Essays in Real Estate Finance

by

Sheharyar Javaid Bokhari

B.A., Mathematics and Economics, Macalester College (2005)
M.A, Economics, Tufts University (2006)

Submitted to the Department of Urban Studies and Planning
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Real Estate and Urban Studies

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

Author ....
Department of Urban Studies and Planning
May 22, 2012

Certified by ....
William Wheaton
Professor of Economics
Thesis Supervisor

Accepted by ....
JoAnn Carmin
Chairman, Department Committee on Graduate Theses

Essays in Real Estate Finance

by

Sheharyar Javaid Bokhari

Submitted to the Department of Urban Studies and Planning
on May 22, 2012, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Real Estate and Urban Studies

Abstract

This dissertation is a collection of three essays in real estate finance. In the first essay, we observe that between 1985 and 2007, the share of household mortgage debt as a proportion of the total value of housing in the US increased substantially from 30% to an all-time high of 50%. With the decline in house prices, these high levels of leverage increased the propensity at which households defaulted. We examine household decisions on mortgage leverage using new extensive loan-level data from Fannie Mae over the sample period 1986 to 2010. We conceptualize a market for leverage per se and develop a theory of leverage demand-and-supply. Empirically, we estimate an interest rate elasticity of leverage demand of -0.37 or, equivalently, a movement along the demand curve from an r-LTV pair of (10%, 72%) to that of (5%, 85%). We find that leverage demand was cyclical and responsive to economic events but without a general trend. By contrast, leverage supply shifts in the form of lower mortgage interest rates were concurrently associated with higher average loan-to-value ratios. We find that in MSAs with higher house prices, households borrowed more and bought equally more expensive houses. That left leverage unchanged but raised households' risk of illiquidity by increasing their loan-to-income ratios. In MSAs with high house price volatility, we find that both leverage demand and supply were lower. We also identify that younger, poorer and less credit-worthy borrowers demand more leverage than their counterparts.

In the second essay, co-authored with David Geltner, we document that loss aversion behavior plays a major role in the pricing of commercial properties, and it varies both across the type of market participants and across the cycle. We find that sophisticated and more experienced investors are at least as loss averse as their counterparts and that loss aversion operated most strongly during the cycle peak in 2007. We also document a possible anchoring effect of the asking price in influencing buyer valuation and subsequent transaction price. We demonstrate the importance of behavioral phenomena in constructing hedonic price indices, and we find that the impact of loss aversion is attenuated at the aggregate market level. This suggests that the pricing and volume cycle during 2001 - 2009 was little affected by loss aversion.

In the third essay, also co-authored with David Geltner, we present a technique to address the problem of data scarcity in the construction of high-frequency real estate price indexes. We introduce a two-stage frequency conversion procedure, by first estimating lower-frequency indexes staggered in time, and then applying a generalized inverse estimator to convert from lower to higher frequency return series. The two-stage procedure can improve the accuracy of high-frequency indexes in scarce data environments. The method is demonstrated and analyzed by application to empirical commercial property repeat-sales data.

Thesis Supervisor: William Wheaton
Title: Professor of Economics

Acknowledgments

To begin at the beginning: I am grateful to my parents for all that I have and to them I dedicate this dissertation. From my father, Jawaid Bokhari, I took a passion for self-learning and from my late mother, Razia Bokhari, I learnt to be inquisitive about all things intellectual. I am also fortunate to have extremely supportive and caring siblings. Despite being so far away, they have been close advisors in all matters.

Obtaining a higher education would not have been possible, if not for the generosity of Macalester College, Tufts University and MIT. I am greatly in their debt for keeping me debt free. I did leverage my access to excellent teachers at these institutions and I'm thankful for a very enjoyable learning experience.

It's a pleasure to thank and acknowledge my dissertation committee of William Wheaton, David Geltner and Walter Torous for their superb mentoring, advice, encouragement and discussions. Much of my interest in real estate and exposure to industry came while working under David Geltner. Bill Wheaton has been great in nudging me to push beyond my boundaries. I am also grateful to both Bill and David for keeping as their teaching assistant over the past couple of years. Walt Torous, while being at UCLA, was highly accessible and gave pragmatic advice and feedback on this work.

I also enjoyed working and socializing with my classmates. This thesis benefitted particularly from the thoughts of Lauren Lambie-Hanson.

Perhaps the best part of my PhD experience was marrying my wife, Sania, who has given all the advice, love, time and support a person could ever ask for.

Finally, I acknowledge financial support for Chapter 3 by the Real Estate Research Institute (RERI) and I thank Real Capital Analytics (RCA) for providing the data used in Chapters 3 and 4. I also thank Fannie Mae for providing their data for Chapter 1.[1]

[1]Nothing in this thesis should be construed as representing the opinions or views of RERI, RCA or Fannie Mae.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The goal of this thesis is to contribute to our understanding of three current issues in the area of real estate finance. The three topics are unified by the underlying microeconomic inquiry of individual decisions and their relation to the performance of real estate markets. The three areas explored in this thesis are:

- Borrower decision on mortgage leverage, it's relationship to policy and to performance of real estate markets.
- Psychological biases in the pricing of real estate and their impact on both individual and market performance.
- Measuring performance of real estate markets when transactions data is scarce.

In the first essay, we make the observation that there has been a substantial increase in household mortgage leverage in the US over the past 25 years. To explain this trend, we develop a theory of how households make decisions on leverage and empirically test it's implications on a large longitudinal dataset provided by Fannie Mae. Our finding is that the increase in leverage can be primarily explained by increases or shifts in the supply of leverage. We document that increases in the national conforming loan limit allowed borrowers to borrow cheaply and to lever more. We also show how leverage varied across markets with different house prices and/or volatilities. Finally, we find that younger, poorer and less credit-worthy borrowers were leveraged more than their counterparts.

In the second essay, we ask whether pricing in real estate markets is affected by psychological biases of buyers and sellers? Using data on commercial real estate, we find strong empirical evidence that real estate transactions (listing prices, transaction prices and time on the market) are affected by both loss aversion on the part of sellers and anchoring on the part of buyers. We show that controlling for these behavioral phenomenon can greatly improve the construction of traditional measures of performance such as a hedonic price index. We also document how loss aversion varied across the real estate cycle and by type of investor.

In the third essay, we recognize the importance of measuring the performance of real estate markets accurately and on a frequent basis. The need for such series is underscored by their demand by both academic research (such as their use in the previous two essays) and by industry participants seeking frequent information on where the market is moving. The challenge in constructing an accurate, high-frequency, localized price index is the thin-ing of data on transactions as one zooms into a local area such as an MSA or a neighborhood. In this paper, we propose a new technique for constructing such price indexes in a scarce data environment and illustrate it's effectiveness using data on commercial real estate.

It should be noted that while the cases examined in each paper are separate, the actual theories and methods developed in these papers are generalizable to both housing and commercial real estate. For instance, one could examine the leverage decisions of commercial real estate investors, where aspects of our theory such as the effects of asset prices on leverage can be empirically tested. Similarly, loss aversion behavior has been documented in the housing market (Genesove and Mayer (2001)) and the technique developed in the third essay is supplemental to existing hedonic and repeat sales techniques used in housing research.

Chapter 2

Why did Household Mortgage Leverage Rise from the mid-1980's until the Great Recession?

2.1 Introduction

In the United States, household mortgage leverage increased dramatically in the run-up to the Great Recession. Figure 2-1 shows that between 1985 and 2007, the share of household mortgage debt as a proportion of the total value of housing in the US increased substantially from 30% to 50%. With the decline in house prices and subsequently slow de-leveraging, that share further increased to 60% by 2010. These high levels of mortgage leverage increased the propensity at which households defaulted on their mortgages and there is evidence that leverage was a primary driver of the recession (Mian and Sufi (2009), Mian and Sufi (2011)). In fact, it is a characteristic of highly leveraged economies that they seldom avoid a financial crisis (Reinhart and Rogoff (2008)). Furthermore, leverage also plays an important role in partly determining asset prices (Geanakoplos (2009), Lamont and Stein (1999)). As leverage goes up and down, asset prices also go up and down and that is damaging to the economy. Given the crucial role of leverage in the economy, it is imperative that we understand how households determine an optimal level of mortgage leverage. Thus, our objective in this article is two-folds. First, we seek to understand why household mortgage leverage rose so dramatically? Is it leverage demand or leverage supply that primarily lead to an increase in leverage? Second, and to enable us to empirically answer the research question, we conceptualize a market for leverage per se and develop a theory of leverage demand-and-supply.

In our theory of leverage, lenders and households endogenously choose an LTV ratio in a competitive market and under the possibility of default that depends on future house prices. The model is set in the tradition of optimal LTV contracts under asymmetric information (Brueckner (2000), Harrison et al. (2004)). While earlier studies have focused on the role of asymmetric information on equilibrium leverage, our paper is different in that it derives separately, an optimal leverage, for lenders and borrowers without invoking asymmetric information. The predictions of the models are intuitive. When house prices are volatile, lenders demand more collateral and rise-averse borrowers demand less leverage. This is because a more volatile house price distribution increases the risk of borrower default, which adversely affects both expected profits and collateral payoffs. On the other hand, higher average prices (holding all else constant) positively affect future payoffs and thus increase both leverage demand and supply. The model also predicts that the poorer or more impatient the borrowers are, the higher the leverage they demand at any given interest rate.

The theoretical model yields structural demand and supply equations to test econometrically. We identify the demand curve by viewing exogenous changes in the national conforming loan limit as supply shifts of leverage (e.g. loans that are jumbo become cheaper to finance over time). We estimate the demand equation and reduced forms of equilibrium interest rates and loan-to-value ratios using extensive loan-level data from Fannie Mae over the sample period 1986 to 2010. We estimate an interest rate elasticity of demand of -0.37, which implies that if the note interest rate dropped from 10% to 5% (see historical drop shown in Figure 2-2), then from an initial LTV ratio of 72% (avg. LTV in 1986), leverage demand would rise by

18.5% to an LTV ratio of 85%. Furthermore, we find that (holding all else constant) leverage demand has historically been cyclical and without any general upward trend. It rises and falls in concurrence with economic conditions, which suggests that it had a limited role in increasing historical household leverage.

We find two pieces of compelling evidence in favor of an increasing leverage supply hypothesis. First, we estimate that a doubling of the national conforming loan limit (CLL) would decrease the average note interest rate by 4%. Holding the demand curve constant, this implies that LTV would rise by 12.5 percentage points, which is substantial, given that the CLL has been increased quite a bit over the past two decades.

Second, the reduced form estimates of equilibrium interest rates and LTV ratios show that note rates have generally fallen while LTV ratios have concomitantly risen over the sample period. Given that demand is found to be cyclical in nature and that outward shifts in demand would instead cause interest rates to rise over time, this finding strongly suggests that leverage supply has been the primary driver of increases in household leverage. For e.g., in 2005, when leverage demand was about 2.8 percentage points higher than it's 1986 level, equilibrium leverage was 13.4 percentage points higher. Thus, a rough estimate would be that leverage supply was 10.6 percentage points higher than it's 1986 level. In the aftermath of the crisis, we find that while leverage demand had collapsed to it's 1980's levels, leverage supply was still high, possibly reflecting the crucial role played by the GSEs in supplying credit when the rest of the market was holding back.

Our results on house prices reveal that greater house prices lead borrowers to not only borrow more but to also buy equally more expensive houses. Although this kept leverage virtually unchanged, it raised households' risk exposure. This is because, controlling for income, a high loan amount implies a higher loan-to-income ratio. This amounts to a greater debt service and exposure to greater risk of illiquidity in the future. We also find that in markets with greater house price volatility, both borrowers and lenders contracted at lower LTV ratios, which is consistent with theory.

Our article is related to the research on mortgage contract choice and demand for mortgage debt (Follain (1990), Jones (1993), Brueckner (1994), Follain and Dunsky (1997), Ling and McGill (1998), Hendershott et al. (2002), Hendershott and LaFayette (1997), Leece (2006), Elliehausen (2010))). We complement this research by using a long historical dataset and go beyond debt demand to look at a market for leverage per se. In particular, the demand-and-supply framework is unique in that it helps us in isolating the determinants of (changes in) leverage.

The rest of the paper proceeds as follows. The next section describes the data and certain stylized facts about leverage. Section 3 presents the theory of leverage. Section 4 outlines the empirical strategy and presents the results, and Section 5 concludes.

2.2 Data and Stylized Facts

Our data are a random sample of single-family home mortgages originated in the U.S. over the period 1986 to 2010 and purchased by Fannie Mae. The raw sample in each year includes approximately 120,000 observations with equal-sized shares of purchase mortgages versus refinance mortgages. For each mortgage we have data summarizing the characteristics of the loan and the underlying property as well as information on the borrower.

The loan-to-value (LTV) ratio is lender submitted and defined as the ratio of the loan amount to the lesser of the sale price or the appraised value of the property. For second mortgages, a combined LTV ratio is calculated using the sum of the current unpaid principal balances of the first and second mortgages. Figures 2-3 and 2-4 present empirical cumulative density functions of LTV ratios for purchases and refinances, respectively. For purchases, we see that close to 40% of the data contains LTV ratios below 80%. There is considerable bunching (over 20% of the data) at the 80% LTV ratio, the threshold beyond which Private Mortgage Insurance (PMI) is required for all conforming loans. We see similar bunching at the 90%, 95% and 100% LTV ratios. Some of the purchase LTV ratios even exceed 100%. For refinances, the bulk of the data (close to 80%) is at LTV ratios below 80%. These distributions are suggestive of interior as well as corner solutions to the borrower's problem of choosing an optimal LTV ratio. However, it does not appear to be the case that all borrowers situate themselves at the PMI threshold.

Next, we sub-divide our sample into three separate time periods, 1986 to 1992, 1993 to 2007 and 2008 to 2010. We choose 1992 as the end point of the first period because that year marked the passage of the Federal Housing Enterprises Financial Safety and Soundness Act (FHEFSSA) which had a major impact on the activities of the Government Sponsored Entities.[1] The second, largely prosperous, period between 1993 and 2007 was highlighted by economic growth, low unemployment and an unprecedented rise in house prices. This was followed by a period marking the beginning of the Great Recession from early 2008 which continued until the end of our data in 2010. We see in Figure 2-5 that the distribution of LTV ratios for purchase mortgages during the 1986 to 1992 period was skewed towards LTV ratios below 80% with relatively few LTV ratios close to 100%. In comparison, the panel for the 1993 to 2007 period shows that the left tail of the distribution became less pronounced with the effect that mass accumulated not only at an 80% LTV ratio but also at very high LTV ratios. In the period following the recession, we see that the mass at 80% LTV rose even further but largely due to a lower density at higher LTV ratios. There was also an increase in the fraction of the data with LTV ratios less than 80%. We see a similar but less pronounced effect for refinances (see Figure 2-6).

The above discussion suggests that the fraction of risky high LTV ratio mortgages increased over time. In Table 2.1, we document that the fraction of mortgages with an LTV ratio of greater than 90% increased from 7% in 1992 to over 15% by 1999. In 2007, such loans made up about a fifth of all mortgages in our sample. A similar pattern arises when we look at the behavior of the sampled mortgages' Debt-to-Income (DTI) ratios over time. The DTI ratio is defined as the fraction of the borrower's monthly income that is relied upon in paying the monthly mortgage debt (see histogram in Figure 2-7). Higher DTI ratios are consistent with riskier mortgages as a greater burden is placed on borrowers' existing incomes to service their mortgages.

[1] For example, among other requirements, FHEFSSA mandated GSEs to reach a target percentage of their mortgage purchases to be secured by homes of low- and moderate-income households.

The fraction of risky high DTI ratio mortgages also increased over time in our sample. For example, in Table 2.2, we see that the share of DTI ratios between 0.42 and 0.65 increased from less than 10% prior to 1995 to 27% in 2000 and eventually peaked at 41% in 2007.

Figure 2-8 provides a histogram of homeowners' FICO scores. Notice that the histogram is noticeably skewed towards lower FICO scores. In Table 2.3 we use this FICO information and classify our sampled mortgages into the following three progressively riskier categories: LTV ratio $> 80\%$ and/or FICO score < 660, LTV ratio $> 90\%$ and/or FICO < 620 and finally LTV ratio $> 95\%$ and/or FICO $<$ 580. The share of the latter two categories generally increased over time, particularly starting from 1993 onwards. The middle category saw a decrease in it's share after 2003 but the share of the most risky mortgages (the LTV $>$ 95 and/or FICO $<$ 580 category) kept rising until 2007, when it peaked at 14% of all loans. Consistent with arguments by Acharya et al. (2011) and others, this analysis suggests that the quality of loans purchased by Fannie Mae deteriorated over time.

Having investigated the characteristics of the home mortgages purchased by Fannie Mae, we turn our attention to summary statistics of borrower characteristics. In Table 2.4 we see that, on average, borrowers' income and FICO scores rose over time. However, if we instead turn to Table 2.5 where these characteristics are summarized for three different LTV ratio categories, we find that generally poorer, younger and riskier (those with low FICO scores) borrowers are leveraged the highest. Finally, looking at the occupancy status of the underlying properties, we note in Table 2.6 that the share of mortgages secured by second homes and investment properties steadily increased over time.

In summary, we note the following stylized facts about the mortgages purchased by Fannie Mae. The CDFs of LTV ratios suggest the existence of an interior solution to the problem of a household's LTV choice. The fraction of risky mortgages increased over time, especially after passage of FHEFSSA in 1992. In addition, we document that borrowers who are younger, poorer and with low FICO scores are leveraged more. Finally, the share of mortgages secured by second homes and investment properties has increased over time.

These stylized facts motivate the research questions to be answered in this article. First, how does the borrower arrive at an optimal leverage ratio? Second, what explains the increase in household mortgage leverage over time? Is it due to an increase in borrower demand for leverage or due to a greater supply of it? We now turn to addressing these questions.[2]

2.3 A Demand and Supply Model of Leverage

In this section, we conceptualize a market for leverage. We set up a partial equilibrium model where a lender and a borrower optimally choose a leverage ratio, given default risk that depends on future house prices. The lender model derived below is a specific form of a more general credit rationing model specified in Jaffee and Stiglitz (1990). It most closely resembles the model of mortgage strategic default developed by Brueckner (2000). The major point of departure from that model is that we assume lenders to be price takers. Thus, the note interest rate is taken as given and the lender maximizes expected profits by choosing an appropriate leverage ratio to supply. The same exception applies to the borrower model, also based on it's counterpart in Brueckner (2000). In addition, we treat borrowers as risk averse (as opposed to risk neutral) and consider ruthless default instead. This latter assumption is not crucial to the model and is left out because our task is to derive empirical predictions that do not depend on variables not available to us in the data (such as the borrower's cost of default that serves as asymmetric information and drives strategic default in Brueckner (2000) and Harrison et al. (2004)).

An advantage in this approach of separately deriving the demand and supply for leverage is that this gives us results on comparative statics that are specific to the lender and the borrower, thus enabling us to better understand the factors affecting the two sides of the market. We first set up the lender's problem below, followed by the borrower model and end this section with a discussion of a leverage market

[2]For further details on the data cleaning process, please refer to the appendix.

equilibrium.

2.3.1 Lender Model

Lenders are assumed to be risk-neutral and functioning in a competitive market. The assumption of a competitive market implies that each lender is essentially a price taker. In the 2-period model derived below, a borrower wishes to buy a house of value V, which is set equal to 1, so that the choice of the loan amounts to choosing a loan-to-value (LTV) ratio. In the first period, the lender lends an amount L to the borrower, earning a payoff of $-L$. In the second period, the borrower must sell the house and pay back both the principal and interest (at the rate r). The ability of the borrower to return the amount $L(1+r)$ depends on the value of the house, P, in the second period. If value of the house exceeds the amount owed, then the lender's payoff is $L(1+r)$. However, if the value of the house is less than the amount owed, we assume that the borrower ruthlessly defaults and the lender's payoff is P, the proceeds from the sale of the house under foreclosure. Furthermore, it will be convenient to assume that second period house prices are uniformly distributed with pdf $f(P)$ and support $[P_L, P_H]$.[3] The lender's expected profits are therefore given as:

$$\pi = -L + \eta \int_{P_L}^{L(1+r)} Pf(P)dP + \eta \int_{L(1+r)}^{P_H} L(1+r)f(P)dP$$

This expected profit equation is the expected utility of the payoffs in the two periods. When the house price, P, is between P_L and what is owed, $L(1+r)$, the borrower defaults and the lender receives the low payoff of P (discounted to the first period by lender patience or discount factor, η). In the case where the house price is higher than what is owed, (i.e. it is between $L(1+r)$ and P_H), the lender's payoff is the high outcome of a full repayment of the principal and interest. Furthermore, under the assumption that the value of the house in the first period is 1, all the variables in the model can be expressed in terms of the numeraire. For e.g., L can be expressed as a fraction of the value of the house and will therefore represent a loan-

[3]The mean is given by $\frac{(P_L+P_H)}{2}$. The range of prices is $P_H - P_L$.

to-value ratio. Next, we derive what the optimal LTV a lender is willing to supply at a given interest rate, r. In other words, our objective is to derive the lender's offer curve for various r-LTV combinations. Maximizing expected profit w.r.t L, we get the following first-order condition:

$$\frac{d\pi}{dL} = -1 + (1+r)\eta[\frac{P_H - L(1+r)}{(P_H - P_L)}] \stackrel{set}{=} 0$$

It is easily verified that expected profit are indeed maximized.[4] The first-order condition yields the optimal amount loaned:

$$L^S = \frac{P_H}{(1+r)} - \frac{(P_H - P_L)}{(1+r)^2\eta} \tag{2.1}$$

Comparative statics of the loan offer curve gives a number of results. First, taking the derivative of (2.1) with respect to r, we get:

$$\frac{dL^S}{dr} = -\frac{P_H}{(1+r)^2} + \frac{2(P_H - P_L)}{(1+r)^3\eta}$$

For the loan offer curve to be upward sloping, $\frac{dL^S}{dr} > 0$. This implies that the interest rate, r, cannot exceed $(1+r) = \frac{2(P_H - P_L)}{P_H\eta}$.This is analogous to having a maximum loan size beyond which the lender's offer curve begins to bend backwards. At some high loan level, the borrower's liability, $L(1+r)$, may exceed the upper bound on house prices P_H. Since this would guarantee default, the lender won't offer any higher loan amounts beyond a certain maximum amount. For all r, the maximum LTV given in (2.1) is $\frac{P_H}{1+r}$. The backward-bending portions are not relevant to us but such a feature is characteristic of the loan offer curves in general models of credit rationing.[5]

This model also makes empirical predictions with regards to house prices. In order to derive the effects of expected prices and the range of prices on leverage supply, let's look first at the individual effects of P_L and P_H separately. If we increase P_L, holding P_H (and all else) constant, then given our assumptions on $f(P)$, this results in

[4]The second order condition is: $\frac{d^2\pi}{dL^2} = -(1+r)^2\frac{\eta}{(P_H - P_L)} < 0$

[5]See Jaffee and Modigliani (1969), Jaffee and Russell (1976) and Jaffee and Stiglitz (1990).

simultaneously higher expected prices but a lesser range of prices. Taking the partial derivative of L^S with respect to P_L,

$$\frac{\partial L^S}{\partial P_L} = \frac{1}{(1+r)^2 \eta} > 0$$

This implies that the supply curve shifts outwards. The interpretation is that in markets with higher average prices and lesser range of prices, the supply of credit will be more. This is not particularly useful at the moment because we are ultimately interested in separating out the pure effects of higher prices and greater range. We will address this shortly but next we look at the effect of increasing P_H.

If we increase P_H holding P_L (and all else) constant, we increase both expected prices and the range of prices. Taking the partial derivative of L^S with respect to P_H,

$$\frac{\partial L^S}{\partial P_H} = \frac{1}{(1+r)} - \frac{1}{(1+r)^2 \eta}$$

This has an ambiguous sign. In order for L^S to be increasing in P_H, i.e. $\frac{\partial L^S}{\partial P_H} > 0$, interest rates would have to satisfy the condition, $(1+r) > \frac{1}{\eta}$. The increase in P_H increases expected profits but also increases the probability of default due to a greater range in realized prices. For the lender to actually increase the supply of credit, he/she has to be compensated by a higher interest rate, as reflected by the condition $(1+r) > \frac{1}{\eta}$. For interest rates lower than this, $\frac{\partial L^S}{\partial P_H} < 0$, and the increase in P_H would result in a fall in supply. Thus, the two parts of the supply curve behave differently depending on the prevailing interest rates.

Now that we have obtained the pure effects of P_H and P_L, we next look at the effect of the range in prices, holding the mean of the distribution constant. If we increase P_H by one unit and decrease P_L by one unit, then the mean of the resulting distribution would be the same as the old distribution. The new distribution would also be riskier in the Diamond-Rothschild-Stiglitz (mean-preserving spread) sense.

Formally, we can see the effect of this as:

$$\frac{\partial L^S}{\partial P_H} - \frac{\partial L^S}{\partial P_L} = [\frac{1}{(1+r)} - \frac{1}{(1+r)^2\eta}] - \frac{1}{(1+r)^2\eta}$$

$$\frac{\partial L^S}{\partial P_H} - \frac{\partial L^S}{\partial P_L} = \frac{1}{(1+r)} - \frac{2}{(1+r)^2\eta}$$

If the above expression were to be positive, it would imply that $(1+r) > \frac{2}{\eta}$. Even for a lender discount factor of 0.99, this would imply interest rates of over 100%, which is an impractical implication. Thus, we can safely conclude that the above expression is negative, implying that with greater risk (higher range of prices, holding mean prices (and all else) constant), lenders would supply lower leverage at every interest rate.

Similar to above, we can analyze the effect of an increase in mean prices while keeping the range of prices constant. If we increase both P_H and P_L by one unit, we increase the mean also by 1 unit but keep the range of prices constant. Formally, the effect of this can been by:

$$\frac{\partial L^S}{\partial P_H} + \frac{\partial L^S}{\partial P_L} = \frac{1}{(1+r)}$$

Thus, the supply of LTV increases with higher expected prices, holding the range of prices constant.

Figure 2-9 illustrates the rate-LTV combinations that a lender is willing to supply at and the above-derived effects on the offer curve. In particular, three offer curves are shown with different sets of values for P_L and P_H. In all three cases, the lender patience factor is set equal to 0.8 ($\eta = 0.8$) and, as mentioned earlier, the value of the house is equal to 1. In the base case represented by the triangle-blue line, the values of P are distributed $P_L = 0.5$, $P_H = 1.1$. This curve gives the upward sloping rate-LTV combinations for this distribution of house prices. In a second case, house prices are distributed $P_L = 0.6$, $P_H = 1.2$, i.e. where expected prices are higher by 0.1, but the range is kept constant. The circled-black line shows this effect that with higher expected prices, the supply of leverage is greater than that in the base case. A third case shows the effect of a mean-preserving spread in the distribution of house prices. In particular, this is the case where $P_L = 0.4$, $P_H = 1.2$, a distribution with the same mean as that of the base case. The effect of this is that with greater risk , the supply of leverage falls or shifts in from the base case to the squared-green line. Note that only the increasing portion of these loan offer curves are the relevant supply functions and not the backward-bending portions.

2.3.2 Household Model

In a competitive market, borrowers are price takers and thus maximize their expected utility over the choice of a leverage or LTV ratio. In the general two-period model below, a household has a first-period wealth and discounted value of all future incomes, Y. The value of the house in the first period is 1, i.e. the purchase decision is made outside of the model. In the first period, the household chooses a loan of amount L and enjoys a payoff of $Y-(1-L)$, where $1-L$ is the down payment and $L \leq 1$ (i.e. unsecured debt is not allowed). Any surplus from the first period is transferred over to the second period, which earns a rate of return of $(1+e)$. In addition, the household has to sell the house for a value of P, realized stochastically and distributed uniformly (with pdf $f(P)$ and cdf $F[P]$). If the value of P exceeds the amount owed, $L(1+r)$, the household pays the amount in full and enjoys a surplus of $P-L(1+r)+(Y-(1-L))(1+e)$. If, however, the value of the house in the second period falls below the amount owed, the borrower defaults and enjoys a surplus of $(Y-(1-L))(1+e)$.

In the model below, households are assumed to be risk-averse.[6] Household's utility functions are assumed to be concave and additively separable over the 2 periods. In particular, let u and v be the first and second period utility functions, respectively, with the properties $u' > 0$, $v' > 0$, $u'' < 0$ and $v'' < 0$. The household expected utility

[6]An entire household is treated as an individual. The terms household and borrower are used interchangeably.

can be written as:

$$\begin{aligned}\Omega &= u[Y-(1-L)] + \delta \int_{P_L}^{L(1+r)} v[(Y-(1-L))(1+e)]f(P)dP \\ &+ \delta \int_{L(1+r)}^{P_H} v[(Y-(1-L))(1+e)+P-L(1+r)]f(P)dP\end{aligned}$$

The first term in the household's expected utility function is the utility of the first-period payoff of $Y-(1-L)$. Since the decision on the loan amount is being made in the first period, the second period expected utility is discounted by the borrower discount or patience factor, $\delta < 1$. The second period expected utility is the probability of default times the bad outcome at default $v[(Y-(1-L))(1+e)]$ plus the probability of no default times the good outcome in the absence of default $v[(Y-(1-L))(1+e)+P-L(1+r)]$. It can be verified[7] that it is optimal to default when $P = L(1+r)$ and thus the probability of default is given by $F[L(1+r)]$.

A few key properties of the above general utility function are that:

1. Utility is increasing in the rate of return on surplus, i.e. $\frac{d\Omega}{de} > 0$.

2. Utility is decreasing in the cost of borrowing, i.e. $\frac{d\Omega}{dr} < 0$.

3. If $e > r$, then $\frac{d\Omega}{de} - \frac{d\Omega}{dr} > 0$. However, if $e < r$, $-(\frac{d\Omega}{de} - \frac{d\Omega}{dr}) < 0$.

The first part of property 3 states that if the rate of return on the surplus exceeds the rate on the loan ($e - r > 0$), the borrower utility will always be increasing no matter what loan rate, r is. It is easily seen from the first-order condition ($\frac{d\Omega}{dL}$) that this leads the borrower to borrow as much as possible:

[7] All derivations and proofs are relegated to the appendix.

$$\begin{aligned}\frac{d\Omega}{dL} &= \underbrace{u'[Y-1+L]}_{>0 \text{ by assumption}} \\ &+ \delta(1+e)\int_{P_L}^{L(1+r)} \underbrace{v'[(Y-(1-L))(1+e)]f(P)dP}_{>0 \text{ by assumption}} \\ &+ \delta\underbrace{(e-r)}_{>0}\int_{L(1+r)}^{P_H} \underbrace{v'[(Y-1)(1+e)+P+L(e-r)]f(P)dP}_{>0 \text{ by assumption}}\end{aligned}$$

No matter what the loan rate, $\frac{d\Omega}{dL} > 0$. Thus, the borrower will choose to borrow as much as possible, and obtain a maximum LTV of $L^D = 1$.[8]

The second part of property 3 states the converse, i.e. if the rate of return on the surplus does not exceed the loan rate, $(e - r < 0)$, then the borrower's utility will always be decreasing. Furthermore, the borrower can be at a higher utility level if he/she allocates the first period surplus for either down payment or consumption instead of transferring it to the second period. This will be higher utility because utility will then only fall by $\frac{d\Omega}{dr}$.[9] Note that this does not imply that the borrower will not borrow. For a given loan rate, the borrower, through his/her choice of L, trades off (at the margin) an increase in current wealth with a fall in expected (discounted) future wealth, conditional on house prices. We now turn to formally showing this result.

