326]**	[0.8246]	[0.5574]	[0.0001]***	[0.1844]	[0.7708]	[0.0000]***	[0.0123]**
STATE	-0.2397	-0.1387	-0.1412	-0.0221	-0.1069	-0.1040	-0.0283	-0.0698	-0.1104	-0.0017	-0.0703
	[0.1431]	[0.1417]	[0.1133]	[0.3783]	[0.1095]	[0.3002]	[0.1738]	[0.3907]	[0.0474]**	[0.9418]	[0.1512]
GR (-)	-0.0214	-0.0124	0.0031	-0.0059	0.0028				0.0133	-0.0182	-0.0180
	[0.2203]	[0.2242]	[0.5527]	[0.0194]***	[0.4306]				[0.0822]*	[0.0000]***	[0.0000]***
ROA (-)	0.3871	0.2239	-0.0135	-0.0952	-0.0046				0.0771	0.0181	0.0734
	[0.2894]	[0.2764]	[0.9521]	[0.2022]	[0.9778]				[0.6348]	[0.7832]	[0.5541]
STOCK (+)	0.4120	0.2384	0.1801	0.0065	0.1376				0.1271	-0.0147	0.0570
	[0.0002]***	[0.0003]***	[0.0082]***	[0.5681]	[0.0034]***				[0.0047]***	[0.2222]	[0.0645]*
INS (-)	-0.0554	-0.0320	-0.0701	-0.0180	-0.0526				-0.1571	0.0000	-0.0966
	[0.4531]	[0.4527]	[0.0298]**	[0.2546]	[0.0411]**				[0.0000]***	[0.9965]	[0.0000]***
INCENTIVE (+)	0.1125	0.0651	0.0198	0.0226	0.0138	-0.0039	-0.0283	-0.0150			
	[0.0049]***	[0.0059]***	[0.5276]	[0.0000]***	[0.5350]	[0.9083]	[0.0000]***	[0.5373]			
Log-Likelihood	-238.8458		-181.3314			-191.6688			-191.8703		
k	11		21			13			17		
n	91		91			91			91		

Note: 1. The variance of regression errors is assumed to be in the form $\sigma_i^2 = \exp(\alpha' \mathbf{x}_i)$ in Specifications 2, 3, and 4.

2. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

Table 4.4: Dividend Models for Monitored Firms

Dependent variable: EXDV

Independent variable	Specification 1		2			3			4		
	Homoscedastic β	Marginal effects	Heteroscedastic β	α	Marginal effects	Heteroscedastic β	α	Marginal effects	Heteroscedastic β	α	Marginal effects
Constant	4.5668	3.2102	0.7958	17.9713	0.6992	1.7167	3.9382	1.3206	1.2158	15.7341	0.9997
	[0.3652]	[0.3705]	[0.7881]	[0.0316]**	[0.7891]	[0.5167]	[0.0002]***	[0.5153]	[0.5844]	[0.0178]**	[0.5801]
EXFFO (+)	0.1851	0.1301	0.5583	-0.0029	0.4972	0.2173	-0.0057	0.1682	0.2912	-0.0051	0.2395
	[0.0022]***	[0.0024]***	[0.0000]***	[0.7359]	[0.0000]***	[0.0172]**	[0.3531]	[0.0154]**	[0.0000]***	[0.2887]	[0.0002]***
Lagged EXFFO	0.1735	0.1220	-0.1020	-0.0328	-0.0127	0.0624	-0.0071	0.0493			
	[0.0177]**	[0.0183]**	[0.1932]	[0.0005]***	[0.8667]	[0.4210]	[0.2626]	[0.4022]			
MV	0.0041	0.0029	-0.0067	0.0020	-0.0105	-0.0017	0.0016	-0.0016	-0.0017	0.0018	-0.0034
	[0.2603]	[0.2621]	[0.1149]	[0.0000]***	[0.0075]***	[0.7105]	[0.0000]***	[0.6138]	[0.6607]	[0.0035]***	[0.2596]
LEVER	-0.0873	-0.0614	-0.0230	-0.0189	0.0242	-0.0166	-0.0061	-0.0116	-0.0167	-0.0175	0.0068
	[0.1313]	[0.1327]	[0.3063]	[0.0000]***	[0.2833]	[0.5933]	[0.1170]	[0.6264]	[0.4291]	[0.0038]***	[0.7252]
STATE	0.2911	0.2046	0.0175	0.0228	-0.0381	0.0147	-0.0119	0.0136	0.0156	0.0009	0.0118
	[0.0444]**	[0.0447]**	[0.8341]	[0.1739]	[0.6528]	[0.8932]	[0.3102]	[0.8698]	[0.8360]	[0.9450]	[0.8536]
GR (-)	0.0186	0.0130	0.0015	-0.0086	0.0215				0.0031	-0.0022	0.0051
	[0.0037]***	[0.0037]***	[0.5913]	[0.0000]***	[0.0000]***				[0.1901]	[0.0038]***	[0.0570]*
ROA (-)	0.1577	0.0037	0.0648	-0.1516	0.4126				0.1138	-0.1327	0.2487
	[0.6188]	[0.6179]	[0.6416]	[0.0006]***	[0.0133]**				[0.1577]	[0.0000]***	[0.0089]***
STOCK (+)	-0.4143	-0.2912	0.0765	-0.0309	0.1397				-0.0401	-0.0186	-0.0112
	[0.0010]***	[0.0010]***	[0.0817]*	[0.0980]*	[0.0024]***				[0.4325]	[0.1494]	[0.8162]
INS (-)	-0.2509	-0.1764	0.0228	0.0323	-0.0559				-0.0453	0.0158	-0.0558
	[0.0098]***	[0.0093]***	[0.6276]	[0.0002]***	[0.1791]				[0.3332]	[0.1398]	[0.1547]
INCENTIVE (+)	-0.0700	-0.0492	-0.0012	0.0121	-0.0293	-0.0057	0.0067	-0.0056			
	[0.0739]*	[0.0758]*	[0.9383]	[0.0012]***	[0.1137]	[0.8661]	[0.0182]**	[0.8237]			
Log-Likelihood	-374.8726		-321.2470			-349.3432			-327.8927		
k	11		21			13			17		
n	123		123			123			123		