For the case $(e - r < 0)$, the borrower's expected utility can now be written as:

$$\Omega = u[Y-(1-L)] + \delta\int_{P_L}^{L(1+r)} v[0]f(P)dP + \delta\int_{L(1+r)}^{P_H} v[P-L(1+r)]f(P)dP \qquad \text{for } e < r$$

This model is interpreted as follows. The first-term, again, is the utility of the first-period payoff of $Y - (1 - L)$. As argued above, any surplus in this period is

[8]This result is also found in Brueckner (1994) where all households demand 100% LTV ratios if the return on equity capital (e) exceeds the costs of mortgage debt (r).

[9]Harrison et al. (2004) make this assumption explicit in their borrower's objective function. It is implicitly assumed in the model by Brueckner (2000).

either consumed or put in the down payment for the reason that that would make the borrower better off. Thus, in the second period, if realized house prices are below $L(1+r)$, the borrower defaults and has a utility of surplus 0, $v(0)$. In the event that realized house prices are between $L(1+r)$ and P_H, the borrower sells the house and pays the lender what is owed and in return enjoys a surplus of $P - L(1+r)$.

Maximizing the borrower expected utility w.r.t to L, the following first-order condition is obtained:

$$u'[Y - 1 + L] - \delta(1+r) \int_{L(1+r)}^{P_H} v'[P - L(1+r))f(P)d(P) \overset{set}{=} 0 \tag{2.2}$$

The interpretation of this equation is that, in equilibrium, a household chooses an L such that it equates the marginal utility of first period wealth to the expected (discounted) marginal utility of second period wealth. The tradeoff in the borrower's decision is that increasing the loan amount (and thus, leverage) increases the first period wealth but it also increases the expected future loan payment, thus decreasing second period wealth, conditional on house prices. From (2.2), the household's optimal loan demand can be implicitly written as:

$$L^D = L^D[r, P, Y, \delta]$$

Comparative statics of (2.2) yields several properties of the optimal leverage demand schedule. These are summarized in the proposition below and all proofs are provided in the appendix.

Proposition 1. *The optimal loan demand L^D has the following properties:*

1.1 As r rises, the minimum LTV demand is $L^D = 1 - Y$ for $Y < 1$ and $L^D = 0$ for $Y \geq 1$.

1.2 In the $\lim_{r \to e}$, the maximum LTV demand is $L^D = 1$. For $r < e$, $L^D = 1$.

1.3 L^D has an interior downward sloping schedule w.r.t loan rate r ($\frac{dL^D}{dr} < 0$).

1.4 L^D is decreasing in borrower patience δ, ($\frac{dL^D}{d\delta} < 0$).

1.5 L^D is decreasing in borrower incomes Y, ($\frac{dL^D}{dY} < 0$)

1.6 If a distribution P first-order stochastically dominates a distribution P', then $L^P \geq L^{P'}$.

1.7 If a distribution P second-order stochastically dominates a distribution P' then $L^P \geq L^{P'}$.

In Figure 2-10, Propositions 1.1 to 1.3 are illustrated in panel (a), and Propositions 1.4 to 1.7 are illustrated in panels (b) to (e). As the interest rate on the loan increases, LTV demanded falls along a downward sloping schedule. The intuition for this result is that a higher interest rate lowers second period wealth and to smooth wealth over the two periods, a borrower would decrease their leverage to simultaneously lower first period wealth and instead increase second period wealth. The minimum LTV can be $1 - Y$ for those with $Y < 1$. It will zero for those that are not constrained by income ($Y > 1$). As the rate on the loan falls and approaches the rate of return on the surplus, e, the maximum LTV demanded would be 1. The same result holds for rates below e. Proposition 1.4 establishes that more patient borrowers, i.e. borrowers that value future wealth more, would have lower demand for leverage. In Proposition 1.5, borrowers with higher Y would find that the marginal increase in first period wealth (from an increase in L) would not be as much as for them as it would be for those with lower incomes. Therefore, they would comparatively lever less. Proposition 1.6 states that if there were two house price distributions and one unambiguously yielded higher average prices, then there would be greater demand for leverage in that market. Finally, Proposition 1.7 states that if there were two house price distributions and one had unambiguously greater risk (but same average prices) than the other, then a risk averse borrower would demand lower leverage in that market.

Equilibrium The highest LTV ratio demanded is 1 while the highest supplied is $\frac{P_H}{1+r}$, which would often be greater 1 (since $P_H > 1$). The lowest leverage demand is $1 - Y$ (or 0) while the lowest supplied at $r = 0$ in (2.1) would be $\frac{\eta P_H - P_H}{\eta} + \frac{P_L}{\eta} < \frac{P_L}{\eta}$. These would represent small loan balances, $L(1+r)$, below P_L that are risk-free (and thus supplied at r = 0). In general, we should not expect these amounts to exceed

$1-Y$ ($Y < 1$). Given that demand is downward sloping and supply is upward sloping between the above fixed points (at high and low rates), this ensures that there is an interior equilibrium point of intersection between demand and supply as shown in Figure 2-11.

2.4 Empirical Analysis

2.4.1 Empirical Strategy

The model of leverage derived in the previous section yields two structural equations describing the demand and supply of leverage, which we specify as follows:

$$L = \beta_0 + \beta_1 r + X\beta_2 + T\beta_3 + \epsilon \tag{2.3}$$

$$r = \gamma_0 + \gamma_1 L + X\gamma_2 + T\gamma_3 + \gamma_4 CLL + \nu \tag{2.4}$$

Leverage demand is given by (2.3), where LTV (L) is a linear function of the note rate (r), a matrix of exogenous variables representing borrower and market characteristics (X), and a matrix of yearly time dummies (T). Similarly, leverage supply is given by (2.4), where the note rate (r) is a linear function of LTV (L) and the same X and T matrices of characteristics and time dummies. In addition, we include the log of the national conforming loan limit (CLL) as a supply-shifter.[10] Similar to the Adelino et al. (2011) paper, where changes in the national conforming loan limit were used as an instrument for changes in the cost of credit, we view these exogenous changes as supply shifts because over time they make the cost of credit to be cheaper for houses that would otherwise require financing via a jumbo loan (or part conforming, part more expensive financing). Since changes in the CLL are based on national appreciation of house prices, it is reasonable to assume that these changes are exogenous to individual mortgages and to local housing market conditions (and thus avoid correlation with

[10] Up until 2008, the conforming loan limit was set nationally with the exception that it was always 1.5 times higher in Hawaii/Alaska. Starting from 2008, there was an additional county-based CLL available for areas that were determined as high cost. With the exception of 1990, where CLL fell by \$150 over the previous year, changes in it have always amounted to an increase in the limit.

our MSA measures in X).

In both of our structural equations, X does not contain any specific demand-only or supply-only variable. This is because our dataset does not contain any variable that is observed only by the borrower and not by the lender, and vice versa. Thus, the only variable excluded from the demand equation is the conforming loan limit (CLL). It is easy to see that (2.3) is just-identified.[11] Furthermore, since there are no demand-only variables, the supply equation, (2.4), is not identified. However, looking at the reduced form, the coefficient on CLL is uncontaminated and it's estimate would help answer how leverage supply has changed due to policy changes in this supply-shift variable:

$$L = \pi_0 + X\pi_1 + T\pi_2 + \beta_1\gamma_4 CLL + u_1 \tag{2.5}$$

$$r = z_0 + Xz_1 + Tz_2 + \gamma_4 CLL + u_2 \tag{2.6}$$

The coefficient on CLL in (2.6) also allows for an indirect least squares estimate of the interest rate elasticity of leverage demand (β_1). This would be obtained by dividing the reduced form coefficient on CLL in (2.5) by that in (2.6).

We estimate (2.5) and (2.6), equation by equation using OLS.[12] We also estimate the leverage demand equation, (2.3), via 2SLS. Since leverage demand is just-identified, a limitation is that we cannot perform a test of over-identifying restrictions. In the case of purchase mortgages, we also separately estimate the numerator and denominator in LTV, i.e. loan and housing demand (V) regressions, using the same exogenous variables. This is because in our theory, we derived results on leverage under the assumption that the housing demand decision was given outside the model. Empirically, the house purchase decision cannot be included in the leverage regressions as it is an endogenous decision. By including it as a separate reduced form

[11]The demand equation, (2.3), has one exclusion restriction and one normalization (coefficient on L is 1). Thus, the sum of the restrictions (2) adds up to the number of endogenous variables (2). Since the order condition is satisfied with equality, (2.3) is just-identified. The rank condition also holds as long as the coefficient on CLL in the supply equation is not zero ($\gamma_4 \neq 0$) (see appendix). The supply equation, (2.4), fails the order condition as there is only one restriction (normalization on r) which is less than the number of endogenous variables (2). Thus, (2.4) is not identified.

[12]It is well known that in a system of linear *seemingly unrelated regression* equations with identical regressors, equation by equation OLS yields efficient parameter estimates.

regression, it allows us to better understand the leverage decision. In the case of refinances, while there is no housing decision being made and V changes stochastically, we will still find it useful to add a reduced form loan demand estimation.

2.4.2 Results

The estimates for the structural leverage demand equation are shown in Table 2.7 while the reduced form estimates are shown in Tables 2.8 (purchases) and 2.9 (refinances). We first discuss results for purchases and then compare any differences for refinances.

Leverage Demand: The 2sls estimates for purchase mortgages are shown in the first column of Table 2.7. The estimate on the note interest rate is -3.07, which is also verified via indirect least squares.[13] Equivalently, a log-log estimate[14] gives an interest rate elasticity of leverage demand of -0.37. This implies that a 10% increase in the note rate leads to a 3.7% decrease in the loan-to-value ratio. A few numerical examples illustrate this effect. From an r-LTV combination of (5%, 100%), if the rate increases to 6%, then this would lead to 7.4% [2 x 3.7%] fall in LTV to 92.7%. Starting from an r-LTV combination of (10%, 70%), a subsequent fall in interest rates to 5% would lead to a 18.5% [5 x 3.7%] increase in LTV to 83%. These are quite plausible estimates of demand elasticity. Furthermore, holding note rate and all else constant, the year dummies can be interpreted as changes in taste or preference of borrowers for LTV. These dummies are used in Figure 2-12 to trace the evolution of LTV demand, starting with an LTV of 72% (sample mean) for 1986. The first thing to note is that there is no upward trend and in fact, LTV demand is cyclical. Historical events that would be expected to negatively effect households and real estate markets are indeed reflected in down-ticks or falls in leverage demand. LTV demand was relatively healthy during periods of steady economic growth and particularly strong in the most recent real estate boom. We now turn to look at leverage supply.

[13]The coefficients on CLL in Table 2.8 provide an indirect estimate by dividing the coefficient in the LTV column by that in the Note Rate column ($\frac{12.4648}{-4.0568} = -3.07$).

[14]There is however a slight attenuation bias in that regression due to the fact that ln(1) is undefined.

Leverage Supply: There are two pieces of evidence that suggest that leverage supply has increased over time. First, as argued in the previous section, the coefficient on CLL in the reduced form note rate regression in Table 2.8 is uncontaminated and gives it's structural marginal effect. Also, since all the variation in that variable is over time, it's coefficient tells us how much leverage supply has changed due to shifts caused by changes in CLL. We find that a doubling or a 100% increase in CLL leads to a fall in the average note rate by 4% [(4.0568/100) x 100]. Over in the LTV column of Table 2.8, this increase in CLL leads to an increase in LTV of 12.5 percentage points. As expected, this combination of falling rates and rising LTV corresponds to an outward shift in the leverage supply curve. Furthermore, examining the Ln Loan and Ln Price columns, a 100% increase in the CLL is associated with an increase in the average purchase price by 44% and the average loan size by 58%. This would be consistent with the intuition that increases in the CLL would make expensive houses more attractive by making them cheaper to finance. These results indicate a strong effect of policy changes in CLL on the market for mortgage leverage.

A second look at how leverage supply has changed over time begins by examining the reduced form yearly time dummies. Since CLL captures the same variation over time, we drop that variable and reestimate the note rate and LTV regressions.[15] These yearly time dummies for the reduced forms and the structural demand equation are shown in Table 2.10. Several observations can be made from the reduced form columns of note rate and LTV. Relative to 1986, the average note rate fell over time while the loan-to-value ratio rose concurrently. Specifically, when compared to 1986, the note rate was 1.75%, 2.4% and 3.47% lower in 1995, 2001 and 2005, respectively. At the same time, the LTV ratio was higher by 5.9%, 13.5% and 13.4% in 1995, 2001 and 2006, respectively. Comparing the same time periods in the structural LTV demand column, we see that demand was 5.2%, 6.1% and 2.7% higher.[16] These results strongly suggest that leverage supply had increased substantially over time. For e.g, in

[15]All coefficients in Table 2.8 are robust to the exclusion of CLL. Therefore, the full regressions are not shown but are available upon request.

[16]Each of these point estimates is statistically significant and significantly different from its counterparts.

2005, leverage supply was roughly 10.6 percentage points higher (compared to demand which was only 2.7 percentage points higher) than it's 1986 level. Furthermore, if demand was the primary reason for the rise in leverage, then we would have expected the note rate to have increased, not fallen, over time. We can also see that the reduced form LTV time dummies stayed relatively high even in 2008 (16%) and 2009 (13%) while demand collapsed to 0.17% (relative to 1986) in 2009. This is likely due to the fact that in the aftermath of the crises, the GSEs were the only suppliers of credit left.

House Prices: We next turn to the MSA Ln House Price Level variable in Table 2.7, constructed by first creating a series of average house price levels (for MSAs) in 2000 using the 5% PUMS sample of the Census, and then extrapolating those price levels using the quarterly MSA house price (repeat-sales) indices published by the Federal Housing Finance Agency (FHFA).[17] Thus, this variable measures, over time, the log house price level both across MSAs and within an MSA. In Table 2.7, we find that a 10% increase in the average house price level leads to a fall in the LTV ratio demanded by an economically small 0.57 percentage points (combined with a virtually insignificant fall in the rate shown in Table 2.8). Examining the Ln Loan and Ln Price regressions in Table 2.8, we find that this 10% increase in house prices leads borrowers to increase the size of their loans by 4.9% and to buy houses that are 5.7% more expensive.[18] Our borrower model predicted that with higher average prices, the borrower should be levered more. However, that result was derived based on a fixed value of house whereas empirically we find that borrowers roughly offset larger loans with equally expensive house purchases (holding all else equal), which implies that they put down more in equity. Moreover, since we control for income in the regression, a higher loan amount implies a higher loan-to-income ratio. This may mean greater debt service and exposure to greater risk of illiquidity in the future. There is corroborating evidence in a study at the aggregate MSA level in which Goetzmann et al. (2011) find that based on past price appreciation, households

[17]Further details on the construction of this variable can be found in the data appendix.

[18]The difference 5.7% - 4.9% = 0.8%, fall in LTV can be shown in a regression where the dependent variable is Ln LTV.

borrowed more and purchased more expensive houses. This subsequently lead to an increase in the loan-to-income ratio, again implying that households were at a greater risk.

House Price Volatility: The 2-year back, MSA De-trended Ln HPI Quarterly Volatility variable in Table 2.7 is simply the de-trended log volatility of the FHFA repeat-sales indices (lagged by 8 quarters).[19] In Table 2.7, a increase of 10% in the past house price volatility leads to a fall in the demand for leverage by 0.2 percentage points, which is consistent with the borrower theory. This latter figure is small because the magnitude of log volatility is less than 0.5, which implies that a 10% increase is a small change (e.g 10% increase from 0.1 is 0.11). Moving to the reduced form estimates in Table 2.8, the coefficient on HPI Volatility is positive in the note rate regression and more negative in the LTV regression, suggesting that the net effect of higher past price volatility is that lenders supply less leverage. This is because greater house price volatility increases the risk of borrower default which adversely affects expected profits. Consistent with our theoretical model, lenders would supply less leverage. Also in Table 2.8, the same 10% increase in past volatility, leads to a fall in the loan amount and the purchase price by 2.5% and 2.3%, respectively.

Borrower Characteristics: Looking at the Borrower's Total Monthly Income Amount variable in Table 2.7, we find that borrowers with monthly incomes higher by $5,000 lever less by an economically insignificant amount (0.41 percentage points) and (in Table 2.8) pay a rate that's only 1.6 bps less. Furthermore, the coefficients in the Ln Loan and Ln Price columns reveal that such borrowers not only carry a loan that's bigger by 11.4% [2.28e-05 x 5,000 x 100%] but they also buy a house that's 11.8% more expensive. This would explain why the leverage ratio would fall by a very small amount.

Next, we find in Table 2.7 that borrowers with bad credit scores demand more leverage. For example, a decrease in the Borrower Credit Score by 100 leads to an LTV ratio that is higher by 6 percentage points and (in Table 2.8) a note rate higher

[19]These results are robust to slightly shorter and slightly larger windows of lag. We do not use longer windows as we lose considerable data. Shorter windows, on the other hand, make the calculation of standard deviation much less reliable.

by 23 bps. Also from Table 2.8, borrowers with credit scores lower by 100 take out loans smaller by 4% and buy houses that are less expensive by 11%. To the extent that credit scores serve as a signal of riskiness and/or reflect the (asymmetric) default costs of a borrower, this result would be consistent with the predictions of models by Brueckner (2000) and Harrison et al. (2004). The reason is that riskier borrowers (those with low default costs) self-select into higher LTV ratios.

Our next finding is that the demand for leverage is monotonically decreasing with age. In Table 2.7, the base age group is 16-to-24 years and we add four age group dummies of 25-to-34, 35-to-49, 50-to-64 and above 64 years. Each age group demands lower leverage relative to the base group and to groups that are younger to it. For instance, the age group 35-to-49 levers 5.9 percentage points less than the base group and about 4.2 percentage points less than the 25-to-34 age group. Furthermore, a test of equality on the leverage ratios for every pair of these dummy variables rejects the null hypothesis that these groups behave the same. To the extent that older borrowers are more patient and value future wealth more than younger borrowers, we would expect to find that the demand for leverage falls with age (consistent with our borrower model).

Gender, Race and Occupancy: In Table 2.7, women demand less leverage than men. However, in Table 2.8, the sign on the Female variable is positive in the note rate regression and slightly more negative sign in the LTV regression. This suggests that women pay more for mortgages than men (leverage supply is less). Since we do not fully control for wealth, we cannot know for sure if these results are robust to unobservables. However, they are consistent with recent work by Cheng et al. (2011) that suggests that women pay higher rates because they are more likely to go to a lender by recommendation whereas men are more likely to search for (and find) a lower rate.

Relative to whites, the demand for leverage is higher for all other races. Again, these results are interesting but inconclusive due to a lack of information on other unobservables that may be correlated with these characteristics. The demand for leverage on second homes and investment properties is greater than that on first homes. Interestingly, the rate on an investment property is about 53 bps higher whereas it is higher by only 6 bps for second homes (Table 2.8). Also, the reduced form LTV estimate is less positive than it's structural counterpart. This implies that leverage supply is lower for investment properties.

Comparison with Refinances: Most results are consistent with the findings for purchases. The conforming loan limit had a similar effect on increasing leverage supply. Leverage demand was similarly cyclical. A 10% increase in average house prices lead borrowers to increase debt by 4.4%, which is slightly less than that for purchases. Finally, an increase in house price volatility reduced leverage supply for refinances and more adversely affected leverage demand than it did for purchases.

2.5 Conclusion

Why did household mortgage leverage rise from the mid-1980's until the Great Recession? We conclude that it is outward shifts in leverage supply that were the primary driver of the increase in leverage. By contrast, we find that leverage demand was cyclical and responsive to major economic events, but without a general upward trend. In this article, we developed a theory of leverage demand-and-supply and estimated an interest rate elasticity of leverage demand of -0.37. Our empirical results document the effects of house prices and borrower characteristics on household leverage. We find that greater house price volatility reduces LTV ratios while greater house prices lead borrowers to borrow more and buy more expensive houses. The effect of the latter was to keep leverage unchanged but it raised households' exposure to risk of illiquidity by increasing their loan-to-income ratios. We find that poorer, more impatient and less credit-worthy borrowers demand more leverage than their counterparts.

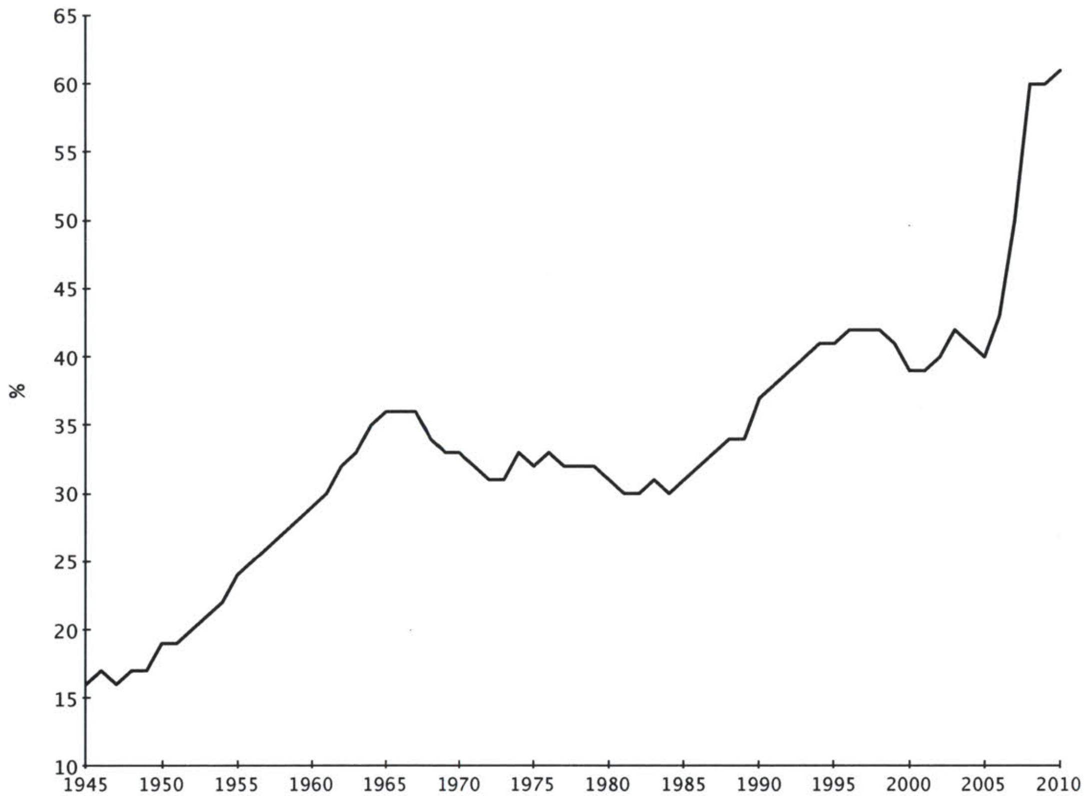


Figure 2-1: Mortgage Debt as a Percentage of Total Value of Housing

Source: Federal Reserve Household Balance Sheet

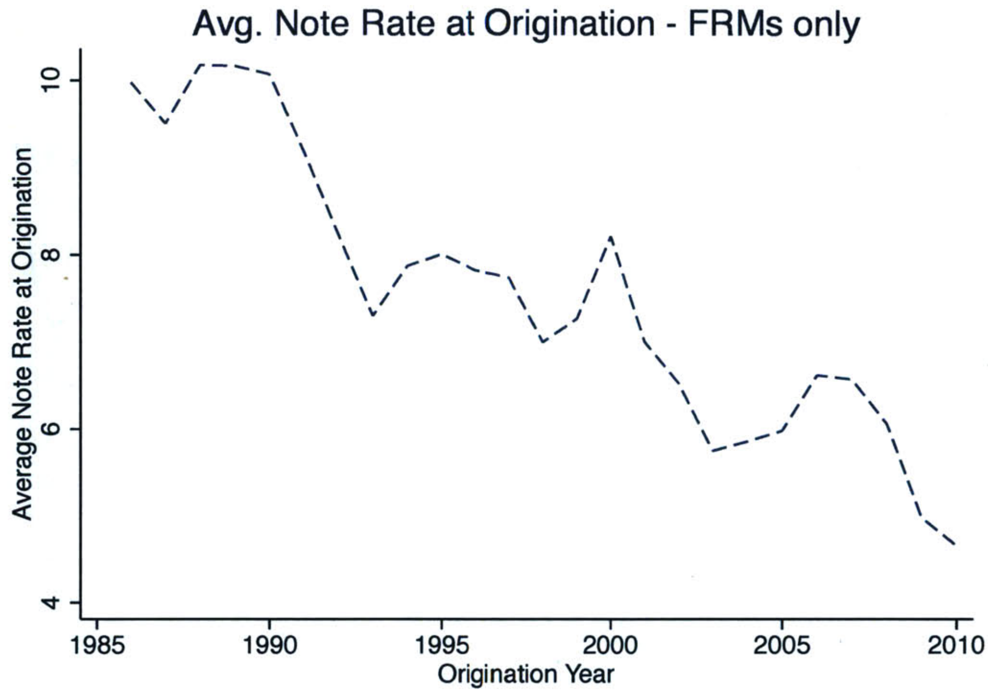


Figure 2-2: Average Note Rate of Fixed Rate Mortgages

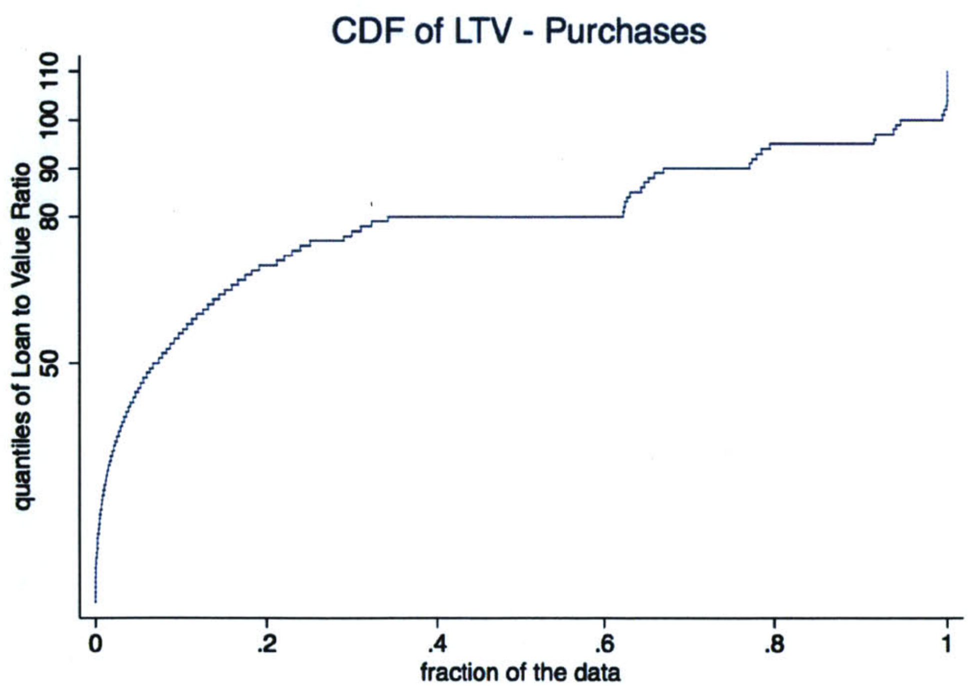


Figure 2-3: Cumulative Density Function of LTV - Purchases

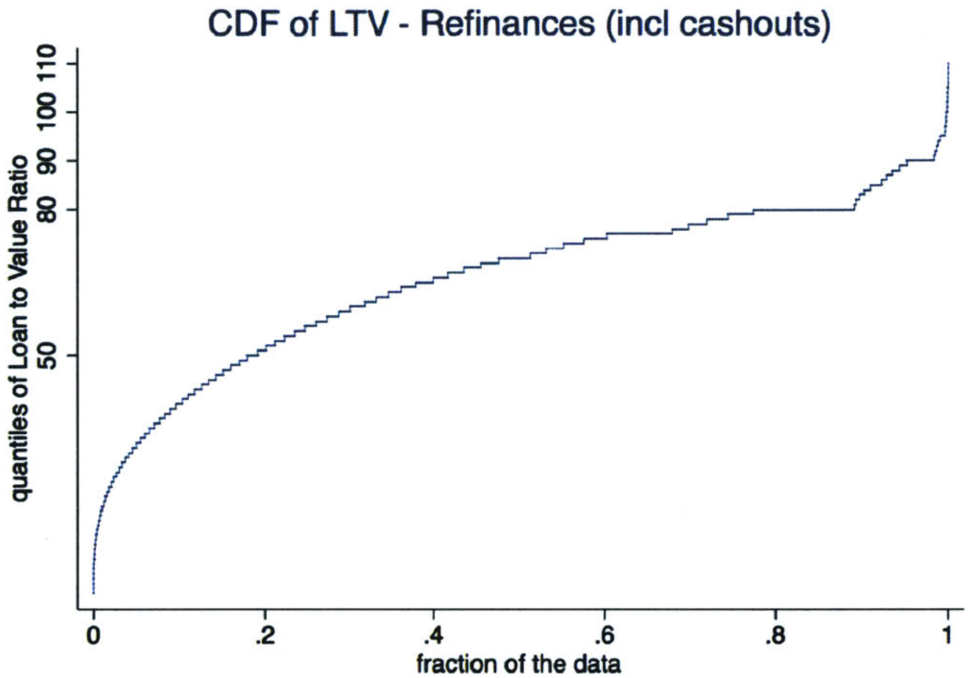


Figure 2-4: Cumulative Density Function of LTV - Refinances (incld. Cashouts)

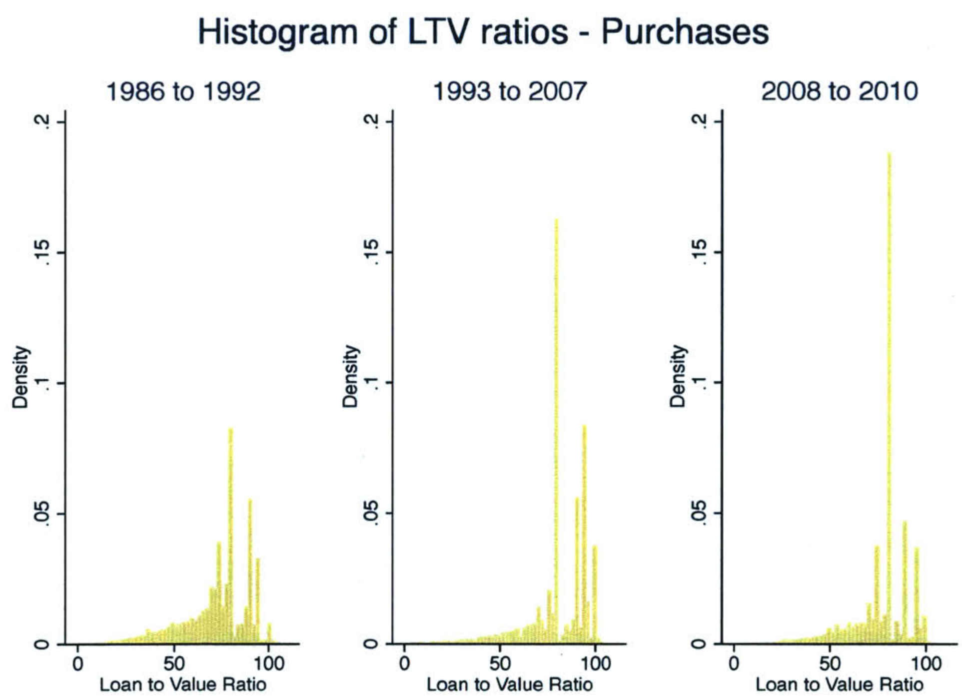


Figure 2-5: Histogram of LTV - Purchases over time

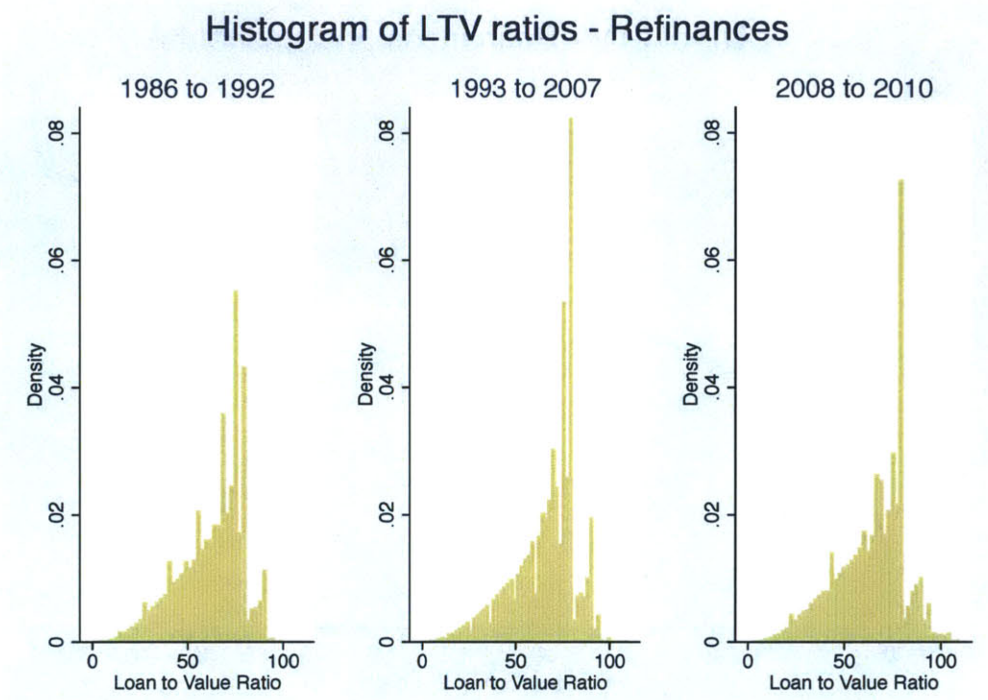


Figure 2-6: Histogram of LTV - Refinances over time

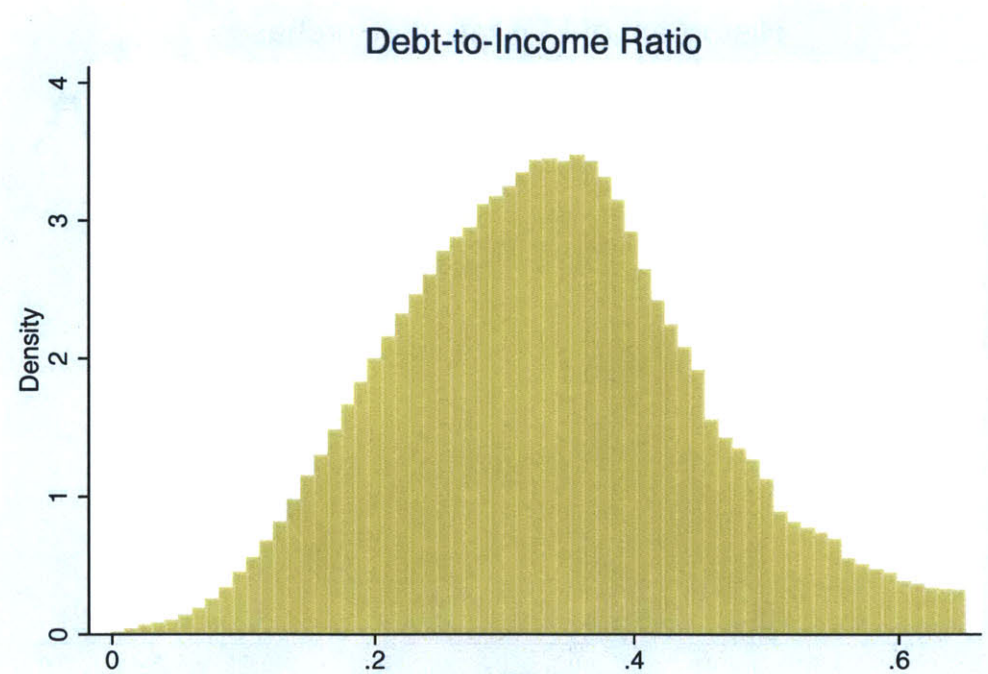


Figure 2-7: Histogram of Debt-to-Income Ratios

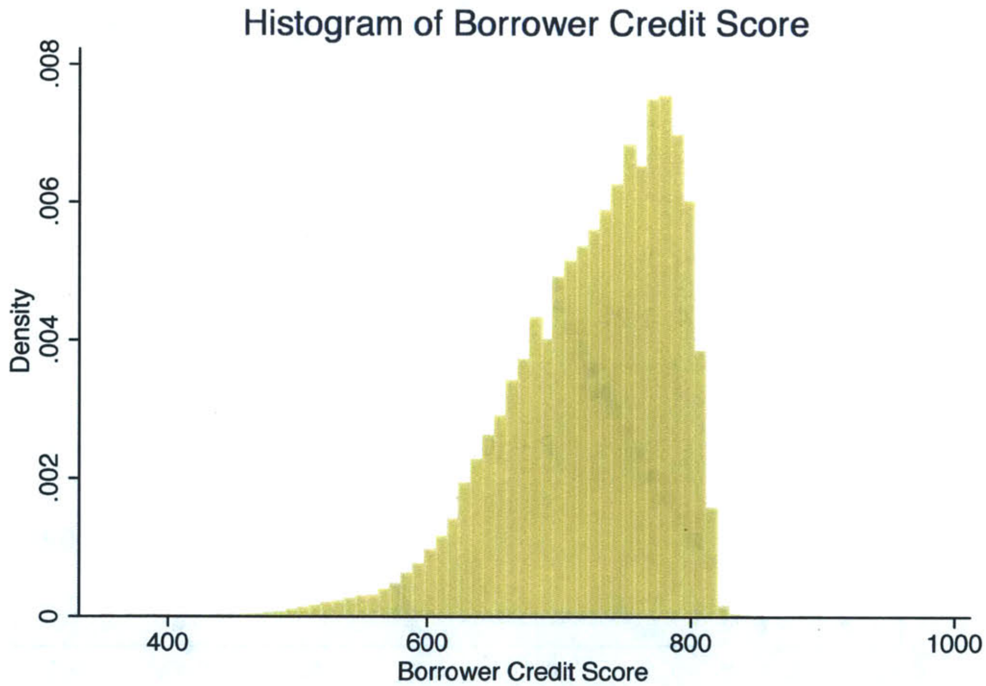


Figure 2-8: Histogram of FICO Scores

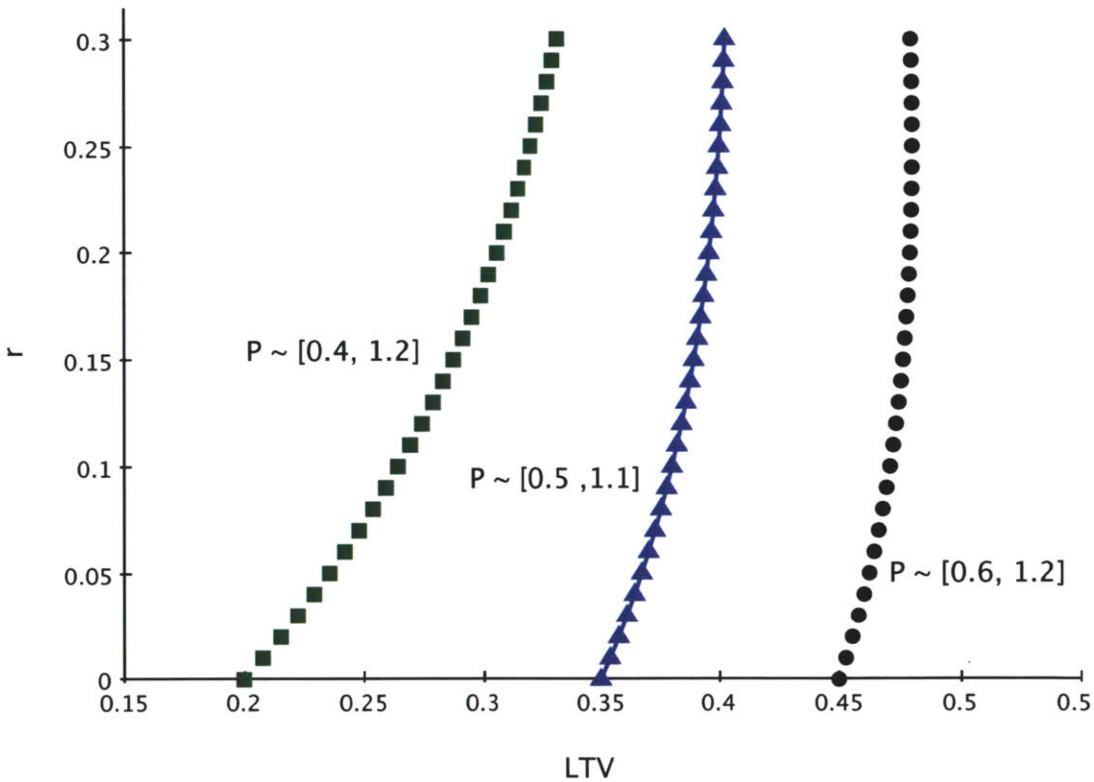
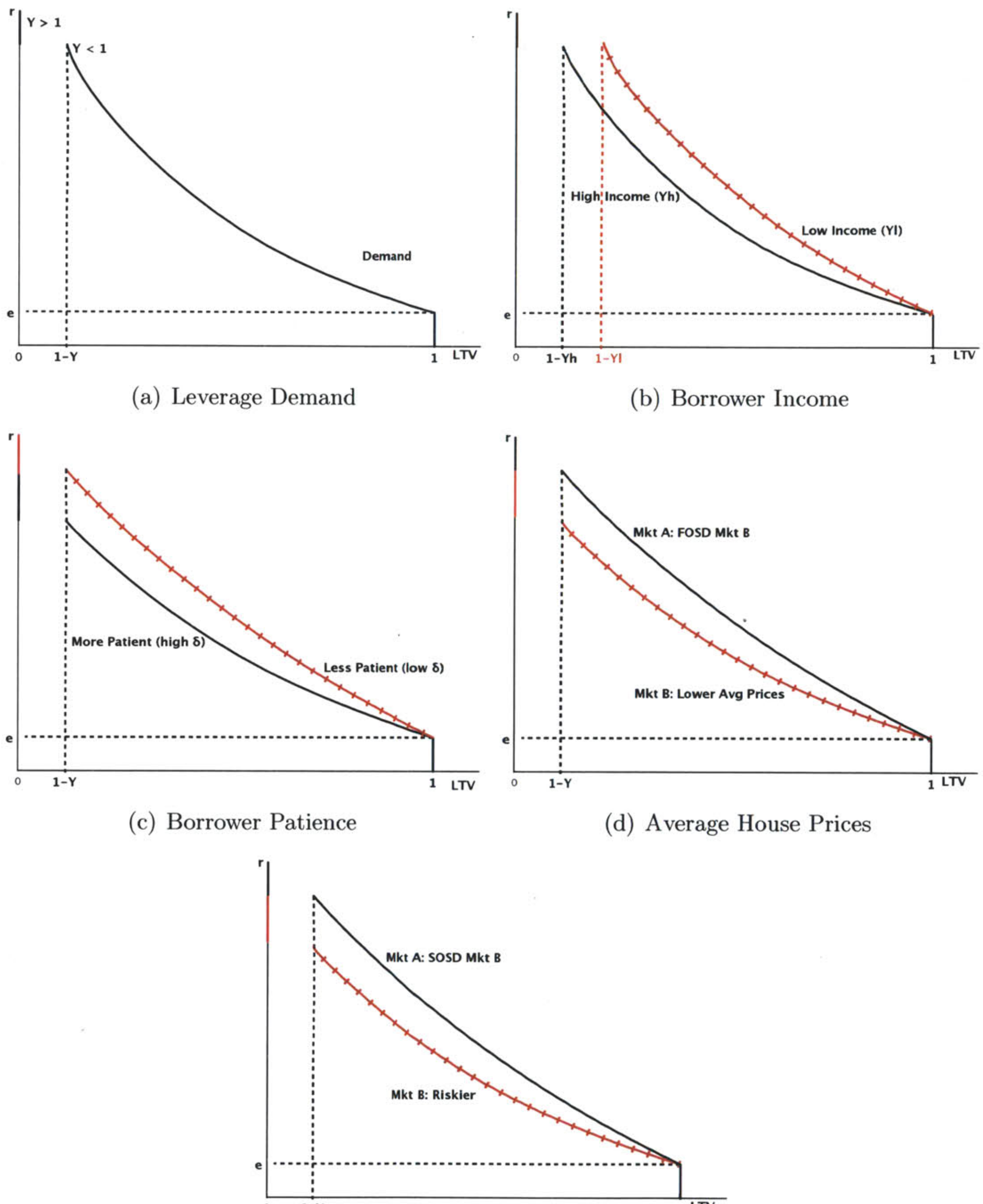


Figure 2-9: LTV Supply



(a) Leverage Demand

(b) Borrower Income

(c) Borrower Patience

(d) Average House Prices

(e) House Price Risk

Figure 2-10: Illustration of Proposition 1

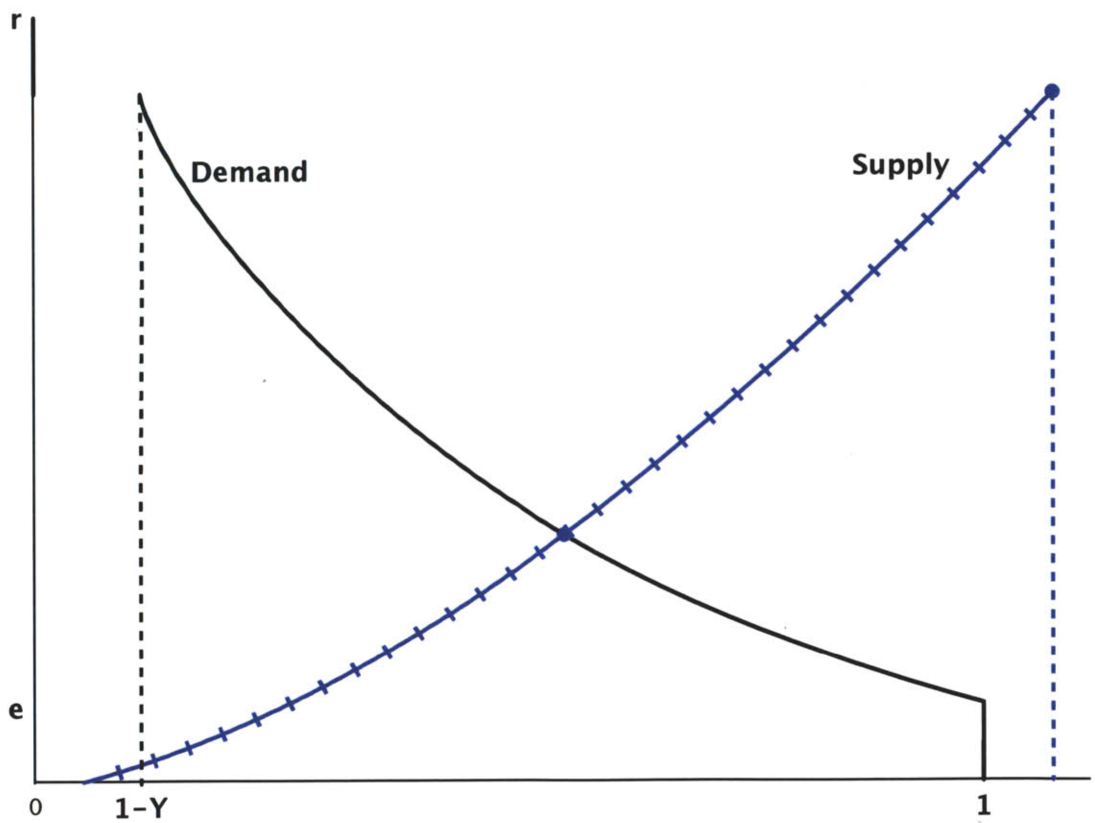


Figure 2-11: Equilibrium Leverage

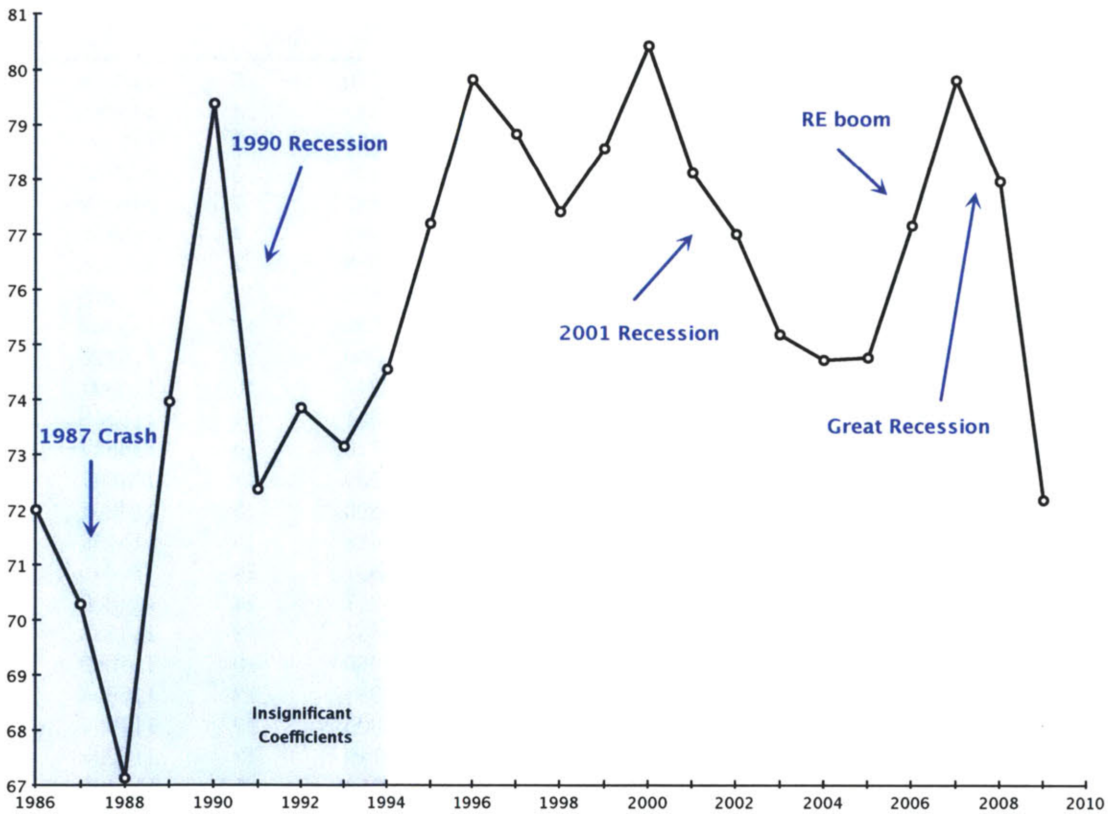


Figure 2-12: LTV Demand - Estimated from Yearly Dummies

Origination Year	LTV Category							
	LTV ≤ 80%		80% < LTV ≤ 90%		90% < LTV ≤ 110%		Total	
	No.	Row %	No.	Row %	No.	Row %	No.	Row %
1986	94940	81	17024	14	5510	5	117474	100
1987	90567	78	20482	18	4943	4	115992	100
1988	87585	74	23839	20	6160	5	117584	100
1989	92402	78	19003	16	6449	5	117854	100
1990	92007	78	18230	15	7882	7	118119	100
1991	93985	79	16390	14	8481	7	118856	100
1992	94661	80	15894	13	7873	7	118428	100
1993	86679	74	17980	15	12619	11	117278	100
1994	80752	69	18619	16	17936	15	117307	100
1995	79281	68	17388	15	20059	17	116728	100
1996	79995	69	17652	15	18244	16	115891	100
1997	82260	71	16207	14	16651	14	115118	100
1998	82153	71	15585	13	18776	16	116514	100
1999	80440	70	15288	13	19139	17	114867	100
2000	80792	71	16579	15	16890	15	114261	100
2001	81038	70	16023	14	18688	16	115749	100
2002	86395	74	13227	11	16827	14	116449	100
2003	88868	76	11215	10	16217	14	116300	100
2004	91089	79	9848	9	13851	12	114788	100
2005	94857	82	9172	8	12060	10	116089	100
2006	94315	81	8557	7	13984	12	116856	100
2007	83432	71	11823	10	21806	19	117061	100
2008	87780	75	14798	13	13719	12	116297	100
2009	102266	86	9850	8	6353	5	118469	100
2010	100090	85	10059	9	7984	7	118133	100
N	2,208,629		380,732		329,101		2,918,462	

Table 2.1: Percentage of Data by LTV Category

Origination Year	DTI Category							
	DTI < 0.26		0.26 ≤ DTI < 0.42		0.42 ≤ DTI ≤ 0.65		Total	
	No.	Row %	No.	Row %	No.	Row %	No.	Row %
1986	264	49	242	45	36	7	542	100
1987	427	48	402	45	62	7	891	100
1988	1614	31	3391	65	201	4	5206	100
1989	1756	23	5394	72	353	5	7503	100
1990	1610	22	5239	73	332	5	7181	100
1991	1441	26	3895	70	208	4	5544	100
1992	3629	36	5892	59	483	5	10004	100
1993	40746	38	61044	57	5854	5	107644	100
1994	34544	32	66365	61	7501	7	108410	100
1995	29919	27	69551	63	10903	10	110373	100
1996	30829	28	68101	62	11341	10	110271	100
1997	32032	30	64091	59	12159	11	108282	100
1998	37393	35	57635	53	13162	12	108190	100
1999	34013	32	51910	49	19228	18	105151	100
2000	26411	25	49846	48	27497	27	103754	100
2001	32073	30	50233	47	25301	24	107607	100
2002	34141	31	47899	44	26401	24	108441	100
2003	33777	31	46515	43	28060	26	108352	100
2004	27925	26	46571	44	31495	30	105991	100
2005	21627	20	49655	47	35115	33	106397	100
2006	18296	17	48738	45	40213	37	107247	100
2007	17322	16	47149	43	45503	41	109974	100
2008	23640	20	49751	43	42122	36	115513	100
2009	34664	29	53333	45	29988	25	117985	100
2010	35564	30	59601	51	22457	19	117622	100
N	555,657		1,012,443		435,975		2,004,075	

Table 2.2: Percentage of Data by Debt-to-Income Category

Year	LTV > 80 &/or FICO < 660	LTV > 90 &/or FICO < 620	LTV > 95 &/or FICO < 580	Other	Total	LTV > 80 &/or FICO < 660	LTV > 90 &/or FICO < 620	LTV > 95 &/or FICO < 580	Other	Total
	No.	No.	No.	No.	No.	Row %	Row %	Row %	Row %	Row %
1986	17110	4856	829	94679	117474	15	4	1	81	100
1987	20642	3953	1283	90114	115992	18	3	1	78	100
1988	24109	4858	1709	86908	117584	21	4	1	74	100
1989	19176	5485	1278	91915	117854	16	5	1	78	100
1990	18449	7318	902	91450	118119	16	6	1	77	100
1991	16591	7925	851	93489	118856	14	7	1	79	100
1992	16159	7338	815	94116	118428	14	6	1	79	100
1993	18236	11053	2034	85955	117278	16	9	2	73	100
1994	19235	15905	2926	79241	117307	16	14	2	68	100
1995	17849	15777	5217	77885	116728	15	14	4	67	100
1996	22235	16798	6334	70524	115891	19	14	5	61	100
1997	22354	15807	5990	70967	115118	19	14	5	62	100
1998	20816	15439	6851	73408	116514	18	13	6	63	100
1999	21936	14756	8220	69955	114867	19	13	7	61	100
2000	25262	15570	7855	65574	114261	22	14	7	57	100
2001	21897	15565	8357	69930	115749	19	13	7	60	100
2002	19101	13420	8090	75838	116449	16	12	7	65	100
2003	16929	11222	8634	79515	116300	15	10	7	68	100
2004	16663	8945	9161	80019	114788	15	8	8	70	100
2005	16475	7479	9174	82961	116089	14	6	8	71	100
2006	16418	7275	11947	81216	116856	14	6	10	70	100
2007	18299	10395	16883	71484	117061	16	9	14	61	100
2008	18940	10440	5872	81045	116297	16	9	5	70	100
2009	11609	5466	1371	100023	118469	10	5	1	84	100
2010	11674	5756	2415	98288	118133	10	5	2	83	100
N	468,164	258,801	134,998	2,056,499	2,918,462					

Table 2.3: Fraction of Risky Mortgages

	FICO			Income			Age		
Origination Year	N	Mean	Sd	N	Mean	Sd	N	Mean	Sd
1986	2300	682	91	697	6408	19679	764	44	16
1987	3785	696	87	1268	5230	4601	1393	44	16
1988	5099	704	79	6229	5147	4306	6433	40	13
1989	3984	698	81	8478	5716	10223	8925	39	11
1990	3839	699	76	8287	5640	4560	9039	40	12
1991	3413	694	82	6541	5404	4496	6756	40	13
1992	5069	712	77	13162	5619	5413	12542	42	12
1993	8320	705	84	110089	5469	5370	98897	42	11
1994	15222	709	75	110977	5143	4418	102663	42	12
1995	13819	692	80	112668	5056	4220	102635	42	12
1996	90529	710	62	112003	5282	5094	100619	42	12
1997	112709	713	60	110154	5588	5907	98961	43	12
1998	114702	718	58	110855	5940	4819	101080	43	12
1999	113028	715	60	108959	6028	5465	100393	43	12
2000	112222	707	63	108130	6023	5315	98785	43	12
2001	114644	715	64	109994	6590	5902	103265	43	12
2002	115791	721	60	112666	6865	6833	105766	43	12
2003	115850	725	57	113348	6962	6785	106167	44	13
2004	114248	722	58	110538	6871	9293	97527	44	13
2005	115698	723	59	110864	7150	6059	96790	44	13
2006	116585	719	61	111057	7603	7473	100114	44	13
2007	116913	718	63	112573	7747	6665	108099	44	13
2008	116183	742	53	115720	8690	7931	106804	45	13
2009	118285	763	41	118178	9298	9091	106372	46	13
2010	117984	766	41	117737	9899	9219	105611	47	13

Table 2.4: Summary Statistics by Year

	FICO			Income			Age		
LTV Category	N	Mean	Sd	N	Mean	Sd	N	Mean	Sd
LTV $\leq$ 80%	1334225	733	56	1546528	7191	7396	1416242	45	12
80% < LTV $\leq$ 90%	198632	713	58	249593	6075	5126	228678	40	11
90% < LTV $\leq$ 110%	237364	686	72	265051	5101	3305	241480	36	11

Table 2.5: Summary Statistics by LTV Category

	Occupancy Status Code							
	First Home		Second Home		Investment Home		Total	
ORIGTN_YEAR	No.	Row %	No.	Row %	No.	Row %	No.	Row %
1986	113074	96	1528	1	2872	2	117474	100
1987	111178	96	1804	2	3010	3	115992	100
1988	111564	95	2576	2	3444	3	117584	100
1989	112240	95	2786	2	2828	2	117854	100
1990	112996	96	2420	2	2703	2	118119	100
1991	114032	96	2243	2	2581	2	118856	100
1992	113718	96	1580	1	3130	3	118428	100
1993	111455	95	1864	2	3959	3	117278	100
1994	109957	94	2110	2	5240	4	117307	100
1995	110765	95	2320	2	3643	3	116728	100
1996	109912	95	2408	2	3571	3	115891	100
1997	108809	95	2625	2	3684	3	115118	100
1998	110198	95	2766	2	3550	3	116514	100
1999	106926	93	3038	3	4903	4	114867	100
2000	104324	91	3174	3	6763	6	114261	100
2001	106477	92	3238	3	6034	5	115749	100
2002	105204	90	3971	3	7274	6	116449	100
2003	104719	90	4732	4	6849	6	116300	100
2004	102076	89	5344	5	7368	6	114788	100
2005	100619	87	5865	5	9605	8	116089	100
2006	101099	87	5902	5	9855	8	116856	100
2007	102664	88	5413	5	8984	8	117061	100
2008	100817	87	6240	5	9240	8	116297	100
2009	104374	88	7397	6	6698	6	118469	100
2010	101437	86	7696	7	9000	8	118133	100
N	2,690,634		91,040		136,788		2,918,462	

Table 2.6: Occupancy Status by Year

Table 2.7: Structural Leverage Demand Estimation (2sls)