Note: 1. The variance of regression errors is assumed to be in the form $\sigma_i^2 = \exp\left(\alpha' \mathbf{x}_i\right)$ in Specifications 2, 3, and 4.

2. * significant at 10% level; ** significant at 5% level; *** significant at 1% level.

first part presents the empirical results for non-monitored firms. The second part discusses the results for monitored firms.

4.4.1 Empirical Results for Non-monitored Firms

Table 4.3 provides the empirical estimation results for non-monitored firms. Likelihood statistics suggest Specification 2 for analysis. From this regression, we are able to tell the dominance between the signaling and agency cost models from the marginal effects (slopes, hereafter). Specifically, $STOCK$ has a significant and positive slope as hypothesized. This slope supports the hypothesis that outside shareholders with more concentration of ownership are more likely to influence insider behavior, thereby reducing agency costs and leading to a lower dividend distribution. As hypothesized, INS has a significant and negative slope consistent with agency-cost hypothesis. This slope is consistent with the hypothesis that outside equity holders own a larger share of the equity, they will demand a higher dividend as part of the optimum monitoring package. $EXFFO$ has a significant and positive slope. The slope supports the hypothesis that more cash flows give mangers more discretion to over-invest in non-positive net present value projects in agency-cost explanations.

Although the slope of $EXFFO$ also is consistent with the hypothesis that expected cash flow increases dividend distribution, the slope cannot distinguish signaling explanations from agency cost explanations. Without this $EXFFO$ result, we have no other supporting evidence for signaling explanations. Specifically, the three variables for expected cash flow volatility all have insignificant and negative slopes. The slope of $LEVER$ has a hypothesized sign, but the slopes of MV and $STATE$ have signs opposite to the hypotheses in signaling explanations.

To sum up, we have evidence supporting agency cost explanations but no distinct evidence supporting signaling explanations for non-monitored firms. Therefore, the evidence indicates that

agency cost explanations dominate signaling explanations when firms are not effectively monitored. As hypothesized, this dominant relationship is reverse to the relationship documented in the existing studies that do not take into account Easterbrook's (1984) rationale or the mandatory payout requirement for REITs.

4.4.2 Empirical Results for Monitored Firms

Table 4.4 exhibits the empirical estimation results for monitored firms. Likelihood statistics again suggest Specification 2 for analysis. This specification provides mixed evidence for both signaling and agency cost explanations. Although $EXFFO$ has a significant and positive slope consistent with both signaling and agency cost explanations, MV has a significant and negative slope inconsistent with signaling explanations. This slope suggests that firms with lower cash flow volatility pay out more dividend than firms with higher cash flow volatility. Nevertheless this slope is consistent with agency cost explanations that do not have a determinate prediction for the relationship between cash flow volatility and dividend distribution.