	LTV - Purchases	LTV - Refinances
Loan Original Note Rate	-3.0725	-3.90
	(0.7753)	(0.8308)
15 YR FRM	-13.6819	-12.0273
	(0.3527)	(0.4015)
20 YR FRM	-6.8572	-3.7932
	(0.2168)	(0.1334)
25 YR FRM	-3.0843	-1.3910
	(0.3930)	(0.1733)
40 YR FRM	5.8035	4.4890
	(0.3627)	(0.4613)
Borrowers Total Monthly Income Amount	-8.28e-05	-2.90e-05
	(7.84e-06)	(5.31e-06)
Borrowers Count	-1.7511	-0.2080
	(0.0641)	(0.0659)
$25 \leq$ Borrower's Age ≤ 34	-1.5857	0.2115
	(0.0890)	(0.2104)
$35 \leq$ Borrower's Age ≤ 49	-5.8772	-4.4202
	(0.0961)	(0.2079)
$50 \leq$ Borrower's Age ≤ 64	-10.0289	-9.3643
	(0.1009)	(0.2078)
Borrower's Age > 64	-13.7921	-15.2030
	(0.1384)	(0.2244)
Borrower Credit Score	-0.0595	-0.0511
	(0.0018)	(0.0019)
MSA Ln House Price Level	-5.7276	-10.3784
	(0.0629)	(0.0902)
MSA Detrended Ln HPI Qtrly Vol, 2 yr bk	-1.0511	-9.9837
	(0.6113)	(0.7568)
Second or Vacation Home	2.2555	0.5683
	(0.1113)	(0.2030)
Investment Property	2.7088	1.3346

Table 2.7: Structural Leverage Demand Estimation (2sls)

	LTV - Purchases	LTV - Refinances
	(0.4179)	(0.3681)
Female	-0.7518	-1.1664
	(0.0466)	(0.0551)
American Indian/Alaskan Native	1.6672	1.6204
	(0.3204)	(0.3440)
Asian/Pacific Islander	0.3042	2.9806
	(0.0828)	(0.1040)
Black (and not Hispanic)	5.2097	4.2897
	(0.1212)	(0.1254)
Hispanic	4.0734	3.0021
	(0.1122)	(0.0954)
Other Race	0.3783	1.1764
	(0.0988)	(0.1023)
1987	-1.7117	-0.6123
	(4.3005)	(4.5062)
1988	-4.8607	-3.9173
	(4.3216)	(7.5002)
1989	1.9601	2.5111
	(4.4727)	(4.5786)
1990	7.3831	-5.3855
	(4.1008)	(5.8132)
1991	0.3751	-0.2471
	(3.6473)	(4.0221)
1992	1.8482	-1.9157
	(2.7883)	(3.0130)
1993	1.1479	-3.1531
	(1.9334)	(2.1004)
1994	2.5485	-3.0586
	(2.2595)	(2.2396)
1995	5.1954	0.0939
	(2.2484)	(2.5327)

Table 2.7: Structural Leverage Demand Estimation (2sls)

	LTV - Purchases	LTV - Refinances
1996	7.8182	3.6810
	(2.4842)	(2.5702)
1997	6.8279	3.7182
	(2.2896)	(2.4734)
1998	5.4118	2.6810
	(1.7118)	(1.8958)
1999	6.5646	3.2855
	(1.9880)	(1.9864)
2000	8.4343	6.5932
	(2.5683)	(2.7942)
2001	6.1209	4.3877
	(1.7229)	(1.7841)
2002	5.0041	1.4278
	(1.4158)	(1.3682)
2003	3.1811	-2.7925
	(0.8087)	(0.7697)
2004	2.7177	-2.1865
	(0.9025)	(0.8374)
2005	2.7608	-0.9531
	(0.9024)	(0.9478)
2006	5.1577	2.5185
	(1.3332)	(1.4415)
2007	7.8040	4.1488
	(1.2961)	(1.4096)
2008	5.9641	2.4414
	(1.0580)	(1.0295)
2009	0.1781	-1.5372
	(0.2508)	(0.2558)
Constant	216.7163	261.9842
	(5.5240)	(6.2962)
R^2	0.22	0.24

Table 2.7: Structural Leverage Demand Estimation (2sls)

	LTV - Purchases	LTV - Refinances
N	511,448	524,826

Table 2.8: Reduced Form Estimates with CLL (Purchases)

	Note Rate	LTV	Ln Loan	Ln Price
15 YR FRM	-0.4400	-12.3299	-0.2618	-0.0503
	(0.0028)	(0.0914)	(0.0039)	(0.0039)
20 YR FRM	-0.0687	-6.6461	-0.1982	-0.0974
	(0.0086)	(0.2041)	(0.0060)	(0.0060)
25 YR FRM	0.0282	-3.1710	-0.1614	-0.1154
	(0.0170)	(0.3841)	(0.0112)	(0.0113)
40 YR FRM	0.0702	5.5879	0.0693	-0.0051
	(0.0139)	(0.3498)	(0.0109)	(0.0109)
Borrowers Total Monthly Income Amount	-3.19e-06	-7.30e-05	2.28e-05	2.36e-05
	(2.61e-07)	(6.73e-06)	(1.84e-06)	(1.90e-06)
Borrowers Count	-0.0636	-1.5555	0.1416	0.1662
	(0.0015)	(0.0402)	(0.0039)	(0.0040)
$25 \leq$ Borrower's Age ≤ 34	-0.0452	-1.4470	0.2043	0.2187
	(0.0038)	(0.0797)	(0.0037)	(0.0038)
$35 \leq$ Borrower's Age ≤ 49	-0.0596	-5.6941	0.2416	0.3181
	(0.0038)	(0.0818)	(0.0053)	(0.0054)
$50 \leq$ Borrower's Age ≤ 64	-0.0428	-9.8975	0.1206	0.2692
	(0.0040)	(0.0931)	(0.0051)	(0.0052)
Borrower's Age > 64	-0.0007	-13.7898	-0.0206	0.1991
	(0.0046)	(0.1365)	(0.0042)	(0.0041)
Borrower Credit Score	-0.0023	-0.0525	0.0004	0.0011
	(0.0000)	(0.0004)	(0.0000)	(0.0000)
MSA Ln House Price Level	-0.0353	-5.6190	0.4898	0.5764
	(0.0019)	(0.0552)	(0.0041)	(0.0043)
MSA Detrended Ln Qtrly HPI Vol, 2 yr bk	0.3739	-2.1997	-0.2547	-0.2267
	(0.0182)	(0.5291)	(0.0196)	(0.0198)
Ln Natl Conforming Loan Limit (CLL)	-4.0568	12.4648	0.5768	0.4483
	(0.2300)	(3.0115)	(0.0627)	(0.0660)
Second or Vacation Home	0.0628	2.0626	-0.2115	-0.2576
	(0.0033)	(0.0979)	(0.0095)	(0.0099)
Investment Property	0.5307	1.0782	-0.5436	-0.5808

Table 2.8: Reduced Form Estimates with CLL (Purchases)

	Note Rate	LTV	Ln Loan	Ln Price
	(0.0031)	(0.0700)	(0.0069)	(0.0071)
Female	0.0057	-0.7692	-0.0768	-0.0640
	(0.0017)	(0.0455)	(0.0022)	(0.0023)
American Indian/Alaskan Native	0.0269	1.5847	0.0048	-0.0188
	(0.0121)	(0.3144)	(0.0098)	(0.0095)
Asian/Pacific Islander	-0.0441	0.4396	0.0786	0.0608
	(0.0026)	(0.0743)	(0.0026)	(0.0027)
Black (and not Hispanic)	0.1099	4.8720	-0.0495	-0.1142
	(0.0042)	(0.0837)	(0.0036)	(0.0037)
Hispanic	0.1115	3.7307	-0.0920	-0.1450
	(0.0030)	(0.0698)	(0.0034)	(0.0035)
Other Race	-0.0026	0.3863	0.0123	0.0048
	(0.0034)	(0.0968)	(0.0031)	(0.0031)
Year Dummies	yes	yes	yes	yes
Constant	59.5546	33.7326	-2.0965	-1.9045
	(2.9770)	(38.9694)	(0.8163)	(0.8601)
R^2	0.81	0.24	0.48	0.52
N	511,448	511,448	511,448	511,448

Table 2.9: Reduced Form Estimates with CLL (Refinances)

	Note Rate	LTV	Ln Loan
15 YR FRM	-0.4794	-10.1575	-0.2884
	(0.0015)	(0.0510)	(0.0027)
20 YR FRM	-0.1193	-3.3278	-0.1661
	(0.0027)	(0.0880)	(0.0025)
25 YR FRM	-0.0001	-1.3906	-0.0998
	(0.0056)	(0.1705)	(0.0047)
40 YR FRM	0.0514	4.2884	0.0928
	(0.0157)	(0.4484)	(0.0123)
Borrowers Total Monthly Income Amount	-2.21e-06	-2.03e-05	2.22e-05
	(2.86e-07)	(4.22e-06)	(2.48e-06)
Borrowers Count	-0.0569	0.0139	0.1078
	(0.0015)	(0.0452)	(0.0047)
$25 \leq$ Borrower's Age ≤ 34	-0.0455	0.3891	0.1410
	(0.0083)	(0.2030)	(0.0066)
$35 \leq$ Borrower's Age ≤ 49	-0.0428	-4.2532	0.1536
	(0.0082)	(0.2008)	(0.0074)
$50 \leq$ Borrower's Age ≤ 64	-0.0108	-9.3221	0.0592
	(0.0083)	(0.2036)	(0.0070)
Borrower's Age > 64	0.0291	-15.3167	-0.1032
	(0.0085)	(0.2192)	(0.0067)
Borrower Credit Score	-0.0022	-0.0423	0.0001
	(0.0000)	(0.0004)	(0.0000)
MSA Ln House Price Level	-0.0791	-10.0698	0.4415
	(0.0018)	(0.0604)	(0.0048)
MSA Detrended Ln Qtrly HPI Vol, 2 yr bk	0.2105	-10.8046	-0.0291
	(0.0194)	(0.7271)	(0.0209)
Ln Natl Conforming Loan Limit (CLL)	-4.1578	16.2152	0.5485
	(0.2601)	(3.2398)	(0.0684)
Second or Vacation Home	0.0563	0.3487	-0.1298
	(0.0054)	(0.1947)	(0.0163)
Investment Property	0.4309	-0.3460	-0.4025

Table 2.9: Reduced Form Estimates with CLL (Refinances)

	Note Rate	LTV	Ln Loan
	(0.0032)	(0.0835)	(0.0075)
Female	0.0152	-1.2255	-0.0880
	(0.0017)	(0.0529)	(0.0029)
American Indian/Alaskan Native	0.0267	1.5164	-0.0321
	(0.0110)	(0.3392)	(0.0097)
Asian/Pacific Islander	-0.0192	3.0556	0.0953
	(0.0029)	(0.1015)	(0.0027)
Black (and not Hispanic)	0.0910	3.9347	-0.0937
	(0.0037)	(0.0981)	(0.0034)
Hispanic	0.0481	2.8145	-0.0852
	(0.0028)	(0.0850)	(0.0051)
Other Race	0.0077	1.1464	-0.0059
	(0.0030)	(0.1007)	(0.0027)
Year Dummies	yes	yes	yes
Constant	61.3190	22.8425	-0.7726
	(3.3661)	(41.9260)	(0.8932)
R^2	0.84	0.26	0.51
N	524,826	524,826	524,826

Table 2.10: Reduced Form (RF) vs Structural Year Dummies

	Note Rate - RF w/o CLL	LTV - RF w/o CLL	LTV - Structural
1987	-0.4464	-0.3400	-1.7117
1988	-0.4279	-3.5460	-4.8607
1989	-0.2160	2.6238	1.9601
1990	-0.1165	7.7410	7.3831
1991	-0.4681	1.8135	0.3751
1992	-1.3147	5.8876	1.8482
1993	-2.2114	7.9424	1.1479
1994	-1.7609	7.9588	2.5485
1995	-1.7598	10.6024	5.1954
1996	-1.4251	12.1967	7.8182
1997	-1.6763	11.9782	6.8279
1998	-2.4226	12.8552	5.4118
1999	-2.0661	12.9127	6.5646
2000	-1.3167	12.4798	8.4343
2001	-2.4084	13.5208	6.1209
2002	-2.8053	13.6236	5.0041
2003	-3.5929	14.2203	3.1811
2004	-3.4723	13.3865	2.7177
2005	-3.4720	13.4285	2.7608
2006	-2.9124	14.1060	5.1577
2007	-2.9602	16.8992	7.8040
2008	-3.2684	16.0064	5.9641
2009	-4.3268	13.4723	0.1781

Chapter 3

Loss Aversion and Anchoring in Commercial Real Estate Pricing: Empirical Evidence and Price Index Implications

3.1 Introduction

This is an empirical paper examining the role of two psychological theories in the marketplace. The first is prospect theory, often suggested as an alternative paradigm to supersede the utility theory of neoclassical economics. It is based on the concept of *loss aversion*, i.e. for equal sized gains and losses around a reference point, individuals give up more utility for the loss than they receive from the gain. Such a reference dependent preference function was not recognized in classical economic theory. Today, despite considerable laboratory evidence in favor of prospect theory, some economists still believe that loss aversion is merely the result of a mistake made by inexperienced individuals and through time they will learn, and their behavior will more closely match predictions from neoclassical models. In the context of the current literature,

[0]Co-authored with David Geltner

the present paper confirms, extends, and enhances the previous evidence regarding loss aversion, including a very influential 2001 paper by Genesove & Mayer which found loss aversion behavior in the Boston housing market of the 1990s.

In this paper, we use market data based on all U.S. sales of commercial property greater than $5,000,000 from January 2001 through December 2009. We find that loss aversion plays a significant role in the behavior of investors in commercial real estate. We thus extend the Genesove-Mayer findings to the commercial property market where the participants are "professionals" operating in a more purely business environment (compared to homeowners). Furthermore, contrary to common belief and some prior evidence, we find the degree of loss aversion to be actually higher the more sophisticated or experienced the investor is.[1]

The second piece of theory tested in this paper is known as the anchoring-and-adjustment heuristic. Specifically, an asking price could serve as an anchor or heuristic used by a buyer to judge the value of a property, and they may not be able to adjust sufficiently away from the anchor to arrive at what would otherwise be a fair market price. As a result, real estate could be mis-priced if sellers play to this behavior by buyers. We find that there is considerable evidence for the predictions of this theory in the marketplace.

A feature of the present paper is that we develop longitudinal price indices of U.S. commercial property that control for and reflect both the anchoring and the prospect theory phenomena. We show that explicitly including these behavioral factors can greatly improve the construction of a traditional hedonic price index. We also combine our behavioral pricing model with price indices to demonstrate the magnitude and nature of the effect of the behavioral pricing phenomena during the dramatic commercial property market cycle of the 2000s.

While we study both loss aversion and anchoring, the focus of the present paper is primarily on loss aversion. One hypothesis has been that loss aversion may play a

[1]Although the motivation or cause of this behavior is beyond the scope of the present paper, discussions we have had with participants in the commercial property market suggest to us that behavioral phenomena among professional investors in that market may be due to the reluctance on the part of agents to realize losses to the stakeholders in their companies. Further research is needed to document the decision-making process for the various investor types.

significant role in real estate's famously dramatic pro-cyclical variation in asset trading volume, causing property markets to be excessively illiquid during down markets. The hypothesis is that loss aversion could cause transaction prices of completed deals (those that are reflected in the market-wide average prices on which price indices are based) to be "sticky" or fail to register much of a drop during the early phase of a sharp downturn in the market, as compared to movements on the demand side of the market ("constant-liquidity" prices). Sticky pricing and related illiquidity certainly seemed to be present in the 2007-09 drop in the U.S. commercial property market. For example, in the early phase of that downturn from 2Q2007 through 1Q2008, the "TBI" index published by the MIT Center for Real Estate dropped 15% on the demand side while actually rising 2% on the supply side leading to only a 7% drop in consummated transaction prices, while trading volume of major commercial property assets fell during that same period from \$136 billion to \$48 billion.[2] However, there is a question how much of this "sticky pricing" behavior is due to prospect theory based "psychological" loss aversion, as distinct from more classical and rational explanations. With this in mind, we examine the impact of prospect theory at the aggregate market level, and find that while our models attest to the economic importance of loss aversion at the individual property level, they suggest that psychological loss aversion had a relatively small impact on overall average transaction prices (and therefore on trading volume) during the recent market peak and downturn. Most of the "sticky pricing" behavior was either explainable by classical explanations, or may be due to other behavioral phenomena besides psychological loss aversion not examined in the current paper.

The rest of the paper is organized as follows. We begin with an overview of prospect theory and the anchoring-and-adjustment heuristic. We then develop an empirical model to test these theories. We then describe the data source and high-

[2]These volume numbers are based on sales of assets greater than \$5,000,000 as reported by Real Capital Analytics Inc (RCA), the datasource used for this study. The TBI demand and supply indices employ the Fisher et al (2003) methodology based on NCREIF property sales to measure movements in reservations prices on the two sides of the market. Hence, the implication is that property owners (the supply side) actually raised their willing-to-receive prices by 2% while potential property buyers dropped their willing-to-pay prices by 15%, resulting in the huge drop in transaction volume.

light some of the features of the data used in this study. The empirical results are then presented in Section 3.5. Finally our analysis of the implications for commercial property price indices and the aggregate market-wide effects of loss aversion are presented in Section 3.6. A final section then concludes the paper with some finishing remarks.

3.2 Prospect Theory and the Anchoring and Adjustment Heuristic

3.2.1 Prospect Theory

Prospect theory, which helped gain a Nobel Prize in economics for Daniel Kahneman in 2002, is characterized by three essential features [Kahneman and Tverksy (1979); Tversky and Kahneman (1991)]. First, gains and losses are examined relative to a reference point. Second, the value function is steeper for losses than for equivalently sized gains. Third, the marginal value of gains or losses diminishes with the size of the gain or loss. Thus, under prospect theory, people behave as if maximizing an S-shaped value function as shown in figure 3-1.

A difficulty in applying prospect theory to empirical studies is that the reference point is seldom observed in the data. An influential exception has been Genesove and Mayer's 2001 study (hereafter "G-M"), examining seller behavior in the Boston housing market using the home purchase price as the reference point. They find evidence that loss aversion explained the behavior of condominium sellers in their choices of asking prices and in their decisions as to whether to accept an offer or not. They find that property owners, faced with a prospective loss, set a higher asking price and in fact do sell at a higher price than other sellers, suffering as a result less sale frequency or, in effect, a longer time on the market. The first contribution of the current study is our attempt to replicate G-M using data on U.S. commercial real estate instead of housing. Thus, a key aspect of the present paper is that it examines the evidence on loss aversion among sellers that are primarily investors (as opposed to being owner-occupiers who are arguably primarily consumers). A commonly held view is that property owners have a sentimental attachment to their homes, as well as being not full-time "in the business" of real estate investment, and as a result could be overly optimistic or overly influenced by emotions in their listing and sales behavior. Therefore, owner-occupants may understandably behave in a loss-averse manner while it remains unclear if commercial property investors, who typically should not have sentimental motivations, would behave in a similar way.

Under prospect theory, a seller with a potential loss compared to his purchase price would be expected to set a higher reservation price than a seller with a prospective gain. The former can avoid or mitigate loss by setting a sufficiently high reservation price and sticking with it until trade goes through. To formalize this intuition, consider the following simple model.

Assume that the utility from sale, $U(P)$ is increasing in price ($U(P)' > 0$) and greater than the utility from no sale ($U(P) > U_0$), for all relevant prices. Also assume that the probability of a sale is decreasing in price ($prob(P)' < 0$). The seller would then maximize their expected utility from a sale by choosing a reservation price P:

$$max_P \quad prob(P) * U(P) + (1 - prob(P)) * U_0$$

$$U(P)' * prob(P) = prob(P)' * (U(P) - U_0)$$

The above first order condition states that the seller would set a price so as to equate (in expectations) the marginal gain from an increase in price to its marginal cost. Next, we examine the behavior of a loss-averse seller compared to that of a risk-neutral seller. This can be illustrated with a simple reference-dependent utility function where the reference point is the price that the seller first paid for the property, P_f

$$\begin{aligned} U(P) &= (P - P_f) \quad \text{if } P >= P_f \\ &= \lambda * (P - P_f) \quad \text{if } P < P_f \end{aligned}$$

where $\lambda > 1$ is the loss aversion parameter. The first-order conditions can then be written as:

$$\begin{aligned} prob(P) &= -prob(P)' * ((P - P_f) + U_0) \quad \text{if } P >= P_f \\ \lambda * prob(P) &= -prob(P)' * (\lambda * (P - P_f) + U_0) \quad \text{if } P < P_f \end{aligned}$$

Figure 3-2 below shows the difference between the loss-averse seller and risk-neutral seller ($\lambda = 1$), where the prior price (reference point) P_f is at a hypothetical value of 50. We make the following points about the behavior of these two types of sellers. When the market value is greater than the purchase price ($P \geq P_f$), there is no effect of loss aversion. There is bunching at $P = P_f$, and when $P \leq P_f$, the marginal benefit and marginal cost from an increase in price disproportionately increase for the loss-averse agent compared to the risk-neutral seller. Thus, when faced with a loss, a loss-averse seller (as compared to the risk-neutral seller) would find it optimal to set a higher price since for that seller the difference between the marginal benefit and the marginal cost of an increase in price is greater at every price level.

The simple prospect theory based value function formalized above and illustrated in figure 3-2 allows us to propose a finer and more rigorous distinction in the definition of loss aversion behavior. In particular, it is only the behavior of agents who are to the left of the prospect theory value function "kink-point" in figure 3-1, or to the left of the reference price of 50 in figure 3-2, who are in a position to exhibit the sort of psychological loss aversion behavior that is distinguished by prospect theory. This "behavioral loss aversion" will be the major focus of the current paper, and we will identify it by looking for empirical evidence of asymmetric pricing behavior between sellers facing a loss (those to the left of the reference point) versus those facing a gain. More broadly, however, agents may display other types of strategic pricing behavior based on a reference point, possibly including the prior purchase price that we consider as the reference point in our model. Some such behavior could be "rational" (or consistent with classical utility theory, not reflective of prospect theory).[3] But asymmetric pricing behavior around the reference point is nevertheless of interest in

[3]For example, rational attempts to avoid realizing a loss relative to an outstanding mortgage balance could be correlated with loss relative to the prior purchase price.

understanding the behavior of market participants and how the commercial property market functions. Nor will we ignore some broader aspects of reference point based pricing behavior, as we will also consider both the symmetric as well as asymmetric component of the reference point's impact on sellers' pricing strategy.

These considerations highlight another behavioral phenomenon that is closely related to loss aversion and predicted by Prospect Theory, known as the *Disposition Effect* - the tendency to sell winners quickly and hold on to losers. This phenomenon has been documented extensively in the finance literature (see, for example, Odean (1998), Feng et al (2005), Locke et al (2000) and Shapira et al (2001)). In the real estate literature, Crane and Hartzell (2009) find evidence for the disposition effect in REITs. They find that managers of REITs are more likely to sell properties that have performed well and accept lower prices when selling profitable investments. In a relevant paper not focused on behavioral factors, Fisher et al (2004) find that there is a greater likelihood of a sale following increases in the national index of commercial real estate returns and for properties that have outperformed that index. But to date, and unlike the case with housing and the G-M study, there has been no empirical documentation of loss aversion behavior *per se* among commercial property market participants in general.

Loss Aversion and Experience

If loss aversion was a fundamental and stable component of preferences as advocated by prospect theory, then it must be the case that the market experience of an individual and loss aversion would be uncorrelated. For instance, if an investor with little experience behaved in a loss-averse manner (during a down market), then that same investor once he has gained experience would behave in the same fashion in a similar situation.

G-M's study of the Boston housing market in the 1990s found that investors in condominiums were less loss-averse than their owner-occupant counterparts. Presumably, condominium sellers are more experienced in the market than homeowners.[4] List

[4]This would be because condominiums are better suited to investment trading than houses, and

(2003) is an example of a recent experimental field study that also supports the notion that loss aversion can be attenuated with market experience. Examining trading rates of sport memorabilia in an actual marketplace, List observed an inefficiently low number of trades by naive traders, consistent with prospect theory.

On the other hand, there is some evidence that even sophisticated traders are sometimes subject to behavioral biases. Haigh and List (2005) provide experimental evidence that CBOT traders are more likely to suffer from myopic loss aversion[5] than student participants. The 2009 Crane-Hartzell study suggests that even experienced REIT managers can exhibit loss aversion behavior.

In this paper, we shed new light on this question by studying the degree of loss aversion across different types of investors as well as among groups of investors that have significant differences in trading experience.

3.2.2 Anchoring-and-Adjustment

Besides prospect theory, psychological anchors could also affect the valuation of real estate. Specifically, an asking price could serve as an anchor or heuristic used by a buyer to judge the value of a property, and they may not be able to adjust sufficiently away from the anchor to arrive at a rational market value[6]. In the context of housing, Northcraft and Neale (1987) took local real estate agents to a house and asked them to appraise it. Each group of agents was given the same information packet about the house that they could use to appraise the property. However, a key difference was that different groups had been given different asking prices. The appraised values turned out to be positively related to the provided anchor, the asking price. Interestingly,

the G-M study incorporated a period of a condominium boom in Boston, attracting considerable speculative investment in the market.

[5]Myopic loss aversion is a term first coined by Benartzi and Thaler (1995) that combines the concepts of loss aversion and mental accounting (see Thaler (1985)) to provide an explanation for the equity-premium puzzle in the stock market. A myopically loss-averse agent would tend to make shorter-term choices and evaluate losses and gains more frequently.

[6]The anchoring heuristic was first demonstrated by Tversky and Kahnemann (1974) in an experiment where they asked participants to estimate a number such as the percentage of African countries that are members of the United Nations. The experiment began with the subjects being given a number (between 1 and a 100) generated by the spin of a wheel. It turned out that the subjects showed a bias in their final estimates toward the number they were originally given.

most participants reported that the asking price should be irrelevant to the appraised value, yet they were nonetheless influenced by it. It should be noted that in that study, group differences in the appraisals could not be explained by individual differences in appraisals methods alone.

In the present study, a hypothesis generated by the Northcraft and Neale study is that any over- or under-pricing (i.e. the extent to which the asking price is above or below the expected sale price) by a seller could influence a buyer's valuation and thus have an effect on the subsequent transaction price. Black and Diaz (1995) tested this hypothesis in a laboratory experimental setting and found that manipulated asking prices influenced both the buyer's opening offer and the eventual transaction price, indicating a strong anchoring effect of the asking price.

However, it is important to note that another theory common in the urban economics literature makes a similar prediction based on neoclassical ("rational" rather than "behavioral") economics. Yavas and Yang (1995) propose a game theoretic model in which they argue that a seller strategically lists an asking price that reflects his bargaining power in an attempt to signal to certain types of buyers. For instance, a seller who can wait for a high-paying buyer may post a high asking price to attract only those buyers that would value his property higher than the going market value. This is clearly a different behavior than anchoring, which would say that *any* asking price influences the valuation of *all* buyers (although only buyers whose influenced valuations are sufficiently high would come forth to negotiate with the seller).

Another type of signaling behavior in the seller's asking price that would be rational and not inconsistent with classical economic theory would be for the seller to use his asking price to signal private information that he has about the (true market) value of the property. Properties are unique, and no one knows the property as well as its current owner.

Due to the nature of the market data used in this study, which reflects the results from the interactions between a buyer and a seller and our inability to observe their respective bargaining powers, we cannot distinguish between "irrational" psychologically based anchoring behavior versus "rational" signaling behavior such as that described above. However, we are able to test their joint predictions.

3.3 Empirical Model

In this section we develop an empirically testable model that reflects the prospect theory and anchoring phenomena described above. The model developed here is similar to that employed by G-M, but extends and enhances their model by explicitly incorporating the prospect theory reference point in the value function.[7]

To test for prospect theory, the structural model specifies that the log asking price, L is a linear function of the expected log selling price in the quarter of entry (when the property is put up for sale), labeled μ, and a variable defining the reference point, $RF*$:

$$L_{ife} = \alpha_0 + \alpha_1 \mu_{ie} + mRF^*_{ife} + \epsilon_{ie} \tag{3.1}$$

Here i indicates the unit, f the quarter of previous or first sale and e the quarter of entry into the market. If there were no behavioral effects: m=0. Furthermore, the expected log selling price is a linear function of a vector of observable attributes of the property ($X_i\beta$), the quarter of entry Q_e and unobservable quality, ν. The unobserved quality is observable to a buyer or a seller but not to the analyst:

$$\mu_{ie} = X_i\beta + Q_e + \nu_i \tag{3.2}$$

We define the reference-point variable, $RF*$ as the difference between the previous log selling price and the expected log selling price:

$$RF^*_{ife} = (P_{if} - \mu_{ie})$$

$RF^*{}_{ife}$ is therefore positive if there is an expected loss. Assuming that equation (3.2)

[7]One advantage of the model presented here compared to the G-M model is that we are able to estimate a single unbiased coefficient measuring loss aversion, whereas G-M were only able to produce upward and lower biased estimates which they used to provide a range.

holds for all periods, the previous log selling price can be written as:

$$P_{if} = \mu_{if} + w_{if} = X_i\beta + Q_f + \nu_i + w_{if}$$

where w_{if} is the over- or under- payment by the current seller to the previous seller at the time of the current seller's purchase.

Thus the true reference-point term is

$$RF^*_{ife} = (\mu_{if} + w_{if} - \mu_{ie}) = (Q_f - Q_e) + w_{ie} \tag{3.3}$$

The interpretation of the first term is the change in the market price index between the quarter of original purchase and the quarter of listing.[8] If $RF^* \leq 0$, then the seller faces a prospective gain but if $RF^* > 0$, then they face a prospective loss.

Combining equations (3.1), (3.2) and (3.3):

$$L_{ife} = \alpha_0 + \alpha_1(X_i\beta + Q_e + \nu_i) + m(Q_f - Q_e + w_{if}) + \epsilon_{ie} \tag{3.4}$$

The specification above cannot be estimated since w_{if} and ν_i are unobserved. However, we proceed by substituting a noisy measure of the reference-point variable in place of the true term:

$$L_{ife} = \alpha_0 + \alpha_1(X_i\beta + Q_e) + mRF_{ife} + \eta_{ie} \tag{3.5}$$

where

$$RF_{ife} = (P_{if} - X_i\beta - Q_e) = (Q_f - Q_e + \nu_i + w_{if}) \tag{3.6}$$

i.e. RF is estimated as the difference between the purchase price and the predicted selling price from a hedonic regression at the quarter of listing. Substituting (3.6)

[8] It is interesting to note that Pryke and Du Gay (2002) find, in their cultural study of the commercial real estate market in the UK after the crash in the late 1980s, that there was a conscious effort by investors to evaluate the performance of their property relative to a market index. Such a phenomenon has also clearly been present in the U.S. since the mid-1990s (e.g., Geltner 2000).

into (3.5), we get:

$$L_{ife} = \alpha_0 + \alpha_1(X_i\beta + Q_e) + m(Q_f - Q_e + \nu_i + w_{if}) + \eta_{ie}$$

where

$$\begin{aligned} \eta_{ie} &= \alpha_1\nu_i + m((Q_f - Q_e + w_{if}) - (Q_e - Q_e + \nu_i + w_{if})) + \epsilon_{ie} \\ &= (\alpha_1 - m)\nu_i + \epsilon_{ie} \end{aligned}$$

Thus, ν_i is an omitted variable, correlated with the reference-point term, RF. Thus m is expected to be biased since RF is correlated with ν_i.