Again, $STOCK$ has a significant and positive slope as hypothesized in agency cost explanations. However, both GR and ROA have significant and positive slopes. The positive slope of GR implies that a firm experiencing or anticipating a rapid growth would tend to distribute more dividend. The positive slope of ROA implies that shareholders feel more pressure to monitor the investment decisions of managers when the firm has a superior historical investment performance. These two slopes are inconsistent with agency cost explanations.

To summarize, we have mixed evidence for both signaling and agency cost explanations for monitored firms. There are two possible explanations for having the mixed evidence here. One potential reason is that we do not have grouped monitored firms sufficiently well so that the firms may still have sufficiently severe agency problem. The other likely reason is that other

ways of maximizing shareholders' wealth may provide more net benefit than signaling expected cash flows or cash flow volatility. One such way is to balance corporate tax savings and shareholders' personal tax costs when REIT manager distribute dividends (Lee and Kau, 1987).

4.5 Conclusions

The REIT empirical literature offers two competing theories for the level of dividend payouts under asymmetric information. WEG (1993) provide evidence supporting agency-cost explanations for REIT dividend policy. On the other hand, BCS (1998) have evidence supporting signaling explanations and argue that signaling explanations dominate agency-cost explanations with evidence from cash flow volatility.

Neither WEG (1993) nor BCS (1998) consider either the mandatory 95% payout requirement or Easterbrook's (1984) monitoring rationale for paying dividends. By introducing proxies for taxable income and monitoring, we provide evidence different from the evidence provided by BCS (1998). In contrast to BCS (1998), our evidence suggests agency-cost explanations dominate signaling explanations for dividends policy of inefficiently monitored REITs.

4.6 References

Bradley, Michael, Dennis R. Capozza, and Paul J. Seguin. (1998). "Dividend policy and cash-flow uncertainty," *Real Estate Economics* 26, 555-580.

Brandon, David L. (1997). "Federal Taxation of Real Estate Investment Trusts," in Garrigan and Parsons (Ed.), *Real Estate Investment Trusts: Structure, Analysis, and Strategy*, McGraw-Hill, New York, 83-130.

Eades, Kenneth M. (1982). "Empirical Evidence on Dividends as a Signal of Firm Value," *The Journal of Financial and Quantitative Analysis* 17, 471-500.

Easterbrook, Frank H. (1984). "Two agency-cost explanations of dividends," *The American Economic Review* 74, 650-659.

Filbeck, Greg, and Donald J. Mullineaux. (1993). “Regulatory monitoring and the impact of banking holding company dividend changes on equity returns,” *The Financial Review* 28, 403-415.

Filbeck, Greg, and Donald J. Mullineaux. (1999). “Agency costs and dividend payments: the case of bank holding companies,” *The Quarterly Review of Economics and Finance* 39, 409-418.

Graham, Carol M., and John R. Knight. (2000). “Cash flows vs. earnings in the valuation of equity REITs,” *Journal of Real Estate Portfolio Management* 6, 17-25.

Greene, William H. (2000). *Econometric Analysis*, fourth edition, New Jersey.

Internal Revenue Service. (2002). Instructions for Form 1120-REIT. Available: http://www.irs.gov/pub/irs-pdf/i1120rei.pdf.

Kale, Jayant R. and Thomas H. Noe. (1990). “Dividends, uncertainty and underwriting costs under asymmetric information,” *The Journal of Financial Research* 13, 265-277.

Lee, Cheng F. and James B Kau. (1987). “Dividend Payment Behavior and Dividend Policy on REITs,” *Quarterly Review of Economics and Business* 27, 6-21.

Noronha, Gregory M., Dilip K. Shome, and George E. Morgan. (1996). “The monitoring rationale for dividends and the interaction of capital structure and dividend decisions,” *Journal of Banking and Finance* 20, 439-454.

Rozeff, Michael S. (1982). “Growth, beta and agency costs as determinants of dividend payout ratios,” *The Journal of Financial Research* 5, 249-259.

Wang, Ko, John Erickson, and George W. Gau. (1993). “Dividend polices and dividend effects for real estate investment trusts,” *Journal of the American Real Estate and Urban Economics Association* 21, 185-201.

VITA

Ming-Long Lee received a Bachelor of Business Administration degree in International Trade from the Soochow University in Taiwan in 1990. In 1993, after his two-year military service, he matriculated in the University of South Australia in Adelaide, Australia, and received a Master of Business degree in property in 1995. After working for a real estate appraisal firm and teaching at a business college in Taiwan, he entered the doctoral program in business administration (Finance) at Louisiana State University in 1998. He has articles published in the Journal of the Land Bank of Taiwan and has an article to be published in a forthcoming issue of the Journal of Property Investment and Finance. He will complete his doctoral degree in the summer of 2002.