To address this omitted variable bias, we can add to our model the residual of the previous selling price, $\nu + w$, as a noisy proxy for unobserved quality, ν :

$$\begin{aligned} L_{ife} &= \alpha_0 + \alpha_1(X_i\beta + Q_e) + \alpha_1(P_{if} - X_i\beta - Q_e) + m(Q_f - Q_e + \nu_i + w_{if}) + u_{ie} \\ &= \alpha_0 + \alpha_1 X_i\beta + \alpha_1 Q_e + \alpha_1(\nu_i + w_{if}) + m(Q_f - Q_e + \nu_i + w_{if}) + u_{ie} \qquad (3.7) \end{aligned}$$

The residual u_{ie} now contains the following terms:

$$\begin{aligned} u_{ie} &= \alpha_1\nu_i + m((Q_f - Q_e + w_{if}) - (Q_f - Q_e + \nu_i + w_{if})) - (\alpha_1 - m)(\nu_i + w_{if}) + \epsilon_{ie} \\ &= (m - \alpha_1)w_{if} + \epsilon_{ie} \qquad (3.8) \end{aligned}$$

Expanding and rewriting equation (3.7) as:

$$\begin{aligned} L_{ife} &= \alpha_0 + \alpha_1 X_i\beta + \alpha_1 Q_e + \alpha_1(\nu_i + w_{if}) + m(Q_f - Q_e + \nu_i + w_{if}) + (m - \alpha_1)w_{if} + \epsilon_{ie} \\ &= \alpha_0 + \alpha_1(X_i\beta + Q_e + \nu_i) + m(Q_f - Q_e + w_{if}) + \epsilon_{ie} \qquad (3.9) \end{aligned}$$

We can see that equation (3.9) is equivalent to equation (3.4) and thus our specification is fully identified and estimable. In section 3.5 where we estimate this model, the reference point variable is broken into two components representing prospective gain or loss. We would expect that the coefficient m on the loss component would be positive (significantly different from 0) and higher in magnitude as well as significantly different from the coefficient on the gain component. Such a result would confirm the presence of loss aversion in the market.[9]

In order to test for anchoring/signaling effects, we have to switch focus from the listing price to the transaction price. Equation (3.9) above is a listing price regression and its residuals represent the extent to which the asking price is above or below the average or typical asking price, after taking into account any possible effects of loss aversion. Thus, in order to test for the effect of the asking price on the eventual sale of the property, the residuals from equation (3.9) are used as a right-hand side variable in a hedonic regression on the achieved sale price of a property. In section 3.5 where we estimate this model, these residuals are referred to as the Anchoring/Degree of Over-Pricing. The coefficient on this variable would have to be significantly different from 0 in order to make any conclusive statement about the presence of psychological anchoring and/or signaling in the marketplace.

3.4 Data

The sales data used in this study comes from Real Capital Analytics (RCA), a New York based firm that is widely used to provide commercial property transactions data among institutional investment firms in the U.S. RCA attempts to collect price and other information about all commercial property sales in the U.S. of greater than \$5,000,000 in price. RCA estimates that they achieve at least 90 percent coverage of that sales population. The sample period is from January 2001 until December 2009. This time period includes the largest and most dramatic rise and fall in the U.S. commercial property market at least since the Great Depression, and therefore provides an ideal sample for the present study.

The dataset covers all four "core" investment property sectors (usage types: of-

[9]Note that by comparing the magnitude of the estimate of m on the sales with a loss with that of the estimate of m on sales with a gain, and considering only the *difference* between those two, we are focusing on a narrow and pure definition of loss aversion as an explicitly "behavioral" phenomenon of prospect theory, as sales that face a gain instead of a loss are "beyond the kinkpoint" of figure 3-1, that is, beyond the reference point of the prospect theory value function.

fice, retail, industrial and apartments) and has information on location and physical attributes. The raw data obtained from RCA consisted of about 100,000 commercial properties. For the purpose of this study, we discard properties that have incomplete information on property location, sales dates, listing dates as well as those with missing information on prices. To filter observations with a greater likelihood of error, we dropped properties that had a first sale before 1988. All properties that were not part of an arms length transaction are not included. Furthermore, to avoid any overstatement in the calculation of the market price appreciation, we exclude properties that were held for less than 1.5 years ("flipped" properties).[10] Finally, properties that sold as part of a portfolio (multiple property) transaction are filtered out, as it is not possible to determine each property's contribution to the portfolio's transaction price.

Table 3.1 provides a summary of the remaining data. There are 6,767 total listings of which 4,782 properties actually sold in the market. The other 1,985 properties were delisted or pulled from the market without a sale. Of the 4782 completed transactions, 3723 were sold at a gain, and 1059 at a loss.[11] All the properties have complete information on the first sale price and the asking price, the two key variables required to compute our empirical model. As Table 3.1 shows, about one-fourth of the total listings over the sample period faced a loss at the time of entering the market. Moreover, properties that sold spent less time on the market (37 weeks) than delisted, unsold properties.

To construct a variable for experience, we exploit the fact that the RCA data contains the names of buyers and sellers. The same investor is often both a buyer and a seller in the market. Thus we calculate trading experience by counting the number of times an investor's name has appeared in either the buyer or seller name

[10]Other filters were also imposed to ensure the integrity and appropriateness of the data. Details are available from the authors upon request. Examination the data suggests that the smaller analysis sample did not differ significantly from the larger raw data sample in several summary measures including average sale price and number of metro areas covered.

[11]This included 1186 properties sold during the 2007 peak of the market cycle, including (perhaps surprisingly) 234 properties sold at a loss (compared to their prior purchase price) during that peak year (suggesting the extent of heterogeneity or idiosyncratic risk at the asset level in commercial property). A further 1068 properties were sold during 2008-09 including 309 at a loss.

lists. The mean number of trades per seller is 101 for sold listings and slightly less at 88 for all listings, which include unsold, delisted properties.

The RCA data also contains information on the type of investor the seller is. There are primarily four groups of investors: "Institutional" (consisting of banks, insurance companies, pension and hedge funds; national and international entities who tend to purchase larger properties); "Private" (consisting of generally smaller and more local companies geared towards operating, developing or investing in commercial real estate); "Public" (consisting of companies that are listed on public markets like REITS and REOCs); and "Users" (consisting of owner-occupiers such as government, educational, and religious institutions or business that own commercial property for their own use). Of these four groups, the institutional investors and the publicly traded companies are the most experienced in the market which is reflected in our dataset as they make up the majority of the 100-plus trades investors over the given time period.

3.5 Empirical Analysis

3.5.1 Effects of Loss Aversion on Asking Prices

Table 3.2 presents our main empirical results of the test of loss aversion behavior. It shows the equation (3.9) regression of asking prices onto prospective gains and losses, as well as onto the estimated value of the property ($X_i\beta$), the residual from a first sale price regression so as to control for unobserved quality, and dummy variables for the quarter of entry into the market for sale (results for the latter omitted, available from the authors). All price variables are measured in logs. The results confirm that losses play a greater role than gains: the coefficient on Loss is higher and different with statistical significance compared to that on Gain. The estimated coefficient of 0.38 on Loss suggests that a 10 percent increase in a prospective loss (referenced on the seller's purchase price), leads the seller to set an asking price approximately

3.8 percent higher than she otherwise would.[12] In other words, commercial property sellers faced with a loss relative to their purchase price tend to post asking prices higher than otherwise-similar sellers not facing a loss, by a magnitude of about 38% of their loss exposure. The comparable finding in the G-M housing study was 25% to 35%.

It is important to note that it is the *difference* between the coefficient on "Loss" and that on "Gain" that suggests a type of psychological loss aversion based on prospect theory that is inconsistent with classical utility-based economic models. Referring back to figure 3-2, sellers facing a gain are beyond the kink-point in the prospect theory value function, and hence their pricing behavior presumably does not reflect prospect theory based loss aversion. However, the fact that sellers facing a gain also price differentially is still interesting from an economic perspective. It suggests a type of pricing strategy influenced by a reference point (in this case the property's prior purchase price). In particular, the positive coefficient on the "Gain" variable suggests that sellers facing a gain set a *lower* price than they otherwise would, while sellers facing a loss set a *higher* price, in the latter case, asymmetrically so. The asymmetry in this behavior likely reflects the "behavioral" loss aversion phenomenon of prospect theory. But even the symmetrical reference point based pricing behavior may involve "behavioral" components. For example, the pricing behavior discovered here would be consistent with the "sell winners/hold losers" behavior referred to as the "disposition effect", and found by Crane & Hartzell (2009) in their study of REIT behavior. We will discuss these issues further in the next section about participants' experience and in section 3.6 where we quantify the historical magnitude of these pricing strategies during the 2000s commercial property market cycle.[13]

[12] More precisely, this is a point elasticity based on log-differences, so the arc elasticity based on a 10% loss might be slightly different. The mean value of the Loss variable among sold properties with a loss was 0.32 which implies a price loss of 27% of the prior purchase price.

[13] We should note that loss aversion, and the disposition effect, may have "rational" components. However, Crane & Hartzell (2009) argue that the disposition effect they find in REITs cannot be explained by rational motivations such as mean reversion in asset prices. Asymmetric loss aversion such as we find here also can be rational if the seller has a mortgage whose balance exceeds the likely current value of the property. Unlike Genesove & Mayer who are able to control for this consideration, we do not have data on sellers' loan balances. However, we note that certain types of institutions typically rely less on property-level debt, including REITs, pension funds, and foreign

Finally, Table 3.2 also shows the coefficient on the residual from the first sale regression, which is a proxy for unobserved quality. This coefficient is positive and statistically significant, implying that controlling for unobserved property heterogeneity is important.

3.5.2 Loss Aversion and Experience

We next divide the data into two groups; investors that engaged in more than a hundred trades were labeled the "more experienced investors" group, and those with less than a hundred trades, the "less experienced investors" group. In Table 3.3, we find that there is no significant difference between the two groups when they are faced with a gain. However, contrary to previous findings in the literature, we find that the more experienced investor group exhibits a *higher* degree of loss aversion than their less experienced counterparts. A test of the equality of coefficients on "Loss - more experienced investors" and "Loss - less experienced investors", significantly rejects the null hypothesis that the two groups behave the same.

It is interesting to note that over 40 percent of the more experienced investors trading group consists of institutional investors. In Table 3.4, we compare the degree of loss aversion across different investor types and find that consistent with the findings in Table 3.3, the coefficient on Loss for institutional investors (0.485) is among the highest. It is not significantly different from equity fund investors (0.515), who also make large investments in commercial real estate. The next most experienced group is publicly traded companies, which, with a coefficient of 0.346 is more loss-averse than private investors (0.26), although the difference is not statistically significant.

The difference in the loss coefficient of private investors and that of institutional as well as equity fund investors is statistically significant, and we find this difference intriguing. Perhaps local knowledge that's available to private investors has a role in explaining this difference, cutting through the psychological behavioral tendency to indulge in loss aversion (possibly by giving such investors a greater self-confidence

investors, and as we will note in the next section, we find that such institutions exhibit even greater than average loss aversion pricing.

to sell at a loss recognizing that it does reflect the current true state of the local market). Or perhaps private investors tend to employ more property-level debt, and their creditors enforce a more ruthless business logic on their sales in the face of loss. Furthermore, large institutions tend to be "agents" managing the capital of other investors ("principals"). Agents may fear judgment by their stakeholders, making them more reluctant to realize losses. Also, investment management incentive fee structures, such as IRR hurdles and promotes, may incentivize agents to realize gains early, and perhaps to postpone loss recognition.

3.5.3 Evidence of Behavioral Effects on Transaction Prices

In this section, we turn our analysis to the final transaction price. It could be argued that since the results in the previous sections were not based on the selling price (they were based only on asking price), the anomalies exhibited by the sellers would disappear once they enter into a bargaining environment with the buyer. We test if loss aversion still plays a role in the final transaction price. We also test if the asking price has any influence on the sale price when the asking price is set above, or below, the market value (as predicted by the hedonic model). This analysis is achieved by taking the residual from the asking price equation (3.9) and including it in the final sale price regression. This residual will be positive and larger in cases where the asking price is higher than normal *relative* to the average asking price (controlling for other characteristics of the sale and effect of loss aversion), and vice versa. This residual from equation (3.9) is termed the "degree-of-overpricing" (DOP). It captures both the signaling aspect of the bargaining process as well as the psychological anchoring-and-adjustment process.

We present the results in Table 3.5 in two different ways. In column 1 of Table 3.5, we include the Loss variable in the same way as in earlier regressions. However, in column 2, we divide the Loss variable into three regimes; LossPre07 (Loss before 2007), Loss07 (the Loss variable for the market transition year of 2007), and LossPost07 (for the post 2007 regime of 2008-09).

The rationale for this breakdown is that the commercial real estate property market arguably passed through three distinct regimes during the past decade. The period through 2006 was characterized by first a stable and strong market then rising to a full-scale boom (or perhaps a bubble) of historic proportions in the latter few years of that period. The year 2007 was a transition year when the market suddenly and dramatically turned, but with such rapidity that market participants were faced with great uncertainty. Finally, by 2008 it had become clear that commercial real estate property prices were in a serious tailspin the likes of which had not been seen even in the previous "crash" of the early 1990s (which had been the worst fall since the Great Depression of the 1930s).

Consider first the transaction price model presented in column 1, which applies to the overall average during the entire 2001-09 sample, and which is therefore directly comparable to the previous results on the asking price. We note that in the transaction price the effect of loss aversion is smaller in magnitude (0.25) than we found it to be in the asking price in Table 3.2 (0.38).[14] Nevertheless, it is still both statistically and economically significant. This suggests that, while the loss aversion effect carries through to actual transactions, there is some degree of learning in the market through the deal negotiation process. Sellers are not able to achieve in actual sale prices as much loss aversion as they attempt to achieve (or signal) in their asking prices.

Interestingly, the anchoring effect, or degree of over-pricing, turns out to be not only statistically significant but larger than the effect of loss aversion. The coefficient of 0.77 implies that a 10 percent increase in the asking price over the market value results in the seller obtaining a higher transaction price by approximately 7.7 percent. This result implies that signaling and/or psychological anchoring is a potentially very powerful influence on the transaction price (within the range of the DOP observed in the data [15]). However, it is important to recall that the data do not allow us to know how much of the DOP effect we are quantifying here is actually anchoring versus signaling true (superior) quality attributes of a property which buyers subsequently discover and agree with the seller about. It seems possible that the signaling effect

[14]The difference would probably be not so big if we restricted the sample in the asking price regression to only the properties that eventually sold.

[15]Bucchianeri and Minson (2011) find a similar result in the housing market.

could quite large, particularly in cases where the asking price deviates widely from the expected market value.

Next consider the results in column 2 of Table 3.5 where we show the transaction price model with the three different regimes of loss aversion. The coefficient on the Loss variable is statistically significant in all three regimes, but of greater interest is the fact that it is *different* between and across each of the three regimes, and the nature of these differences is quite interesting.[16] First, during the stable and growing market regime of 2001-06 we find that the coefficient of 0.28 on LossPre07 is similar to the overall average result from column 1 discussed above (0.25). This might be viewed as reflecting the "normal" or typical effect of loss aversion in the commercial property market transaction prices. (Note that this coefficient is statistically significantly different from the coefficient of 0.155 on the Gain variable, again confirming the power of the prospect theory based behavioral loss aversion phenomenon even in transaction prices.) But of particular interest is what then happens to the loss aversion phenomenon in the following two exceptional market regimes.

First came the transition period of 2007 at the peak of the market cycle when the turnaround first hit and the market was dealing with great uncertainty. During this period loss aversion in the achieved transaction prices grew greatly in magnitude, to 0.38, significantly different from its prior and more "normal" level of 0.28. This reflected an extreme aversion of sellers (property owners) to facing the possibility of the dramatic change in fortunes that was occurring in 2007. They reacted by simply avoiding agreeing to any sales that did not reflect substantially greater than normal loss aversion. And they apparently succeeded in finding buyers who exhibited a larger-than-normal tendency to reach up toward the sellers' higher loss-averse asking prices (relative to the otherwise-expected market value). The result, of course, was a dramatic drop off in consummated sales volume in the latter part of 2007.

Finally this transition period was followed by an even more curious behavior. During the drastic downfall in the market of 2008-09 loss-aversion actually *weakened*

[16] Though not reported here, we also examined this issue in a slightly different manner, by dividing the dataset into two sub-samples: before 2007 and after 2007, with results substantially the same as those reported here.

to less-than-normal levels, falling statistically significantly below the "normal" level (the 0.16 coefficient on LossPost07 is less than the 0.28 coefficient on LossPre07 with statistical significance). Furthermore, the coefficient on LossPost07 is *not* significantly different from the coefficient on Gain, suggesting that in some sense there was perhaps very little loss aversion (of the prospect theory based behavioral form) during the most dramatic downturn in the market. We hypothesize that this suggests an ability for what one might term "extreme reality" to "break through" psychological behavior and enforce a more rational or "cold-eyed" business behavior. After all, loss aversion is based on a sort of psychological "wishful thinking" or illusion, an illusion that can indeed be realized to some extent in normal times (but only at the cost of lost liquidity and greater time on the market). It may be more difficult cognitively to keep up this type of thinking in the face of the magnitude of downturn that the market faced in 2008-09. It is also possible, of course, that the *demand side* of the market collapsed to such an extent in 2008-09 that loss-aversion behavior on the part of sellers could no longer be effective in consummated transaction prices. Indeed, this could be the actual market mechanism by which the sellers are forced to face reality; they simply can no longer find any buyers at all who will deal at prices that reflect loss aversion.

In summary, the results reported in column 2 of Table 3.5 suggest a wide temporal variation in loss aversion over the market cycle (at least when the cycle is as strong as it was during the 2000s decade). During "normal" times, loss aversion results in average transaction prices slightly higher than they would otherwise be (with concomitantly lower volume). During transition periods of a major market turning point (from up to down), we see that sellers facing a prospective loss during the year 2007 were able to obtain higher prices more so than they normally could (on a reduced volume of closed deals). We conjecture that the uncertainty in the market during that year made it difficult for the demand side to determine the true market value. Then, during 2008-09, the demand side revised downwards drastically its reservation prices, making it unrealistic for potential loss-averse sellers to continue holding out. The coefficient on LossPost07 is similar in magnitude to the coefficient on Gain, implying that loss-averse sellers could not do any different than other sellers in the market. This finding gives a unique perspective on how the market can correct behavioral anomalies. To our knowledge, this type of behavior has not been discovered previously in the literature.

3.5.4 Effects of Loss Aversion on Time on the Market

Consistent with the transaction price evidence of the preceding section, we would expect that if sellers facing a prospective loss have a higher reservation price, then they must also experience a longer time on the market, or equivalently, a lower hazard rate of sale. The hazard rate is defined as the probability that a property sells in any given week, given that the seller has listed the property for sale but it hasn't sold as yet. In this section, we estimate the effect of loss aversion on the hazard rate of sale. The hazard rate is specified as $h(t) = h_0(t)\exp(\alpha X)$, where X is a vector of covariates (in particular, "determinants" of sale propensity), with α being the vector of coefficients. The variable measuring the time spent on the market is the listing duration in weeks. For sold listings, duration is defined by the weeks elapsed between the date of entry into the market and the date of eventual sale. For properties that were delisted without sale (or "pulled" properties), their time on the market is measured by the weeks that elapsed between the dates of entry and exit from the market. In other words, they are treated as being censored at exit.

In Table 3.6, we estimate a Cox proportional hazard model (the classical methodology for calibrating this type of model). As expected, the coefficient on Loss is negative (-0.32) and statistically significant. This indicates for example that investors facing a 10 percent loss (when entering the market and therefore tending to post a higher asking price due to loss aversion behavior), experience approximately a 3% [$= \exp(0.32 * 0.1) - 1$] reduction in the weekly sale hazard (with a resulting concomitant increase in the expected time to sale). The comparative result in the G-M housing study was 3% to 6% (in other words, a slightly greater TOM effect in the 1990s Boston housing market). We also note that the positive and significant coefficient on Estimated Value indicates that higher quality or larger properties have a higher hazard of sale (or shorter time on the market).

3.6 Implications for Price Indices and Aggregate Market Pricing

This section presents two extensions to the previous literature on loss aversion. First, we consider the effect of incorporating the behavioral phenomena, described in sections 3.2 & 3.3 and quantified in section 3.5, on the construction of hedonic price indices. We will see that hedonic index construction can be greatly improved by incorporating the behavioral variables[17] Second, we consider the magnitude of the impact of prospect theory based loss aversion on the aggregate market in terms of the effect on the market-wide average realized transaction price. This effect will also be pictured relative to an historical price index so as to enable a better visualization of the relevant context. Finally, we will demonstrate the magnitude and nature of the entirety of the reference point based pricing that we have quantified in our transaction price model, including symmetric as well as asymmetric effects, and including the three-regime model of the 2000s market cycle.

3.6.1 Hedonic Index Construction

Consider first the role of the behavioral variables in the construction of a hedonic price index. Figure 3-3 presents a direct comparison of two hedonic indices based on the same database, the 4782 repeat-sales observations with sufficient hedonic data to employ the behavioral variables, the same database used in the empirical analysis presented in section 3.5. The hedonic indices presented here are pooled models of the Court-Griliches form in which the hedonic variables are treated as constant across

[17] It should be noted that the price indices developed here are not directly comparable to Moody's/REAL Commercial Property Price Indices (CPPI) that are based on RCA data and published by Moody's Investors Service. The indices developed here are based on hedonic models (with an implied "representative property"), rather than the same-property repeat-sales model employed in the CPPI. The estimation dataset is also somewhat different, including here less than 5,000 sold properties greater than $5,000,000 containing the necessary hedonic and behavioral variables, versus the CPPI which is based on approximately 17,000 repeat-sales of generally over $2,500,000 in price and some different data filters than what is employed here. Finally, for practical reasons the CPPI is a "frozen" index that is not revised with new data, whereas the present analysis of course includes all of the historical effects in the current dataset. The indices presented here have been developed solely for academic research purposes.

time and the price index is thus constructed purely from the coefficients on the time-dummy variables. The smoother index indicated by the blue squares includes the behavioral variables in the hedonic price model, while the more choppy index indicated by the green triangles is based on an otherwise identical hedonic model only without the behavioral variables. The model underlying the blue index is that of column 1 of Table 3.5 and therefore does not reflect the differential effect of loss aversion over time. The comparison with the green index shows that including the behavioral variables greatly improves the index, as the blue index clearly is less noisy than the green.[18] The index without the behavioral variables also does not capture as much of the downturn in prices, as it falls only 26% from 3Q07 to 2Q09 compared to 33% for the blue index.

Next, consider figure 3-4. We note that we had to eliminate a large number of otherwise potentially usable sales observations from the RCA database in order to construct the behavioral index, because we needed information on the reference point (prior sale price) and the asking price, data which was not available for most RCA transaction observations. The result is that the behavioral index must be estimated with barely more than one-tenth the sample size that could otherwise be used for a hedonic index. Figure 3-4 compares the behavioral index (the same blue index indicated by squares as in figure 3-3) with a straight hedonic index based on a much larger dataset. The index in figure 3-4 indicated by the red squares is estimated without behavioral variables, like the green index of figure 3-3, only now based on the full dataset of 45,870 single-sale observations. The much greater sample size tames the noise that we saw in the non-behavioral index of figure 3-3, but the red-squares hedonic index in figure 3-4 is still arguably not as good as the behavioral index even though the latter is based on a much smaller dataset. Without the behavioral variables the hedonic index fails to adequately capture the 2007-09 market downturn, dropping only 19% instead of the 33% of the blue index.

[18]This comparison can be quantified more rigorously by comparing the standard errors on the index log levels across the two models. As seen in Table 3.7, the average standard error in the index without the behavioral controls is almost seven times that of the index with the behavioral controls.

3.6.2 Aggregate Market Impact of Behavioral Loss Aversion

As we noted at the outset, an interesting question is the extent to which behavioral loss aversion is responsible for a salient feature of commercial property markets, namely, their tendency to lose liquidity during down markets. Clearly, the drastic drop in trading volume that occurs at the beginning of a price downturn results from a pulling apart of the demand and supply sides of the market. In principle, loss aversion could explain at least part of the tendency of property owners (the supply side) to hold up their reservation prices rather than following the demand down to pricing at which the market would maintain its normal liquidity.

Figure 3-5 presents a graphical image of the light our study can shed on this question, using the Moody's/REAL CPPI as a sort of "benchmark" to frame the history of the U.S. commercial property market.[19] Four price indices are presented in the chart. The solid black line is the CPPI, indicating the history of the average realized prices in the market, prices which therefore reflect whatever effect loss aversion behavior has at the aggregate level in the actual marketplace.[20] The red-squares and blue-diamonds indices track the movements of, respectively, the demand and supply sides' reservation prices, as measured by the TBI, presented here *relative* to the realized market transaction prices tracked by the CPPI.[21] Note that the potential buyers' reservation prices on the demand side began dropping after 2Q2007 but a gap between demand and supply did not open up until 1Q2008, increasing dramatically further by 1Q2009 to over 25% of the then-prevailing average transaction price. This was

[19]The Moody's/REAL CPPI is a useful benchmark because it is widely followed by industry participants and is viewed as presenting a good picture of actual transaction price movements in the market. Like the rest of this study, the CPPI is based on repeat-sales data from RCA. However, as noted previously there are some database differences that should be kept in mind.

[20]The CPPI depicted in figure 3-5 has been reset to have a starting value of 1.0 as of 1Q2001, as that is the first date in our analysis.

[21]MIT Center for Real Estate: http://web.mit.edu/cre/research/credl/tbi.html. The indices depicted in figure 3-5 are based on the "Liquidity Metric" published by the MIT/CRE. This metric is constructed as the percentage difference, each period, between the demand (supply) index and the TBI price index, normalized to starting values set so that the average value level of all the indices is equal over the entire 1984-2009 history of the TBI. The result is a reasonably good measure of the percentage gap between demand and supply prices as a fraction of current transaction prices, such that when the gap is zero there is approximately normal (long-term average) liquidity (trading volume) in the market. In figure 3-5 we simply take that same gap (each period) and apply it to the CPPI in the same way that the MIT/CRE applies it to the TBI.

accompanied by a drop in RCA major-asset trading volume from $136B in 2Q07 to $48B in 1Q08 to a nadir of $10B in 1Q09. In the early part of this drop the increase in the demand-supply gap was due in part to actual *increases* in the property owners' supply side reservation prices. The question is to what extent this behavior in the marketplace can be attributed to the type of prospect theory based psychological loss aversion behavior we have focused on in the present paper.

One way to answer this question is presented by the green line (indicated by "x" hash marks) in figure 3-5. This index presents a picture of how different the prevailing average market transaction prices would have been had there been no behavioral loss aversion of the type modeled in Table 3.5 of section 3.5. This "adjusted" market index is constructed using a three-regimes model of loss aversion similar to what was discussed in section 3.5.

The index is constructed as follows. First, a behavioral transaction price model similar to that in Table 3.5 is estimated, including three regimes for both the "Gain" and "Loss" variables. The regimes are defined as before: Pre-2007, 2007, and Post-2007 (ending with 4Q2009); and as before we see that the Loss coefficient is significantly larger than the Gain coefficient in the first two regimes but not in the third. Next, we take the difference, within each regime, of the Loss coefficient minus the Gain coefficient. As noted in section 3.5, this difference reflects the purely "behavioral" loss aversion phenomenon associated with the kink-point in the prospect theory value function.[22] We then multiply this difference by the average magnitude of the Loss variable within each regime to arrive at the magnitude of impact on the average transaction price within each regime conditional on the sold property facing a loss.

[22]In other words, there may be some influence of prior purchase price on sellers' asking prices that is "rational" (consistent with classical economic theory), included in the magnitude of the coefficient on the properties that have no loss (the "Gain" coefficient). For example, sellers booking a gain might rationally decide to ask a lower price than they otherwise would in order to apply some portion of their potential profits to selling more quickly, or selling a larger number of properties, rather than taking it all in price appreciation. Similarly, sellers facing a loss may rationally (and symmetrically) decide to do the opposite, asking a higher price in order to trade off some otherwise-expected book loss against taking a longer time to sell or selling fewer properties. This rational component of pricing behavior is not the loss aversion behavior rooted in prospect theory and that we are primarily focusing on in the current paper, though it also is an interesting phenomenon which we will consider further shortly.

Finally, we multiply that conditional loss impact times the proportion of transactions actually facing a loss (within our 4782-observation dataset), within each regime. We then subtract these regime-specific market-weighted impact factors from the CPPI, to produce a "loss-aversion adjusted" price index that presumably reflects something like what the overall market-wide average prices would have been had there been no prospect theory based loss aversion behavior.[23]

If behavioral loss aversion played a substantial role in "sticky pricing", keeping sellers' reservation prices high and thereby successfully influencing transaction prices to not fall far enough to maintain normal liquidity, then we would see the loss-aversion adjusted price index in figure 3-5 drop rapidly and substantially below the actual CPPI, with the loss-aversion-adjusted index tracing a path similar to that of the demand-side reservation price index (the red-squares index in figure 3-5) during the 2007-09 period. But we don't see this. In fact, the impact of behavioral loss aversion is relatively minor. As thusly computed, loss aversion increased aggregate market-wide average prices by only 0.7% during the 2001-06 regime, by 1.2% in 2007, and by 0.4% during 2008-09.

Thus, prospect theory based "behavioral" loss aversion did not apparently have a large impact on the broad macro-behavior (average transaction prices and volumes) within the U.S. commercial property market during the great cycle of the 2000s decade. However, as noted in section 3.5, behavioral phenomena are evidently important at the disaggregate, individual property level, and they may represent much of the strategic pricing behavior in the marketplace. Figure 3-6 presents one way to picture this, in a framework similar to what we have just used to evoke the macro-level impact.

[23]Technically, we first take the log of the CPPI level, then subtract the described loss-aversion adjustment factors, then exponentiate to convert back to straight levels.

3.6.3 Magnitude & Nature of Reference Point Based Pricing Behavior

The chart in figure 3-6 is again keyed on the Moody's/REAL CPPI. The index indicated by purple circles above the CPPI is the total effect of loss aversion pricing on properties facing a loss. Unlike the previous loss-aversion-adjusted market index in figure 3-5, the loss-properties price effect index in figure 3-6 reflects *total* loss aversion. It shows how much higher are the prices achieved by sellers facing a loss due to the pricing strategy of such sellers, including both the "rational" and "behavioral" components of that strategy. Because they are facing a loss, such sellers set prices higher than they otherwise would. (This is what is implied by the coefficient on the "Loss" variable in the price model in Table 3.5 discussed in the previous section.) The difference above the market-average CPPI price level indicated in the loss-properties price effect index is computed as the entire coefficient on the "Loss" variable times the mean magnitude of loss among sold properties that were facing an expected loss when they entered the market. It thus includes the component of loss-properties pricing that is symmetric with that of properties facing a gain (as well as the asymmetric "extra" component reflected in figure 3-5). The indicated difference between the loss-properties price effect index and the CPPI reflects the three regimes described previously, and is approximately 9%, 13%, and 7.5% above the average price indicated by the CPPI in each of the three regimes respectively: 2001-06, 2007, and 2008-09. This is therefore how much higher were the prices achieved (on average) by sellers facing a loss, compared to what they would have obtained if they were facing neither a loss nor a gain, over the 2000s market cycle.

Correspondingly, the orange-diamonds index showing the gain-properties price effect indicates how much lower was the average price obtained by properties sold conditional on the fact that they were facing an expected gain when they entered the market. As noted, such sellers take a lower price than they otherwise would (i.e., if they were facing neither a gain nor a loss). Such sellers are beyond the "kink-point" in the prospect theory value function, and their pricing strategy could reflect "rational" motivations consistent with classical economic utility theory, for example, to sell properties faster or to sell a larger volume of such properties. However, as noted in section 3.5, there could also be "behavioral" components in this pricing strategy, such as the "disposition effect" found by Crane and Hartzell (2009) in their study of REITs property sales. The gain-properties price effect index is computed similarly to the loss-properties price effect index, as the CPPI price level *minus* the hedonic transaction price model coefficient on the "Gain" variable times the mean magnitude of the "Gain" variable (among sold properties facing a gain), within each temporal regime.[24] This price reduction effect is estimated to be approximately 8%, 10%, and 8% below the average transaction price in the market as a whole indicated by the CPPI in each of the three regimes respectively: 2001-06, 2007, and 2008-09. The gain-properties price effect is thus only slightly less in magnitude than the loss-properties price effect. This is because, even though the coefficient on "Gain" is generally smaller than the coefficient on "Loss" (as reported in Table 3.5), the average magnitude of "Gain" is much greater than the average magnitude of "Loss" (at least during the 2001-09 cycle). Thus, the pricing impact is nearly symmetrical between the two cases (gainers vs losers).

The gap in pricing achieved by sold properties facing a loss versus those facing a gain, relative to what they would otherwise fetch (as a percentage of that average market price), is thus seen to be approximately 18%, 23%, and 15.5% within the three temporal regimes of the 2000s commercial property cycle (summing the magnitude of the two gaps noted above). These are clearly very substantially different pricing strategies employed by sellers facing a loss versus those facing a gain. Note also that the gap widened during the peak and turnaround year of 2007, and closed during the severe downturn of 2008-09, reflecting the previously-noted greater-than-normal strength of loss-aversion during the 2007 turning point and the weaker-than-normal loss-aversion of the severe downturn. As noted, not all of the difference between

[24]For the gain-properties index the difference is subtracted (versus added for the loss-properties index), because the "Gain" coefficient is positive and the "Gain" variable is defined to have a negative sign (prior purchase price minus expected current sale price: same definition as that for "Loss" in the case of loss-properties, giving the "Loss" variable a positive sign).

gainers and losers pricing is due to prospect theory based psychological loss aversion behavior, and the effect on overall average market prices is partly attenuated by the offsetting nature of the two effects, although the greater proportion of gainers compared to losers in combination with the nearly symmetrical price gap suggests that the gain properties price effect exceeds that of the loss properties in the overall market average price.[25]

The magnitude of the gap between the loss-properties and gain-properties pricing portrayed in figure 3-6 suggests the importance of differential pricing strategy among sellers facing a loss versus those facing a gain. While some of this strategy may reflect purely rational profit-maximizing concerns, the findings of the present paper combined with those of Crane & Hartzell (2009) on the disposition effect in REITs suggests that at least an important part of the revealed pricing gap may be reflect psychological behavior.

3.7 Conclusion

Using data on U.S commercial property sales of greater than $5,000,000 during the January 2001 - December 2009 period, this paper has explored the effect of loss aversion and anchoring on both asking prices and realized transaction prices, and we have developed historical price indices that are controlled for loss aversion behavior. This study has replicated and extended and enhanced the seminal study of Genesove & Mayer that discovered loss aversion behavior in housing markets, and has also added importantly to the Crane & Hartzell findings about the disposition effect in REITs. We confirm not only that such behavior exists also in the commercial property market, but indeed that loss aversion is of similar magnitude and impact as in the housing market that G-M studied. We furthermore find, contrary to some prior

[25]Note that the loss properties and gain properties price effect indices in figure 3-6 do not reflect the market weights of the two types of sales (gainers and losers). The indices simply represent the pricing differential of each type of sale relative to the market average. A combined and weighted market index adjusted to remove the effect of all reference point based pricing (both symmetric and asymmetric) would lie above the CPPI by 6.1%, 7.4%, and 4.2%, respectively during 2001-06, 2007, and 2008-09. Such an index would have fallen about 2% farther peak-to-trough in the 2007-09 downturn than the CPPI (as a percent of peak value).

literature, that loss aversion behavior in asking prices is actually greater among more experienced investors and among larger more "professional" institutions such as funds and REITs than among smaller private investors. The loss aversion behavior carries through to higher transaction prices (on average), and longer time on the market.

During the particularly dramatic commercial market cycle of the 2000s decade we find the effect of loss aversion behavior varied interestingly, first increasing in the early stage of the market peak and turning point, then collapsing in the face of the overwhelming reality of lack of buyers on the demand side. We explore the role of behavioral variables in the construction of hedonic price indices, finding that they can greatly improve such indices. We use our three-regime model to analyze the nature and magnitude of prospect theory based behavioral loss aversion on market-wide average prices at the aggregate level. We find that this impact is small and appears not to be the major source of the pulling apart of buyer (demand) and seller (supply) reservation prices that caused the severe illiquidity of the 2008-09 market collapse. However, we also use the same three-regime behavioral pricing model to develop indices of the pricing strategy of sellers facing a loss juxtaposed with that of sellers facing a gain. This illustrates the magnitude and cyclical nature of the differential pricing strategy of "losers" and "gainers" during the historic property market cycle of 2001-09, and may be at least partly reflective of a "disposition effect". We see both the substantial relative magnitude of this pricing strategy difference, as well as the way that it changed during the cycle, increasing during the peak and turning point year of 2007, and then attenuating during the subsequent severe downturn. The magnitude of the pricing strategy difference between gainers vs losers is substantial, as much as 23% during the peak and turning point year of 2007 and falling to 15.5% in the crash of 2008-09.

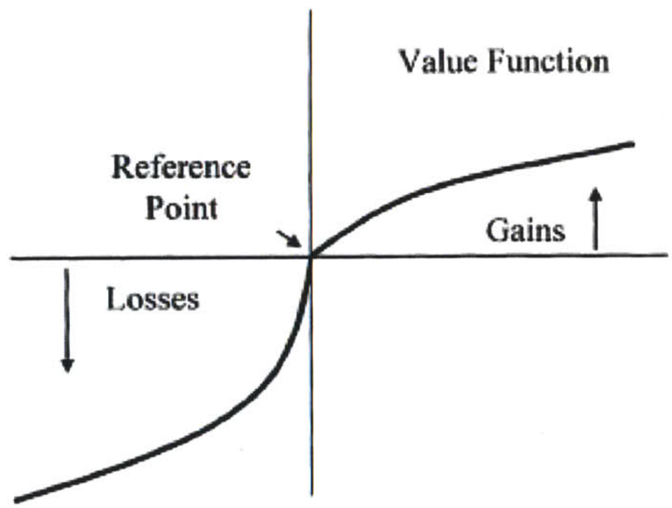


Figure 3-1: Value Function of Prospect Theory

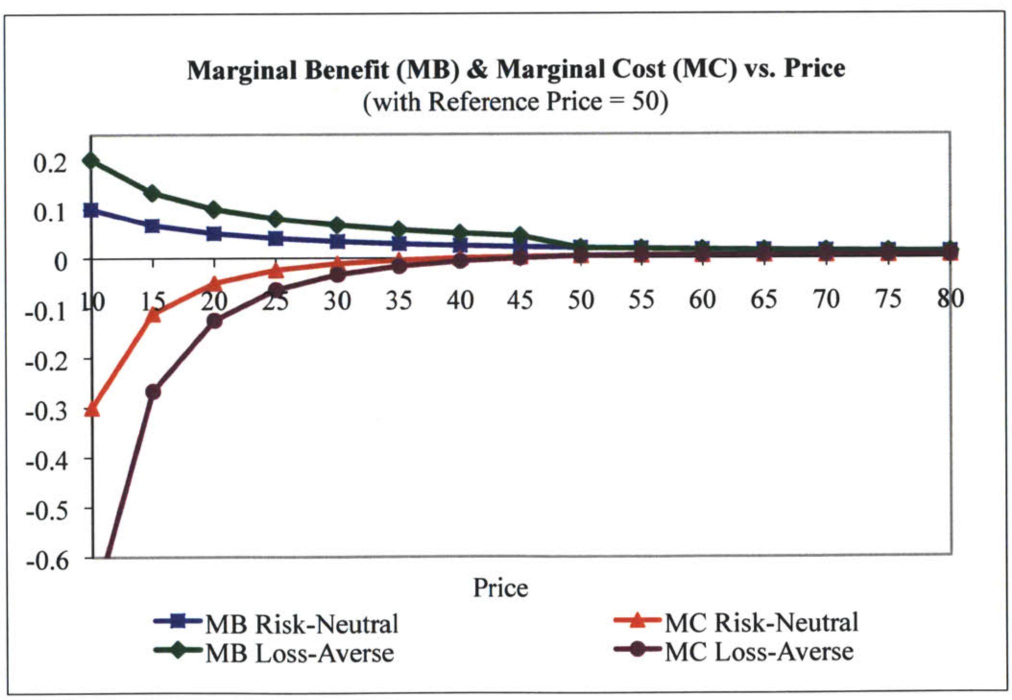


Figure 3-2: Marginal Benefit and Marginal Cost of an Increase in Price

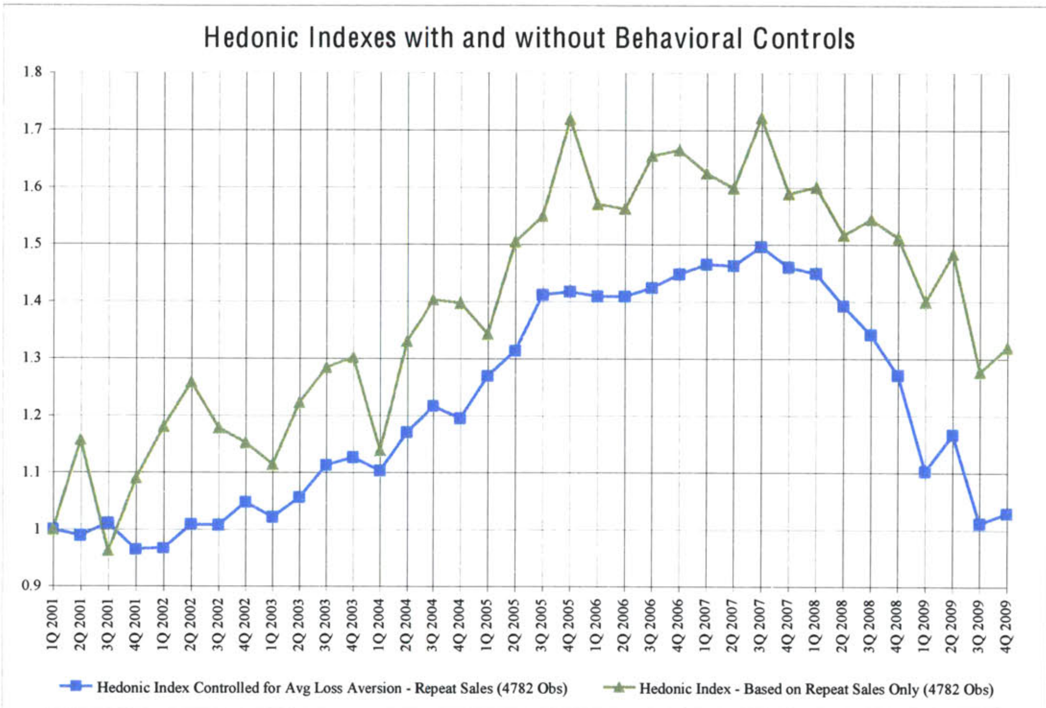


Figure 3-3: Hedonic Indexes with and without Behavioral Controls

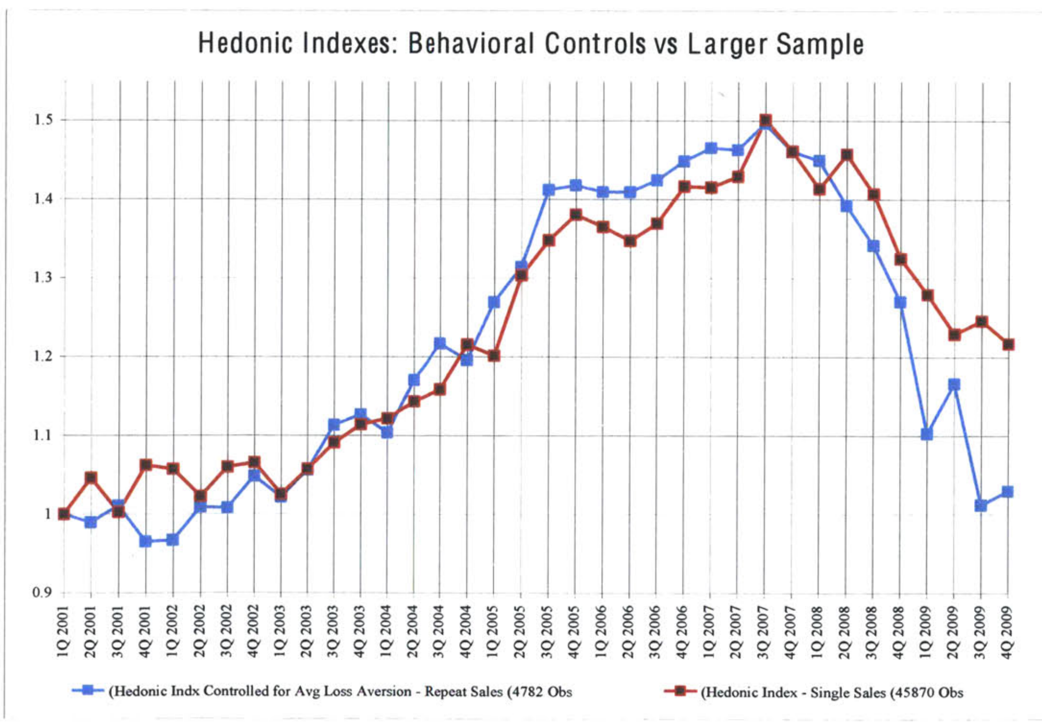


Figure 3-4: Hedonic Indexes: Behavioral Controls vs Larger Samples

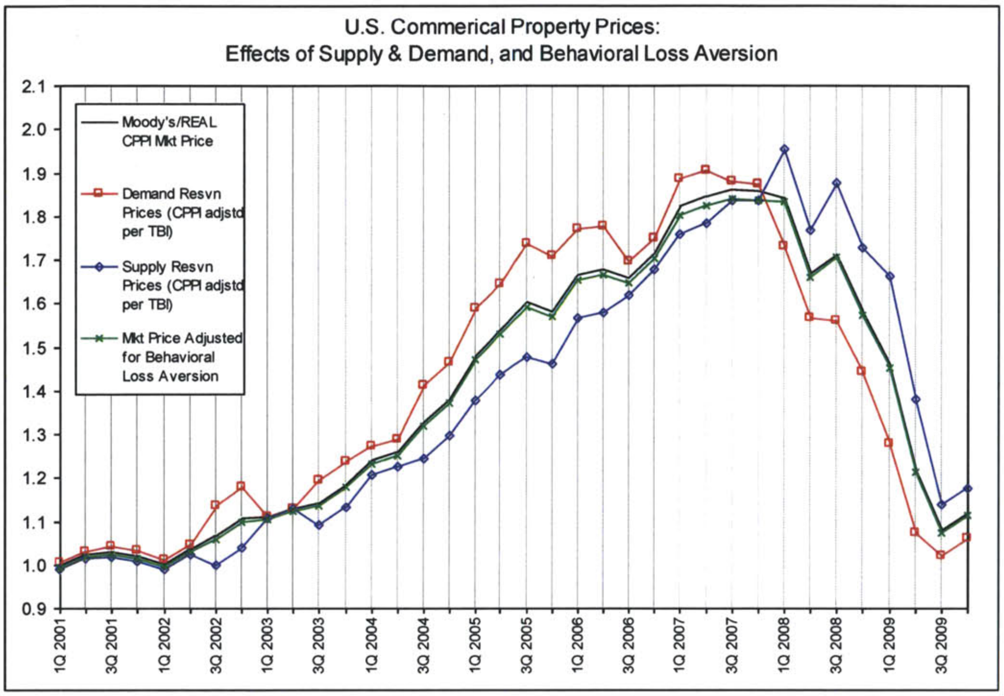


Figure 3-5: US Commercial Property Prices: Effects of Demand/Supply and Loss Aversion

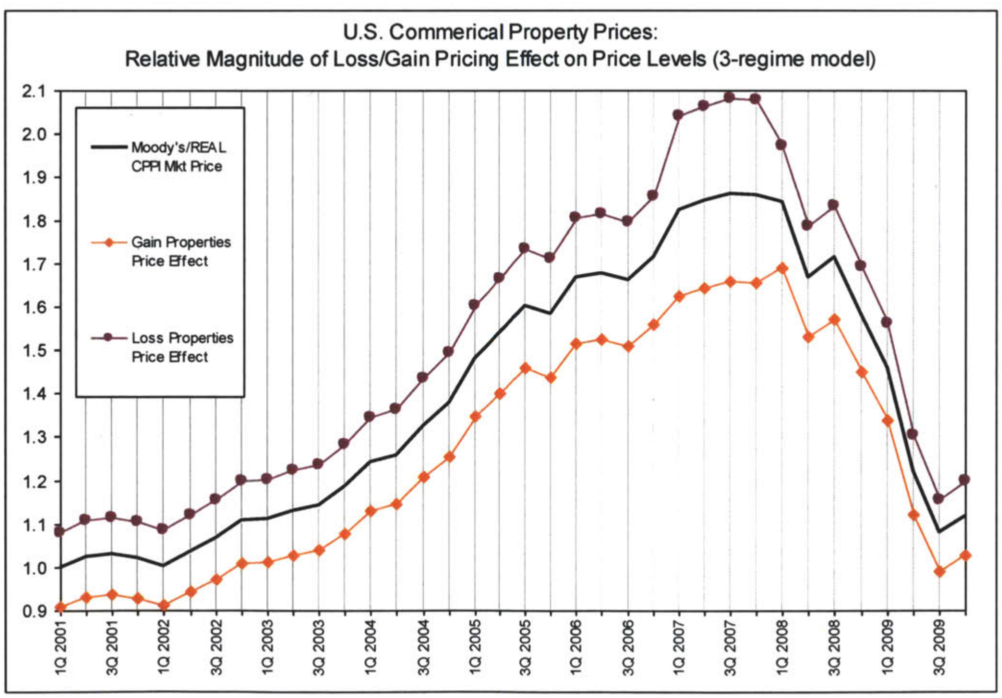


Figure 3-6: US Commercial Property Prices: Relative Magnitude of Loss/Gain Pricing Effect on Price Levels

	All Listings	Sold Listings
Number of Observations	6767	4782
Second Sale Price	-	\$31,775,301
		(77,510,785)
First Sale Price	\$20,537,628	\$22,084,851
	(50,984,144)	(5,359,336)
Asking Price	\$29,863,672	\$32,857,296
	(71,117,741)	(78,344,285)
Square Footage	176551	186250
	(218,790)	(217,986)
Percent with First Sale Price		
> Expected Selling Price	25%	22%
Months Since Last Sale	48	57
	(42)	(44)
Time on the Market (weeks)	50	37
	(38)	(31)
Trading Experience	88	101
	(174)	(185)

Table 3.1: Data Summary

	LogAskingPrice
Gain	0.223
	(0.031)
Loss	0.380
	(0.057)
Estimated Value	0.943
	(0.006)
Residuals from first sale	0.346
	(0.031)
Months since first sale	0.001
	(0.000)
Dummies for Quarter of Entry	yes
Constant	8.028
	(0.116)
R^2	0.87
N	6,767

Table 3.2: Loss Aversion and Asking Prices

	LogAskingPrice
Gain - Less experienced traders	0.219
	(0.031)
Gain - More experienced traders	0.233
	(0.036)
Loss - Less experienced traders	0.350
	(0.065)
Loss - More experienced traders	0.463
	(0.055)
Estimated Value	0.942
	(0.006)
Residuals from first sale	0.347
	(0.031)
Months since first sale	0.001
	(0.000)
Dummies for Quarter of Entry	yes
Constant	8.061
	(0.117)
R^2	0.87
N	6,767

Table 3.3: Loss Aversion and Trading Experience

	LogAskingPrice
Gain	0.193
	(0.029)
Loss - Institutional Investors	0.485
	(0.066)
Loss - Private Investors	0.260
	(0.071)
Loss - Equity Fund	0.515
	(0.082)
Loss - Public Investors	0.346
	(0.106)
Loss - User/Other Investors	0.335
	(0.213)
Estimated Value	0.935
	(0.006)
Residuals from first sale	0.381
	(0.029)
Months since first sale	0.000
	(0.000)
Dummies for Quarter of Entry	yes
Constant	8.072
	(0.096)
R^2	0.87
N	6,767

Table 3.4: Loss Aversion and Investor Types

	LogSecondSalePrice	LogSecondSalePrice
Gain	0.155	0.154
	(0.013)	(0.013)
Loss	0.245	
	(0.030)	
Anchoring/Degree of over-pricing	0.774	0.771
	(0.019)	(0.019)
Estimated Value	0.961	0.961
	(0.004)	(0.004)
Residuals from first sale	0.467	0.465
	(0.014)	(0.014)
LossPre07		0.280
		(0.049)
Loss07		0.383
		(0.045)
LossPost07		0.160
		(0.038)
Dummies for Quarter of Sale	yes	yes
Constant	7.788	7.791
	(0.054)	(0.054)
R^2	0.96	0.96
N	4,782	4,782

Table 3.5: Loss Aversion and Anchoring/Signaling

	TOM (in weeks)
Gain	0.024
	(0.092)
Loss	-0.321
	(0.127)
Estimated Value	0.275
	(0.017)
Residuals from first sale	0.254
	(0.095)
Months since first sale	0.007
	(0.001)
Dummies for Quarter of Entry	0.007
N	6,737

Table 3.6: Hazard Rate of Sale

	Hedonic Index with Average Loss Aversion (4782 Obs)			Hedonic Index based on Repeat Sales Data (4782 Obs)			Hedonic Index based on Single Transactions Data (45870 Obs)		
Quarters	Cum. Log Levels	Std. Errors	T-Statistic	Cum. Log Levels	Std. Errors	T-Statistic	Cum. Log Levels	Std. Errors	T-Statistic
1Q 2001	-0.111	0.05	-2.21	0.094	0.292	0.32	0.02	0.059	0.33
2Q 2001	-0.122	0.04	-3.02	0.24	0.287	0.84	0.065	0.059	1.1
3Q 2001	-0.1	0.041	-2.48	0.056	0.279	0.2	0.022	0.059	0.38
4Q 2001	-0.147	0.042	-3.48	0.18	0.275	0.66	0.08	0.058	1.37
1Q 2002	-0.145	0.048	-3.03	0.26	0.286	0.91	0.075	0.059	1.28
2Q 2002	-0.103	0.045	-2.3	0.324	0.281	1.15	0.042	0.058	0.72
3Q 2002	-0.103	0.04	-2.59	0.258	0.277	0.93	0.078	0.058	1.34
4Q 2002	-0.064	0.044	-1.47	0.236	0.278	0.85	0.083	0.058	1.44
1Q 2003	-0.09	0.041	-2.17	0.203	0.273	0.74	0.045	0.058	0.78
2Q 2003	-0.056	0.043	-1.32	0.295	0.278	1.06	0.075	0.058	1.29
3Q 2003	-0.004	0.044	-0.09	0.344	0.278	1.24	0.107	0.058	1.84
4Q 2003	0.008	0.037	0.21	0.358	0.271	1.32	0.127	0.057	2.23
1Q 2004	-0.013	0.041	-0.31	0.225	0.272	0.83	0.134	0.057	2.35
2Q 2004	0.046	0.038	1.21	0.379	0.272	1.4	0.153	0.057	2.69
3Q 2004	0.085	0.037	2.27	0.433	0.27	1.6	0.167	0.057	2.94
4Q 2004	0.067	0.037	1.83	0.429	0.27	1.59	0.215	0.057	3.79
1Q 2005	0.127	0.038	3.36	0.389	0.27	1.44	0.203	0.057	3.59
2Q 2005	0.162	0.039	4.13	0.503	0.27	1.86	0.285	0.057	5.03
3Q 2005	0.234	0.038	6.23	0.533	0.269	1.98	0.318	0.056	5.64
4Q 2005	0.238	0.037	6.51	0.636	0.27	2.36	0.342	0.056	6.07
1Q 2006	0.232	0.036	6.4	0.546	0.27	2.03	0.331	0.056	5.87
2Q 2006	0.232	0.036	6.47	0.541	0.27	2	0.318	0.056	5.63
3Q 2006	0.242	0.036	6.69	0.598	0.269	2.22	0.334	0.056	5.91
4Q 2006	0.259	0.036	7.23	0.604	0.269	2.24	0.368	0.056	6.52
1Q 2007	0.271	0.036	7.43	0.58	0.269	2.15	0.367	0.056	6.5
2Q 2007	0.269	0.035	7.7	0.564	0.27	2.09	0.377	0.056	6.68
3Q 2007	0.292	0.037	7.97	0.638	0.27	2.36	0.426	0.057	7.53
4Q 2007	0.268	0.036	7.42	0.558	0.27	2.07	0.399	0.057	7.03
1Q 2008	0.26	0.037	7.01	0.565	0.27	2.09	0.366	0.057	6.4
2Q 2008	0.22	0.036	6.09	0.511	0.27	1.89	0.396	0.057	6.91
3Q 2008	0.183	0.038	4.77	0.529	0.271	1.95	0.361	0.058	6.24
4Q 2008	0.128	0.039	3.29	0.507	0.271	1.87	0.301	0.058	5.16
1Q 2009	-0.014	0.047	-0.29	0.431	0.275	1.57	0.266	0.06	4.42
2Q 2009	0.043	0.041	1.04	0.489	0.274	1.78	0.226	0.06	3.78
3Q 2009	-0.1	0.047	-2.11	0.339	0.274	1.23	0.24	0.061	3.93
4Q 2009	-0.082	0.04	-2.05	0.372	0.272	1.37	0.216	0.059	3.65

Table 3.7: Cumulative Log-Level Hedonic Indexes - Estimates and Comparison

Chapter 4

Estimating Real Estate Price Movements for High Frequency Tradable Indexes in a Scarce Data Environment

4.1 Introduction and Background

In the world of transaction price indexes used to track market movements in real estate, it is a fundamental fact of statistics that there is an inherent trade-off between the frequency of a price-change index and the amount of "noise" or statistical "error" in the individual periodic price-change or "capital return" estimates.[1] Geltner & Ling (2006) discussed the trade-off that arises, as higher-frequency indexes are more useful, but *ceteris paribus* are more noisy and noise makes indexes less useful. More generally, the fundamental problem is transaction data scarcity for index estimation, and this is a particular problem with commercial property price indexes, because commercial transactions are much scarcer than housing transactions.[2] How-

[0]Co-authored with David Geltner

[1]The terms "noise" and "error" are used more or less interchangeably in this paper.

[2]There are over 100 million single-family homes in the U.S., but less than 2 million commercial properties.

ever, the greater utility of higher frequency indexes has recently come to the fore with the advent of tradable derivatives based on real estate price indexes.[3] Tradability increases the value of frequent, up-to-date information about market movements, because the lower transactions and management costs of synthetic investment via index derivatives compared to direct cash investment in physical property allows profit to be made at higher frequency based on the market movements tracked by the index. Higher-frequency indexes also allow more frequent "marking" of the value of derivatives contracts, which in turn allows smaller margin requirements, which increases the utility of the derivatives.[4]

In the present paper we propose a two-stage frequency-conversion estimation procedure in which, after a first-stage regression is run to construct a *low frequency* index, a second-stage operation is performed to convert a staggered series of such low-frequency indexes to a *higher-frequency index.* The first-stage regression can be performed using any desirable index-estimation technique and based on either hedonic or repeat sales data.[5] The proposed frequency conversion procedure is optimal in the sense that it minimizes noise at each stage or frequency. We find that while the resulting high-frequency index does not have as high a signal/noise ratio (SNR) as the underlying low-frequency indexes, it adds no noise in an absolute sense to what is in the low-frequency indexes, and it generally has less noise than direct high-frequency estimation. The 2-stage procedure thus preserves essentially all of the advantage of the low-frequency estimation while providing the additional advantage of a higher-frequency index.

---

[3]Over-the-counter trading of the IPD Index of commercial property in the UK took off in 2004 and after growing rapidly through 2007 the market remained active through the financial crisis of 2008-09. Trading on the appraisal-based NCREIF Property Index (NPI) of commercial property in the US commenced in the summer of 2007. The Moody's/REAL Commercial Property Price Index, launched in September 2007 based on Real Capital Analytics Inc (RCA) data, is also designed to be a tradable index and is, like the Case-Shiller house price index, a repeat-sales transaction price-based index.

[4]For example, margin requirements in a swap contract are dictated by the likely net magnitude of the next payment owed, which is essentially a function of the periodic volatility of the index, and volatility (per period) is a decreasing function of index frequency (simply because there is less time for market price change deviations around prior expectations to accumulate between index return reports that cover shorter time spans). Lower margin requirements allow greater use of synthetic leverage which facilitates greater liquidity in the derivatives market.

[5]It could even be based on appraisal data if the reappraisals occur staggered throughout the year.

The rest of the paper is organized into the following sections. In section 4.2, we briefly review the existing literature on estimating real estate price indexes and introduce the frequency-conversion technique, which we label as the "Generalized Inverse Estimator" (GIE) based on the mathematical procedure it employs. In sections 4.3 and 4.4 respectively, we discuss the hypothesized merits of the proposed procedure and provide an empirical comparison between the proposed technique and other popular methods of high frequency price index estimation. We conclude in a final section that the frequency conversion procedure tends to be more accurate at the higher frequency than direct high-frequency estimation in a data scarce environment.

4.2 Prior Work and the Proposed Frequency Conversion Procedure

Goetzmann (1992) introduced into the real estate literature what is perhaps the major approach to date for addressing small-sample problems in price indexes, namely, the use of biased ridge or Stein-like estimators in a Bayesian framework. Other approaches that have been explored in recent years include various types of parsimonious regression specifications that effectively parameterize the historical time dimension (see e.g. Schwann (1998), McMillen et al (2001), and Francke (2009)), as well as procedures that make use of temporal and spatial correlation in real estate markets (see for instance, Clapp (2004) and Pace et al (2004)). Some such techniques show promise, but are perhaps more appropriate in the housing market than in commercial property markets. Spatial correlation is more straightforward in housing markets, and the need for transparency in a tradable index can make it problematical to estimate the index on sales outside of the subject market segment. Another concern that is of particular importance in indexes supporting derivatives trading is that the index estimation procedure should minimize the constraints placed on the temporal structure and dynamics of the estimated returns series, allowing each consecutive periodic return estimate to be as independent as possible, in particular so as to avoid lag bias and to capture turning points in the market as they occur even if these are inconsistent with prior temporal patterns in the index.[6] Most of the previously noted recent techniques are unable to fully address these issues.

Bayesian procedures such as that introduced by Goetzmann (1992) can have the desirable feature of not inducing a lag bias and not hampering the contemporaneous representation in the index of turning points in the market. Such a technique, applied if necessary at the underlying low-frequency first stage, can therefore complement the frequency-conversion procedure we propose, and we explore such synergy in the present paper, finding that the frequency-conversion technique can further enhance indexes that are already optimized by such Bayesian methods.

We now introduce the frequency conversion procedure.[7] For illustrative purposes, we derive a quarterly-frequency index from four underlying staggered annual-frequency indexes. Other frequency conversions are equally possible in principle (e.g., from quarterly to monthly, or semi-annual to quarterly). Also for illustrative purposes and because they represent a scarce-data environment, we use a repeat sales database in this paper to assess the frequency-conversion technique. However, as mentioned earlier, the application of the proposed procedure is in principle not limited to any particular type of dataset, sample-size or choice of first-stage estimation methodology.

As noted, commercial property transaction price data in particular is scarce (e.g., compared to housing data). To the extent the market wants to trade specific segments, such as, say, San Francisco office buildings, the transaction sample becomes so small that we may need to accumulate a full year's worth of data before we have enough to produce a good transactions-based estimate of market price movement. This is the type of context in which we propose the following two-stage procedure to produce a

[6]This is particularly important to allow the derivatives to hedge the type of risk that traders on the short side of the derivatives market are typically trying to manage. For example, developers or investment managers seek to hedge against exposure to unexpected and unpredictable downturns in the commercial property market.

[7]It should be noted that the procedure introduced here is similar to what has been recently suggested in the regional economics literature, where Pavia et al (2008) provide a method for estimating quarterly accounts of regions from the national quarterly and annual regional accounts. The two methods are similar in that both use the generalized inverse, in the regional economics case to construct a quarterly regional series with movements that closely track the underlying annual figures.

quarterly index.

4.2.1 The Proposed Methodology

We begin by estimating annual indexes in four versions with quarterly staggered starting dates, beginning in January, April, July, and October. Label these four annual indexes: "CY", "FYM", "FYJ", and "FYS", to refer to "calendar years" and "fiscal years" identified by their ending months. Each index is a true annual index, not a rolling or moving average within itself, but consisting of independent consecutive annual returns.[8] It is important to use time-weighted dummy variables in the low-frequency stage in order to eliminate temporal aggregation. For example, for the calendar year (CY) index beginning January 1st, a repeat-sale observation of a property that is bought September 30 2004 and sold September 30 2007 has time-dummy values of zero prior to CY2004 and subsequent to CY2007, and dummy-variable values of 0.25 for CY2004, 1.0 for CY2005 and CY2006, and 0.75 for CY2007.[9] The result will look something like what is pictured in figure 4-1 for an example index based on the Real Capital Analytics repeat-sales database for San Francisco Bay area office property. If properly specified, these annual indexes generally have no lag bias and essentially represent end-of-year to end-of-year price changes.[10] Each of these indexes also has as little noise as is possible given the amount of data that can be accumulated over the annual spans of time. It is of course important for the low-frequency indexes to minimize noise and, while not the focus of the present paper, the annual-frequency indexes depicted here employ the previously noted Goetzmann Bayesian ridge regression method, which as noted does not introduce a general lag bias.[11]

[8] That is, independent *within* each index. Obviously, there is temporal overlap across the indexes.

[9] This specification, attributable to Bryon & Colwell (1982), eliminates the averaging of the values within the years, and effectively pegs the returns to end-of-year points in time. See Geltner & Pollakowski (2007) for more details.

[10] However, it should be noted that in the early stages of a sharp downturn in the market, loss aversion behavior on the part of property owners can cause a data imbalance that can make it difficult for an annual-frequency index to fully register the downturn at first. This consideration will be discussed shortly.

[11] Note that according to the Goetzmann Bayesian approach, the ridge is not necessary when the resulting indexes are sufficiently smooth without the ridge. This turns out to be the case for the

Next, a frequency-conversion is applied to this suite of annual-frequency indexes to obtain a quarterly-frequency price index implied by the four staggered annual indexes. We want to perform this frequency conversion in the most accurate way possible, with as little additional noise and bias as possible. How can we use those staggered annual indexes to derive an up-to-date quarterly-frequency index? Looking at the staggered annual-frequency index levels pictured in figure 4-1, one is tempted to try to construct a quarterly-frequency index by simply averaging across the levels of the four indexes at each point in time. (Try to fit a curve "between" the four index levels.) But such a process would entail a delay of three quarters in computing the most recent quarterly return (while we accumulate all four annual indexes spanning that quarter), which for derivatives trading purposes would defeat the purpose of the higher-frequency index. Such a levels-averaging procedure would also considerably smooth the true quarterly returns (it would effectively be a time-centered rolling average of the annual returns).

The approach we propose to the frequency-conversion procedure is a matrix operation which can be conceived of as a second-stage "repeat-sales" regression at the quarterly frequency using the four staggered annual indexes as the input repeated-measures data.[12] Each annual return on each of the four staggered indexes is treated like a "repeat-sale" observation in this "second-stage regression". If we have T years of history, we will have 4T-3 such repeated-measures observations (the row dimension of the second-stage regression data matrix), and we will have 4T quarters for which we have time-dummies (the columns dimension in the regression data matrix, the quarters of history for which we want to estimate returns). We are missing "1st-sales" observations for the first three quarters of the history, the quarters that precede the starting dates for all of the annual indexes other than the one that starts earliest in time (the CY index in our present example), as the staggered annual indexes each must start one quarter after the previous. Obviously, with fewer rows than columns

eight regional indexes examined later in the present paper, but not for the MSA-level indexes such as the San Francisco Office index depicted in figure 4-1.

[12]As noted, the input annual indexes from the first stage do not need to be repeat-sales indexes, in principle. They could be any good transactions-price based type of index, such as a hedonic index. They provide "repeated measures" in the sense that they provide repeated observations of index levels or price changes over time.

in the estimation data matrix, our regression is "under-identified", that is, the system has fewer equations than unknowns.[13] Basic linear algebra tells us that such a system has an infinite number of exact solutions (that is, quarterly index return estimates that will cause the predicted quarterly values to exactly match the repeated-measures observations on the left-hand-side of the second-stage regression, i.e., a regression R^2 of 100%, a perfect fit to the data which is the low-frequency index returns). However, of all of those infinite solutions, there is a particular solution that minimizes the variance of the estimated parameters, i.e., that minimizes the additional noise in the quarterly returns, noise added by the frequency-conversion procedure. This solution is obtained using what is called the "Moore-Penrose pseudoinverse" matrix of the data (see original papers by Penrose (1955, 1956) and its applications in Albert (1972)). We shall refer to this frequency-conversion method as the "Generalized Inverse Estimator", or GIE for short. This estimator is the "Best Linear Minimum Bias Estimator" (BLMBE) (see (Chipman, 1964) for proof).

How good is the GIE as a frequency-conversion method? It adds effectively no noise and very little bias to the underlying annual-frequency returns. Appendix B shows a way to see that the bias resulting from such an estimation decreases as the number of index periods to be estimated (T) becomes large, approaching zero as the history approaches an infinite number of periods. From the simulation analysis in figure 4-2, it can be seen that with even small values of T, the amount of bias is small and economically insignificant. In the simulation in figure 4-2 the history consists of less than seven years, converting to 27 quarters. Figure 4-2 depicts a typical randomly-generated history of true quarterly market values (the thick black line, which in the real world would be unobservable), the corresponding staggered annual index levels (thin, dashed lines, here without any noise, to reveal any noise added purely by the frequency-conversion second stage), and the resulting second-stage GIE-estimated quarterly index levels (thin red line with triangles, labeled "ATQ" for "Annual-to-

[13]We cannot simply drop out the first three quarters from the second-stage index, as that will then impute the first three annual returns entirely to the first quarter (only) of the index history and thereby bias the estimation of all of the quarterly returns.

Quarterly").[14] Clearly, the derived GIE quarterly index almost exactly matches the true quarterly market value levels. The slight deviation reflects the bias. Numerous simulations of random histories and varying market patterns over time give similar results to those depicted in figure 4-2. The GIE-based frequency-conversion adds only minimal and economically insignificant bias to the staggered annual indexes, while increasing the index frequency to quarterly. Unlike other techniques that require a procedure for choosing an optimal value for a parameter, the GIE is already optimal in the class of linear estimators, and it is relatively simple to compute (see Appendix A).

4.2.2 An Illustration of the Setup for the 2nd Stage Regression

To clarify and summarize the proposed procedure, consider this illustration. Suppose the following staggered annual returns were estimated from the 1st stage regression:

Annual Period	CY Returns	Annual Period	FYM Return	Annual Period	FYJ Return	Annual Period	FYS Return
1Q00-4Q00	a	2Q00-1Q01	b	3Q00-2Q01	c	4Q00-3Q01	d
1Q01-4Q01	e						

Then, as shown in the next table below, the left-hand side variable for the 2^{nd} stage regression will be the stack of annual returns (left-most column in the data table) and the right-hand side variables would be time-dummies that are set equal to 1 for the four quarters that make up a particular annual return observation (the other columns in the table below).

As seen above, there are more quarterly returns to estimate than there are staggered annual return observations. Specifically, for the ATQ frequency conversion,

[14] In figure 4-2 the first (CY) annual index starts arbitrarily at a value of 1.0, and the subsequent three staggered annual indexes are pegged to start at the interpolated level of the just-prior annual index at the time of the subsequent index's start date. This is merely a convention and does not impact the quarterly return estimates, as all indexes are only indicators of relative price movements across time, not absolute price levels.

Stacked-Returns	1Q 2000	2Q 2000	3Q 2000	4Q 2000	1Q 2001	2Q 2001	3Q 2001	4Q 2001
a	1	1	1	1	0	0	0	0
b	0	1	1	1	1	0	0	0
c	0	0	1	1	1	1	0	0
d	0	0	0	1	1	1	1	0
e	0	0	0	0	1	1	1	1

there are always 3 extra parameters to estimate. Appendix A outlines the method used for this estimation and Appendix B shows a way to see that the bias resulting from such an estimation decreases as T becomes large. Indeed, intuitively, the reader can convince oneself that as T becomes large, the percentage difference between T and T + 3 decreases. Hence, the system gets closer to being effectively exactly identified and thus the bias goes down over time. From the simulation analysis in figure 4-2, it can be seen that with even small values of T, the amount of the bias is small. It should also be noted that since the 2nd-stage regression fits the stacked returns observations exactly, any noise in the estimation of the staggered annual returns gets carried over to the estimated quarterly returns, but no new noise is added. Thus, there is a direct relationship between statistically reliable 1st and 2nd stage estimations. For this reason the frequency conversion procedure should be viewed as a complement or "add-on" to good lower-frequency index estimation, not a substitute for such current best practice.

4.2.3 General Characteristics of the Resulting Derived Quarterly Index

Based on the foregoing argument, the 2-stage derived quarterly index (which we shall refer to here as the "GIE/ATQ", for "annual-to-quarterly") offers the prospect of being more precise than a directly-estimated single-stage quarterly index, as it is based fundamentally on a year's worth of data instead of just a quarter's worth. However, though better than direct quarterly estimation in terms of precision or noise minimization, in one sense the 2-stage index cannot be as "good" as the corresponding underlying *annual-frequency* indexes, if we define the index quality by the signal/noise ratio. But the GIE/ATQ will provide information more frequently than the annual indexes, and this may make a lower signal/noise ratio worthwhile. To see this, consider the following.

In the signal/noise ratio (SNR) the numerator is defined theoretically as the periodic return volatility (longitudinal standard deviation) of the (true) market price changes, and the denominator is defined as the standard deviation of the error in the estimated periodic returns.[15] The GIE/ATQ frequency-conversion procedure gives a SNR denominator for the estimated quarterly index which is no larger than that of the underlying annual-frequency indexes (due to the exact matching of the underlying annual returns noted in the previous section). That is, the standard deviation of the error in the second-stage quarterly return estimates is no larger than that in the first-stage annual return estimates, as evident in the simulation depicted in figure 4-2 by the fact that the ATQ adds essentially no error. But the numerator of the SNR is governed by the fundamental dynamics of the (true) real estate market. These dynamics dictate that the periodic return volatility will be smaller for higher frequency returns. For example, if the market follows a random walk (serially uncorrelated returns), the quarterly volatility will be 1/SQRT $= 1/2$ the annual volatility. This means that, even though the theoretical SNR denominator does not increase at all (no additional estimation error), the SNR in the ATQ would still be one-half that in the underlying annual indexes. If the market has some sluggishness or inertia (positive autocorrelation in the quarterly returns, as is likely in real estate markets) then the SNR will be even more reduced in the ATQ below that in the annual indexes.

[15]The theoretical SNR cannot be observed or quantified in the real world, where the true market returns cannot be observed, and hence the true market volatility (SNR numerator) cannot be observed. Empirical estimates of the theoretical SNR are confounded by the fact that the volatility of any empirically estimated index will itself be "contaminated" by the noise in the estimated index (the denominator in the SNR). Furthermore, the denominator of the theoretical SNR should equal the theoretical cross-sectional standard deviation in the return estimates, which is not exactly what is measured by the regression's standard errors of its coefficients. To see this, consider conceptually a "perfect" index whose return estimates always exactly equal the unobservable true market returns each period. The regression producing such an index would have zero in the denominator of its theoretical SNR and yet would still have positive standard errors for its coefficients for any empirical estimation sample, as there is noise in the estimation database (cross-sectional dispersion in the price changes), causing the regression to have non-zero residuals in the data. In spite of these practical limitations, the theoretical SNR is a useful construct for conceptual analysis purposes (and also in simulation analysis, where "true" returns can be simulated and observed).

Importantly, the SNR of the GIE/ATQ can still be greater than that of a directly-estimated (single-stage) quarterly index. To see this, suppose price observations occur uniformly over time. Then there will be four times as much data for estimating the typical annual return in the annual-frequency indexes compared to the typical quarterly return in the directly-estimated quarterly-frequency index. By the basic "Square Root of N Rule" of statistics, this implies that the directly-estimated quarterly index will tend to have SQRT 2 times greater standard deviation of error in its (quarterly) return estimates than the annual indexes have in their (annual) return estimates. Thus, the SNR for the direct quarterly index will have a denominator twice that of the annual indexes and therefore twice that of the GIE/ATQ 2-stage quarterly index. Of course, either way of producing a quarterly index will still be subject to the same numerator in the theoretical SNR, which is purely a function of the true market volatility. Thus, the GIE/ATQ will have a lower SNR than the underlying annual-frequency indexes, but it will have a higher theoretical SNR than a directly-estimated quarterly index. In data-scarce situations, this can make an important difference.[16]

To return to the essential point of the contribution of this technique, while the GIE/ATQ does not have as good a SNR as the annual indexes, it does provide more frequent returns than the annual indexes (quarterly instead of annual), and thereby does provide additional information.[17] Thus, there is a useful trade-off between the staggered annual indexes and the derived quarterly index. The GIE/ATQ gives up some SNR information usefulness in the accuracy of its return estimates, but in return provides higher frequency return information.

[16]The fact that the GIE/ATQ theoretical SNR is greater than that of direct quarterly estimation does not mean that in every empirical instance it would necessarily be greater. It should be noted that formal definition and computation of a "standard error" for the GIE is not straightforward. As noted, the regression is under-identified, which means there are no "residuals" in the second stage.

[17]Among the four staggered annual indexes we do get new information every quarter, but that information is only for the entire previous 4-quarter span, which is not as useful as information about the most recent quarter itself, which is what is provided by the ATQ. For example, a turning point in the most recent quarter will not necessarily show up in the most recent annual index, as the latter is still influenced by market movement earlier in the 4-quarter span it covers.

4.2.4 An Illustrative Example of Annual-to-Quarterly Derivation in Data-Scarce Markets

To gain a more concrete feeling for the above-described methodology and application, consider one of the smaller (and therefore more data-scarce) markets among the 29 Moody's/REAL Commercial Property Price Indexes that are based on the RCA repeat-sales database: San Francisco Bay metro area office properties.[18]

First consider direct quarterly repeat-sales estimation of a San Francisco office market price index. Figure 4-3 depicts a standard Case-Shiller version of such an index based on the 3-stage WLS procedure first proposed by Case and Shiller (1987). This index is labeled "CS" in the chart and is indicated by the thin blue line with solid diamonds. The scarcity of the data gives the CS index so much noise that the resulting spikiness practically obscures the signal of the fall, rise, and fall again in that office property market subsequent to the dot-com bust, the following recovery, and the 2008-09 financial crisis and recession. The thicker green line marked by open diamonds labeled CSG adds the Goetzmann Bayesian ridge noise filter to the basic CS approach. The CSG index allows most of the market price trajectory signal to come through. But this index still may be excessively noisy for supporting derivatives trading, where index noise equates to "basis risk" that undercuts the value of hedging and adds spurious volatility that will turn off synthetic investors.

Now observe how the 2-stage procedure works in the San Francisco office market example. Figure 4-4 depicts the CSG-based direct-quarterly index which we just described together with the GIE/ATQ-based 2-stage quarterly index (indicated by the red line marked by open triangles). Figure 4-4 also shows the staggered annual-frequency indexes that underlie the ATQ (as thin fainter solid lines without markers). These indexes are themselves CSG-based repeat-sales indexes of the same methodology as the direct-quarterly index, only estimated at the annual rather than quarterly frequency. Thus, the annual indexes that underlie the ATQ index already employ

[18] The Moody's/REAL Commercial Property Price Index is produced by Moody's Investor Services under license from Real Estate Analytics LLC (REAL). During the 2006-09 period the San Francisco Office index averaged 12 repeat-sales transaction price observations (second-sales) per quarter, and in the most recent quarter (3Q09) there were only 2 observations.

the state-of-the-art noise filtering of the Goetzmann procedure. In figure 4-4, note how the ATQ index is generally consistent with the annual returns that span each quarter.[19] However, the quarterly index picks up and quantifies the changes implied by *changes* in the staggered annual indexes. For example, while the CY annual index ending at the end of 4Q2007 was positive (up 14.7%), it was less positive than the immediately preceding FYS annual index ending in 3Q2007 (up 26.7%). The resulting derived ATQ quarterly index indicated a downturn in 4Q2007 (-2.1%). Meanwhile, the direct-quarterly CSG index picked up a sharp downturn in 3Q2007, a quarter sooner than the ATQ, but the CSG then indicated a positive rebound of +1.6% in 4Q2007, which is probably noise. The smoother pattern in the ATQ index suggests less noise and therefore less spurious quarterly return estimates.

At first it may seem odd that the derived quarterly index can be negative when all of the staggered annual indexes that underlie it are positive. The intuition behind a result such as the above example is that an annual index could still be increasing as a result of rises during the earlier quarters of its 4-quarter time-span, with a drop in the last quarter that does not wipe out all of the previous three quarters' gains. When the most recent annual index is rising at a lower rate than the next-most-recent annual index, it can (although does not necessarily) indicate that the most recent quarter was negative. The derived quarterly return (ATQ) methodology is designed to discover and quantify such situations in an optimal (i.e., "BLMBE") manner. As noted, simple curve-fitting of the annual indexes introduces excessive smoothing, and will not be able to pick up in a timely manner the kind of turning point just described.

The San Francisco office market depicted in figure 4-4 offers an excellent example of both the strengths and weaknesses of the GIE/ATQ method versus direct quarterly estimation using state-of-the-art methods such as the CSG index depicted in the

[19]In fact, as noted previously, the ATQ returns are *exactly* equivalent to the corresponding underlying annual-frequency returns over each 4-quarter span of time covered by each of the staggered annual indices periodic returns. The depicted ATQ index level in the figure does not exactly touch each annual index periodic end-point value only because of the arbitrary starting value for the annual indexes. Note that the ATQ and CY indexes exactly match at the end of each calendar year, as both these indexes have the same starting value of 1.0 at the same time (2000Q4). The same would be true of the other three annual indexes if we set their starting values equal to the level of the ATQ at their starting points in time during 2001.

chart. Even though it uses Bayesian noise filtering, the CSG index is relatively noisy, as indicated by its spikiness during much of the history depicted (even when the transaction data was most plentiful). The CSG index differs importantly from the staggered annual indexes estimated from the same repeat-sales data. Arguably, the direct quarterly index does not as well represent what was going on in the San Francisco Bay office market during a number of individual quarters of the 2001-2008 period. For example, down movements of -9.1% in 2Q2005, -1.7% in 1Q2006, and -4.3% in 3Q2006 seem out of step with the strong bull market of that period, while up movements of +6.3% in 2Q2001 in the midst of the Bay Area's tech bust, and +1.6% in 4Q2007 and +1.2% in 3Q2009 seem out of step with the big downturn of 2007-09. The staggered annual indexes and the ATQ seem to better represent the tech-bust-related fall in the Bay Area office market during 2001-03 and the strong bull market of the 2004-07 period, and indeed in this particular case the ATQ appears visually to be about as good as the annual indexes (by the smoothness of the index lines' appearance), in addition to being more frequent. The directly-estimated quarterly index has considerably greater quarterly volatility than the ATQ, a likely indication of greater noise in the former index. On the other hand, in spite of the anomalous uptick in 4Q2007, the directly-estimated quarterly index shows some sign of slightly temporally leading the ATQ and the annual-frequency indexes. This is most notable in the direct quarterly index's beginning to turn down in 3Q2007, one quarter ahead of the ATQ in the 2007-09 market crash. Thus, there is some suggestion in our San Francisco office example that the GIE/ATQ method may not be quite as "quick" as direct quarterly estimation at picking up a sharp market downturn, although it appears to rapidly catch up.

4.3 Hypothesized Strengths & Weaknesses of the Frequency Conversion Approach

The preceding section presented a concrete example of both the strengths and weaknesses of the 2-stage/frequency-conversion procedure for providing higher-frequency market information in small markets. The suggestion is that the advantage for the 2-stage approach over direct (single-stage) high-frequency estimation would lie in the GIE's greater precision (less noise). However, even though the 2-stage procedure is more accurate in theory than direct high-frequency estimation, either procedure may be more accurate in a given specific empirical instance, particularly if the effective increase in sample size is small, which would be the case if the change in frequency is not great. In the empirical analysis in this paper the increase in going from quarterly to annual estimation is a fourfold increase in frequency (doubling of the "square root of N"), and we shall see what sort of results obtain.

While precision is a potential strength of the ATQ, there may be a weakness as well. The preceding examination of the San Francisco office index during the 2007-09 market downturn suggested that perhaps direct single-stage quarterly estimation is better at capturing the early stages of a sharp downturn in the market. Recall that the directly-estimated index turned down one quarter sooner than the ATQ in the San Francisco office market in 2007. In other words, the hypothesis would be that direct quarterly estimation might show a slight temporal lead ahead of annual estimation (and the resulting ATQ) in such market circumstances. This could result from the effect of loss aversion behavior on the part of property owners during the early stages of a sharp market downturn. Property owners react conservatively, not revising their reservation prices downward (perhaps even ratcheting them upwards, effectively pulling out of the asset market). Unless and until property owners are under pressure to sell in a down market, the result is a sharp drop-off in trading volume.

This has two impacts relevant for transaction price index estimation. First, the relatively few transactions that do clear during the early stages of the downturn reflect relatively positive or eager buyers. This dampens the price reduction actually realized in the market (as reflected in the prices observable in closed transactions). But it does not prevent a directly-estimated high-frequency index from reflecting that market price reduction (such as it is), as best such an index can do so (given the data scarcity, which increases the noise in the index), in the sense that the index does not have a lag bias.

The second effect of the fall-off in sales volume, however, poses a particular issue for annual-frequency indexes as compared to higher-frequency directly-estimated indexes. An annual index reflects an entire 4-quarter span of time in each periodic return, and in the downturn/loss-aversion circumstance just described the most recent part of that 4-quarter time span has markedly fewer transaction observations than the earlier part of the span. Thus, the data used to estimate the annual index's most recent annual return is dominated by the earlier, pre-downturn sales transactions. Even though the annual index uses Bryon-Colwell-type time-weighted dummy variables (as described previously), the sparser data in the more recent part of the time span may make it difficult for the annual index to fully reflect the recent market movement. Such a difficulty in the annual indexes would then carry over into the quarterly GIE/ATQ indexes derived from them.

4.4 An Empirical Comparison of Frequency Conversion versus Direct Estimation in Data-Scarce Markets

The RCA repeat-sales database and the Moody's/REAL Commercial Property Price Indexes based on that data present an opportunity to begin an empirical comparison of the two approaches. As noted, computation of index estimated returns standard errors is not straightforward for the GIE/ATQ, and "apples-to-apples" comparisons of estimated standard errors across the two procedures is not attempted in the present paper.[20] However, there are two statistical characteristics of an estimated real estate asset market price index that can provide practical, objective information about the quality of the index. These two characteristics are the volatility and the first-order autocorrelation of the index's estimated returns series. Based on statistical considerations, we know that noise or random error in the index return estimates will tend to increase the observed volatility in the index returns and to drive the index returns' first-order autocorrelation down toward negative 50%.[21]

Considering this, it would seem reasonable to compare the two index estimation methodologies based on the volatilities and first-order autocorrelations of the resulting estimated historical indexes. Lower volatility, and higher first-order autocorrelation, would be indicative of an index that is likely to have less noise or error in its individual periodic returns. For example, in the San Francisco office index that we considered previously in figure 4-4, the GIE/ATQ index has 4.4% quarterly volatility, versus 6.8% in the directly-estimated CSG index that seemed more noisy.

Among the Moody's/REAL Commercial Property Price Indexes there are 16 indexes (including the San Francisco office index we have previously examined) that are currently published at only the annual frequency (with four staggered versions, as described above), because the available transaction price data is deemed to be insufficient to support quarterly estimation. An examination of the relative values of the quarterly volatilities and first-order autocorrelations resulting from estimation of quarterly indexes by the two alternative procedures across these 16 market segments can provide an interesting empirical comparison of the two procedures in a realistic setting.

The 16 annual-frequency Moody's/REAL indexes include eight at the MSA level and eight at the multi-state regional level. The eight MSA-level indexes are: four

[20]For one thing, consider that the second-stage regression itself has no residuals, as it makes a perfect fit to the staggered lower-frequency indexes that are its dependent variable. Furthermore, as noted, the objective of a price index regression is not the minimization of transaction price residuals *per se*, but rather the minimization of error in the coefficient estimates (the index's periodic returns). While bootstrapping or simulation could be employed, the present paper focuses on the empirical analysis to follow.

[21]These are basic characteristics of the statistics of indexes. (See, e.g., Geltner & Miller et al (2007), Chapter 25.)

different property sectors (apartment, industrial, office, retail) for Southern California (Los Angeles and San Diego combined), three other MSA-level office indexes (New York, Washington DC, and San Francisco), and one other apartment index (for Southern Florida, which combines Miami, Ft Lauderdale, West Palm Beach, Tampa Bay, and Orlando). The eight multi-state regional indexes include the four property sectors each within each of two NCREIF-defined regions: the East and the South.[22]

Table 4.1 summarizes the comparison of the precision of the two approaches based on a volatility and autocorrelation comparison of the two quarterly index procedures (labeled "FC" and "DirQ" in the figure , for "Frequency-Conversion" and "Direct-Quarterly"). The volatility test is defined as the ratio of the FC quarterly volatility divided by the DirQ quarterly volatility. The autocorrelation test is defined by the arithmetic difference between the FC first-order autocorrelation minus that of the DirQ. Both tests are applied separately to the entire 33-quarter available history 2001-2009Q3 and to the more recent 19-quarter period 2005-09Q3. The RCA repeat-sales database "matured" to a considerable degree by 2005, with many more repeat-sales observations available since that time (until the recent crash and liquidity crisis). The comparison is made for each of the 16 indexes and also averaged across the eight MSA-level and eight regional-level indexes.

This comparison indicates that the 2-stage GIE/ATQ frequency conversion approach provided lower volatility and higher 1st-order autocorrelation in almost all cases, suggesting that this approach is more precise (less noisy). Of the 64 individual index comparisons (16 indexes X 2 time frames X 2 tests), the GIE/ATQ performed better than the DirQ in all but one case (the AC test for the Southern California Industrial Index during 2005-09).

However, while the frequency conversion procedure seems clearly to be less noisy at the quarterly frequency on the basis of the table 4.1 comparison, recall that we

[22]The East Region includes all the 15 states north and east of country-regionGeorgia, StateTennessee, and Ohio. The South Region includes the 9 states encompassed inclusively between and within Florida, Georgia, Tennessee, Arkansas, Oklahoma, and Texas. There is thus some geographical overlap between the MSA-level and regional-level indexes, in the sense that three of the eight MSA-level indexes are also within two of the regional-level indexes. The New York and Washington DC office indexes are within the East Office regional index, and the South-Florida Apartments index is within the South Apartment regional index.

raised a possible weak point about the 2-stage procedure in its ability to quickly and fully reflect the early stages of a sudden and sharp market downturn, such as occurred during 2007-09 in the U.S. commercial property markets. We suggested that during such times property-owner loss-aversion behavior could cause the underlying annual-frequency indexes to experience difficulty fully reflecting a late-period price drop in the market.

Table 4.2 presents some empirical evidence relevant to this point from the same 16 Moody's/REAL market-segment indexes examined in table 4.1. The table shows the percentage price change from the 2007 peak (within each index) through the most recently available 3Q2009 data as tracked in each market by the frequency-conversion index and the directly-estimated quarterly index. The table also presents two measures of the lead-lag relationship. In the left-hand columns are the calendar quarters of the peak for each index, and in the right-most column is the lead-minus-lag cross-correlation between the two indexes. We see that the direct quarterly index turned down first in six out of the 16 indexes (but only with a one quarter difference in five out of the six cases), while the frequency conversion index beat the direct quarterly in two cases (Washington DC & New York Metro office), with the two methods indicating the same peak quarter in eight cases. In the last column, if the lead-minus-lag cross-correlation is negative, it indicates that the correlation of the direct-quarterly index with the 2-stage index one quarter later is greater than the correlation of the converse, suggesting that the direct-quarterly index leads the frequency-conversion index. This is the case in 13 out of the 16 indexes, which suggests that the direct-quarterly index does show some tendency to slightly lead the frequency-conversion index in time.

4.5 Conclusion

This paper has described a methodology for estimating higher frequency (e.g., quarterly) price indexes from staggered lower-frequency (e.g., annual) indexes. The application examined here is to provide higher-frequency information about market movements in data-scarce environments that otherwise require low-frequency indexes. The proposed frequency-conversion approach takes advantage of the lower frequency to, in effect, accumulate more data over the longer-interval time periods which can be used to estimate returns with less error. Then it applies the Moore-Penrose pseudoinverse matrix in a second-stage operation in which the staggered low-frequency indexes are converted into a higher-frequency index. Linear algebra theory establishes that this frequency conversion procedure exactly matches the lower-frequency index returns and is optimal in the sense that it minimizes any variance or bias added in the second stage. Numerical simulation and empirical comparisons described here confirm that the two-stage frequency-conversion technique results in less noise than direct high-frequency estimation in realistic annual-to-quarterly indexes for practical U.S. commercial property price indexes such as the Moody's/REAL CPPI annual indexes (e.g., situations with second-sales observational frequency averaging in the mid-20s or less per quarter). The result is higher-frequency indexes that, while they have signal/noise ratios lower than the underlying low-frequency indexes, nevertheless add higher frequency information that may be useful in the marketplace, especially in the context of tradable derivatives. The only major drawback is that the frequency-conversion procedure may tend to slightly lag behind direct quarterly estimation, particularly during the early stage of a sharp market downturn. The lag appears to generally be no more than one quarter.

Finally, we would propose two strands of possibly productive directions for future research. First, throughout this paper no consideration was given to the covariance structure among the observations. Thus, more efficient estimators may exist if reasonable distribution assumptions were made and accounted for in the estimation of the high frequency series. Second, exploration of approaches that employ multiple imputation techniques in a Bayesian framework or a Markov Chain Monte Carlo context might lead to a better way of estimating high frequency indexes in a data scarce environment and in quantifying the noise that remains in the resulting indexes. With this in mind the current paper presents only a first step which may be improved upon by subsequent researchers, but which in itself appears to already have some practical value.

Figures 4-5 and 4-6 present charts of the GIE/ATQ and direct-quarterly indexes for all 16 annual-frequency Moody's/REAL Indexes.

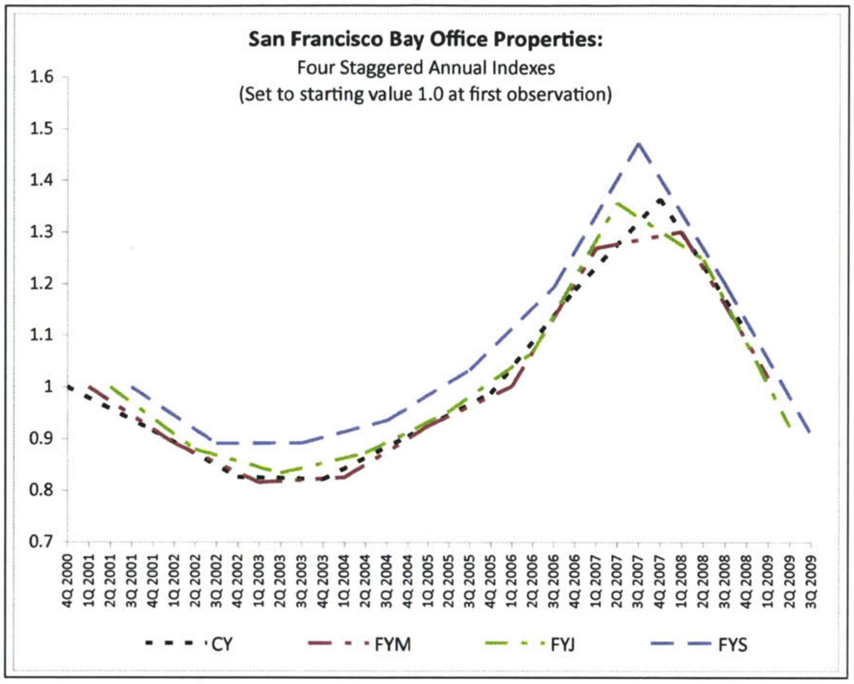


Figure 4-1: Staggered Annual Indexes

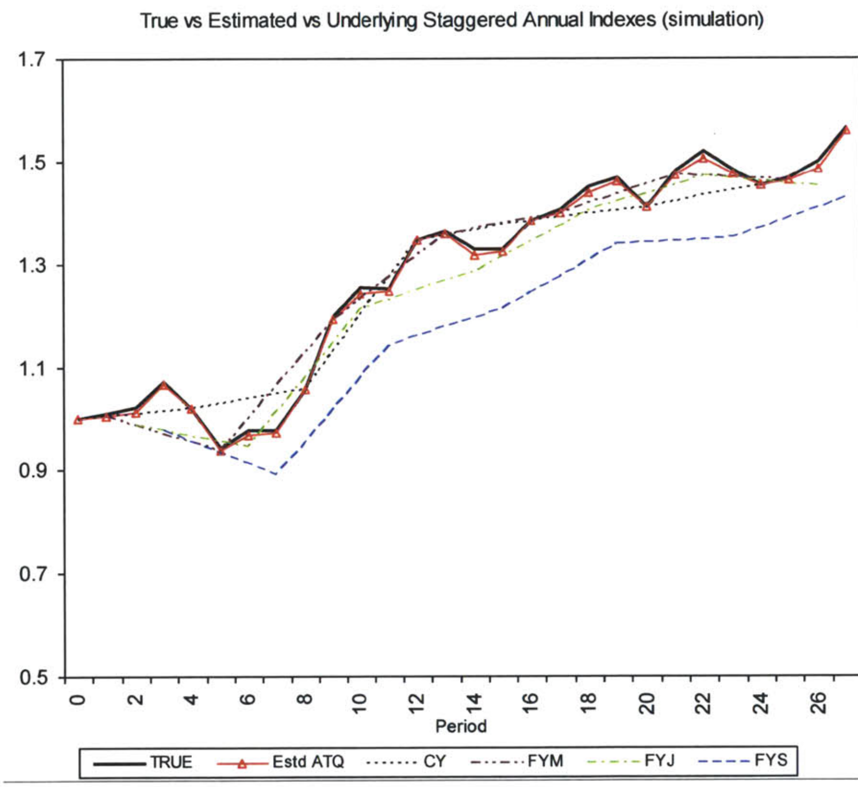


Figure 4-2: Simulation of True vs Estimated Quarterly Price Indexes: Annual to Quarterly

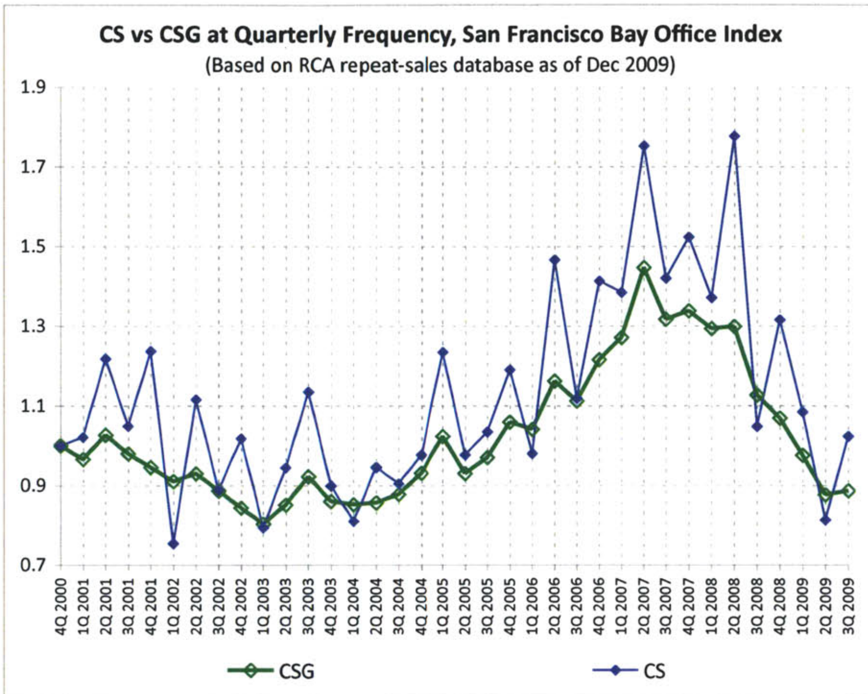


Figure 4-3: Direct Quarterly Estimation: Case-Shiller 3-stage WLS estimation (CS) versus Case-Shiller enhanced with Goetzmann Bayesian procedure (CSG)

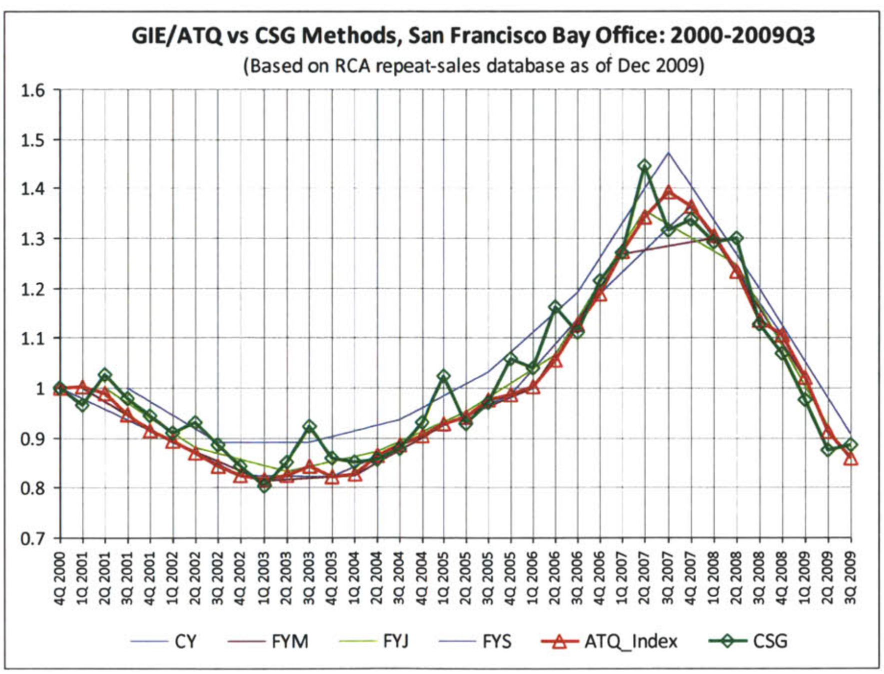


Figure 4-4: CSG direct quarterly estimation and derived GIE/ATQ estimation, together with the staggered annual indexes

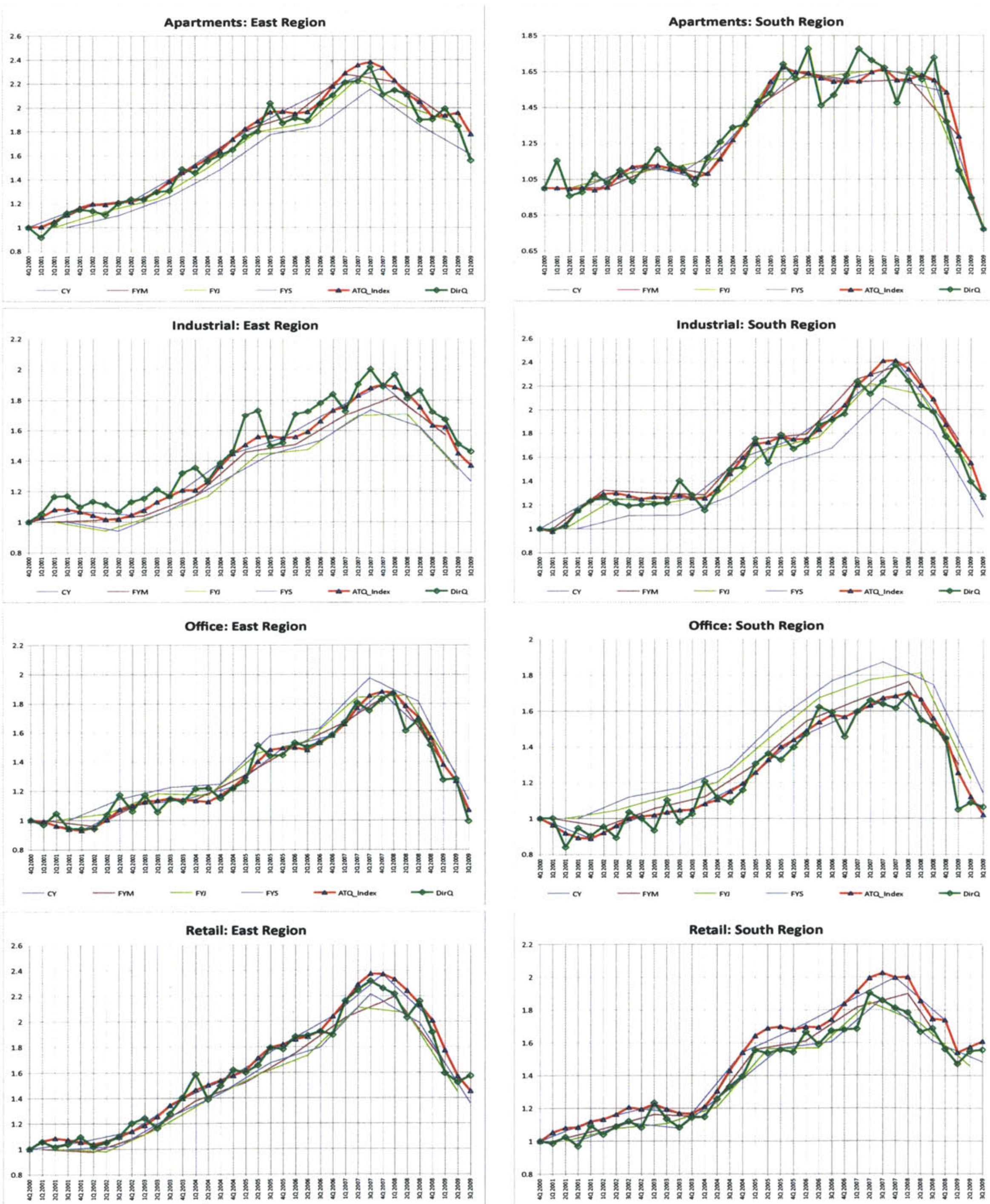


Figure 4-5: ATQ Index Suite - East and South Regions

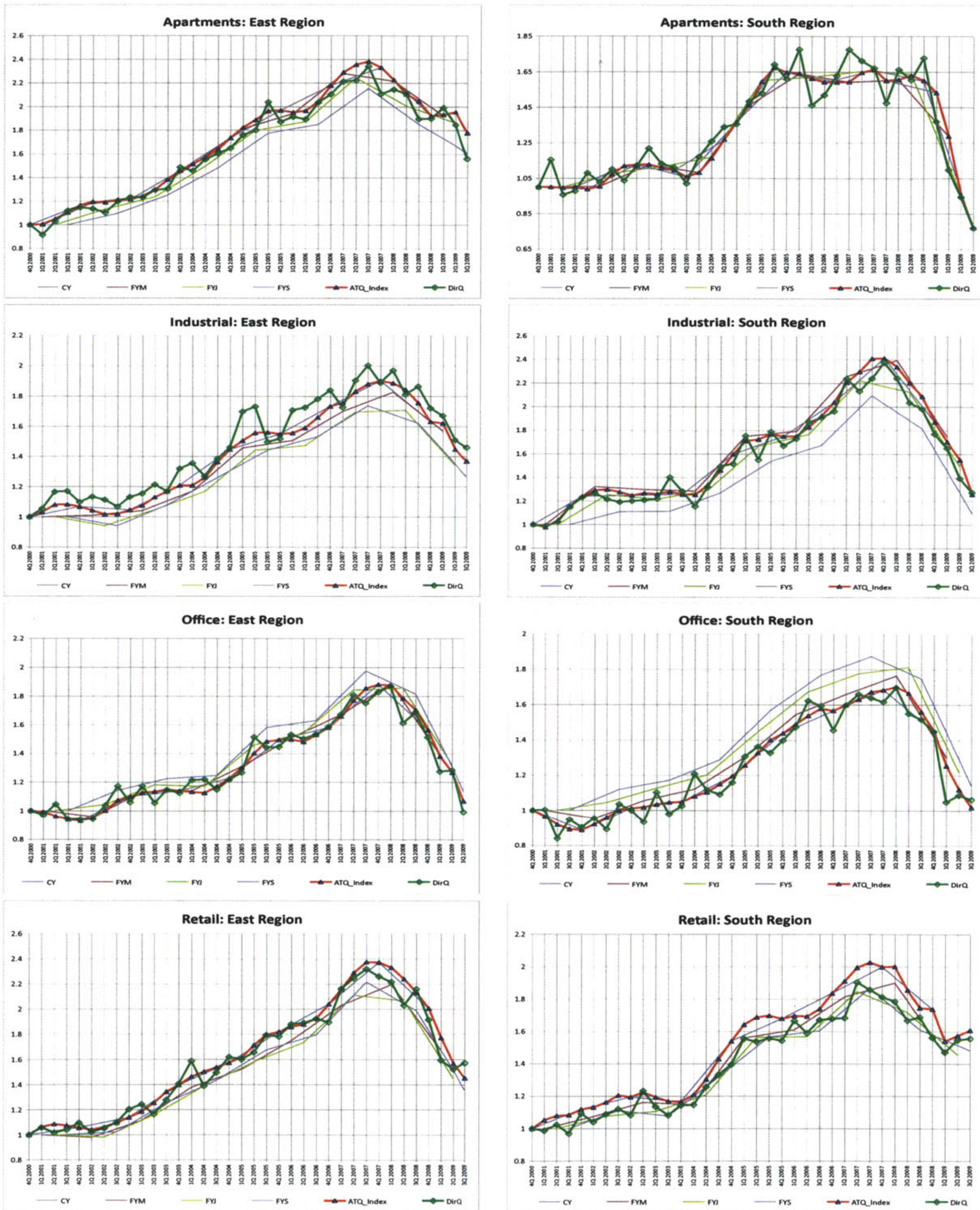


Figure 4-6: ATQ Index Suite - MSAs

	Data*:	Vol Test**:		AC(1) Test***:	
MSA-level indexes:					
Index:		2005-09	2001-09	2005-09	2001-09
NY Office	20 (4)	**0.55**	**0.47**	**80%**	**98%**
DC Office	14 (5)	**0.69**	**0.57**	**71%**	**90%**
SF Office	12 (2)	**0.67**	**0.64**	**91%**	**82%**
SC Office	27 (7)	**0.63**	**0.58**	**77%**	**92%**
SC Industrial	30 (15)	**0.59**	**0.4**	*-48%*	**77%**
SC Retail	16 (5)	**0.61**	**0.55**	**60%**	**85%**
SC Apts	43 (16)	**0.64**	**0.5**	**88%**	**109%**
FL Apts	18 (7)	**0.62**	**0.62**	**76%**	**77%**
Average:	23 (8)	**0.63**	**0.54**	**62%**	**89%**
Regional-level indexes:					
Index:		2005-09	2001-09	2005-09	2001-09
E Office	55 (19)	**0.69**	**0.6**	**87%**	**103%**
S Office	39 (13)	**0.65**	**0.49**	**86%**	**113%**
E Industrial	37 (14)	**0.54**	**0.56**	**67%**	**69%**
S Industrial	30 (8)	**0.75**	**0.73**	**91%**	**77%**
E Retail	29 (14)	**0.81**	**0.67**	**67%**	**83%**
S Retail	38 (9)	**0.78**	**0.69**	**50%**	**78%**
E Apts	56 (21)	**0.59**	**0.56**	**62%**	**74%**
S Apts	58 (23)	**0.75**	**0.68**	**71%**	**83%**
Average:	43 (15)	**0.7**	**0.62**	**73%**	**85%**

*Avg number of 2nd-sales obs/qtr 2006-09. Database was "immature" with considerably fewer 2nd-sales observations prior to 2005. (In parentheses number of obs in most recent 3Q09 qtr.)

** Ratio of FC volatility/DirQ volatility: <1 ==> **FC better**; >1 ==> *DirQ better.*

*** Difference: AC(1)FC - AC(1)DirQ. >0 ==> **FC better**; <0 ==> *DirQ better.*

Table 4.1: Volatility and Autocorrelation Tests: Frequency Conversion (GIE/ATQ) versus Direct Quarterly (CSG) Estimation

	2007 Peak*:		Downturn Magnitude: Peak-3Q09:		Lead – Lag Correl**:
Index:	FC	DirQ	FC	DirQ	
NY Office	**3Q07**	**4Q07**	-43.50%	-53.70%	**8%**
DC Office	**4Q07**	**1Q08**	-24.40%	-26.30%	**6%**
SF Office	*3Q07*	*2Q07*	-38.30%	-38.70%	*-7%*
SC Office	*4Q07*	*3Q07*	-39.50%	-40.90%	*-22%*
SC Industrial	4Q07	4Q07	-31.30%	-35.90%	*-32%*
SC Retail	*4Q07*	*3Q07*	-33.70%	-34.90%	*-32%*
SC Apts	3Q07	3Q07	-26.00%	-27.80%	*-17%*
FL Apts	3Q07	3Q07	-53.40%	-56.20%	*-55%*
Average:			-36.30%	-39.30%	*-19%*
Index:	FC	DirQ	FC	DirQ	Lead-Lag
E Office	1Q08	1Q08	-43.00%	-46.90%	**0%**
S Office	1Q08	1Q08	-40.10%	-37.30%	*-31%*
E Industrial	*4Q07*	*3Q07*	-28.00%	-27.20%	*-26%*
S Industrial	3Q07	3Q07	-47.90%	-43.20%	*-8%*
E Retail	3Q07	3Q07	-38.90%	-32.20%	*-21%*
S Retail	*3Q07*	*2Q07*	-20.60%	-18.20%	*-47%*
E Apts	3Q07	3Q07	-25.30%	-33.40%	*-41%*
S Apts	*3Q07*	*1Q07*	-53.60%	-56.60%	*-22%*
Average:			-37.20%	-36.90%	*-24%*

*Peak qtr before 2007 downturn. Sooner is presumably better, so:
FC earlier ➔ FC leads; *DirQ earlier ➔ DirQ leads.*
** Difference: Correl(FC(t),DirQ(t+1)) – Correl(DirQ(t),FC(t+1)):
Positive ==> FC leads; *Negative ==> DirQ leads.*

Table 4.2: Downturn Magnitude and Lead Minus Lag Cross-Correlation: Frequency Conversion (GIE/ATQ) versus Direct Quarterly (CSG) Estimation

Appendix A

Appendix to Chapter 2

A.1 Mathematical Appendix

Optimal Default Rule

Let P^* be the price below which the borrower defaults. The borrower objective function can be written as:

$$\Omega = u[Y - (1 - L)] + \delta \int_{P_L}^{P^*} v[(Y - (1 - L))(1 + e)]f(P)dP$$
$$+ \delta \int_{P^*}^{P_H} v[(Y - (1 - L))(1 + e)) + P - L(1 + r)]f(P)dP$$

Maximizing Ω w.r.t P^* and applying Leibniz's rule, we get:

$$\frac{d\Omega}{de} = \delta[v[(Y - (1 - L))(1 + e)]f(P^*) - v[(Y - (1 - L))(1 + e) + P^* - L(1 + r)]f(P^*)] \overset{set}{=} 0$$

The P^* that makes this expression equal to zero is $P^* = L(1 + r)$

Expected Utility w.r.t to e and r

Based on the optimal default rule derived above in (i), the borrower's objective function can be written as follows:

$$\Omega = u[Y - (1 - L)] + \delta \int_{P_L}^{L(1+r)} v[(Y - (1 - L))(1 + e)]f(P)dP$$
$$+ \delta \int_{L(1+r)}^{P_H} v[(Y - (1 - L))(1 + e) + P - L(1 + r)]f(P)dP$$

Maximizing w.r.t to e, we find that $\frac{d\Omega}{de} > 0$:

$$\frac{d\Omega}{de} = \delta[Y - 1 + L] \int_{P_L}^{L(1+r)} \underbrace{v'[(Y - (1 - L))(1 + e)]f(P)dP}_{>0 \text{ by assumption}}$$
$$+ \delta[Y - 1 + L] \int_{L(1+r)}^{P_H} \underbrace{v'[(Y - (1 - L))(1 + e) + P - L(1 + r)]f(P)dP}_{>0 \text{ by assumption}} \quad \text{(A.1)}$$

Similarly, maximizing w.r.t to r and applying Leibniz's rule we get:

$$\frac{d\Omega}{dr} = \delta \frac{d}{dr}[L(1 + r)]v[(Y - 1 + L)(1 + e)]f(L(1 + r))$$
$$- \delta \frac{d}{dr}[L(1 + r)]v[(Y - 1 + L)(1 + e) + L(1 + r) - L(1 + r)]f(L(1 + r))$$
$$- L\delta \int_{L(1+r)}^{P_H} v'[(Y - (1 - L))(1 + e) + P - L(1 + r)]f(P)dP$$

Simplifying, we find that $\frac{d\Omega}{dr} < 0$

$$\frac{d\Omega}{dr} = -L\delta \int_{L(1+r)}^{P_H} \underbrace{v'[(Y - (1 - L))(1 + e) + P - L(1 + r)]f(P)dP}_{>0 \text{ by assumption}} \quad \text{(A.2)}$$

From (A.1) and (A.2), it follows that $\frac{d\Omega}{de} - \frac{d\Omega}{dr} > 0$ and $-(\frac{d\Omega}{de} - \frac{d\Omega}{dr}) < 0$.

Optimal LTV when $e > r$

For the case where $e > r$, we maximize the borrower objective function in (ii) above by choosing an optimal L: The first order condition is as follows:

$$\begin{aligned}
\frac{d\Omega}{dL} &= u'[Y-1+L] + \delta\frac{d}{dL}[L(1+r)]v[(Y-1+L)(1+e)]f[L(1+r)] \\
&+ \delta(1+e)\int_{P_L}^{L(1+r)} v'[(Y-(1-L))(1+e)]f(P)dP \\
&- \delta\frac{d}{dL}[L(1+r)]v[(Y-1+L)(1+e)+L(1+r)-L(1+r)]f[L(1+r)] \\
&+ \delta(e-r)\int_{L(1+r)}^{P_H} v'[(Y-1)(1+e)+P+L(e-r)]f(P)dP
\end{aligned}$$

Simplifying, we see that $\frac{d\Omega}{dL} > 0$ no matter what the loan rate:

$$\begin{aligned}
\frac{d\Omega}{dL} &= \underbrace{u'[Y-1+L]}_{>0 \text{ by assumption}} \\
&+ \delta(1+e)\int_{P_L}^{L(1+r)} \underbrace{v'[(Y-(1-L))(1+e)]f(P)dP}_{>0 \text{ by assumption}} \\
&+ \delta\underbrace{(e-r)}_{>0}\int_{L(1+r)}^{P_H} \underbrace{v'[(Y-1)(1+e)+P+L(e-r)]f(P)dP}_{>0 \text{ by assumption}}
\end{aligned}$$

Thus, the borrower will choose to borrow as much as possible, which is the maximum allowable LTV of 1.

Proof of Proposition 1

FOC: First, we derive the first-order condition for the borrower's objective function when $e < r$. The objective function is:

$$\Omega = u[Y-(1-L)] + \delta\int_{P_L}^{L(1+r)} v[0]f(P)dP + \delta\int_{L(1+r)}^{P_H} v[P-L(1+r)]f(P)dP$$

for $e < r$

Maximizing w.r.t L and applying Leibniz's rule is:

$$\begin{aligned}
\frac{d\Omega}{dL} &= u'[Y-1+L]+\delta\frac{d}{dL}[L(1+r)]v[0]f[L(1+r)]+\delta\int_{P_L}^{L(1+r)} 0 \\
&- \delta\frac{d}{dL}[L(1+r)]v[0]f[L(1+r)]-\delta(1+r)\int_{L(1+r)}^{P_H} v'[P-L(1+r)]f(P)dP
\end{aligned}$$

Simplifying, we verify (2.2), the first-order condition in the text:

$$u'[Y-1+L]-\delta(1+r)\int_{L(1+r)}^{P_H} v'[P-L(1+r))f(P)d(P) \overset{set}{=} 0 \quad (2.2)$$

1.1: As r increases, the minimum LTV would be such that $u'[Y-1+L]=0$, i.e when $L=1-Y$ for $Y<1$. If $Y>1$, the LTV demand would fall to 0.

1.2: As r goes to e, the model reduces to that given in (iii) above and L will be equal to 1, the maximum allowable leverage.

1.3: To prove that L^D is a downward sloping curve w.r.t loan rate r, we need to show that $\frac{dL^D}{dr}<0$. By taking the total derivative of the first-order condition (2.2) w.r.t L and r, we can derive $\frac{dL^D}{dr}=-\frac{\Omega_{Lr}}{\Omega_{LL}}$. We proceed to derive Ω_{LL} and Ω_{Lr}:

$$\begin{aligned}
\frac{d^2\Omega}{dL^2} &= u''[Y-1+L]+\delta\frac{d}{dL}[L(1+r)]v'[L(1+r)-L(1+r)]f[L(1+r)] \\
&+ \delta(1+r)^2\int_{L(1+r)}^{P_H} v''[P-L(1+r)]f(P)dP
\end{aligned}$$

Simplifying and using the assumption that the derivative of the utility function at zero wealth is zero (i.e $v'(0)=0$), we find that $\frac{d^2\Omega}{dL^2}<0$:

$$\frac{d^2\Omega}{dL^2}=\underbrace{u''[Y-1+L]}_{<0 \text{ by assumption}}+\delta(1+r)^2\int_{L(1+r)}^{P_H}\underbrace{v''[P-L(1+r)]f(P)dP}_{<0 \text{ by assumption}} \quad (A.3)$$

Next, taking the derivative of the first order condition w.r.t r, we get:

$$\begin{aligned}
\frac{d^2\Omega}{dLdr} &= (-) - \delta\frac{d}{dr}[L(1+r)]v'[L(1+r) - L(1+r)]f[L(1+r)] \\
&+ \delta(1+r)^2\int_{L(1+r)}^{P_H} v''[P - L(1+r)]f(P)dP \\
&- \delta\int_{L(1+r)}^{P_H} v'[P - L(1+r))]f(P)d(P)
\end{aligned}$$

Using the assumption that $v'(0) = 0$, we find that $\frac{d^2\Omega}{dLdr} < 0$:

$$\frac{d^2\Omega}{dLdr} = \delta[(1+r)^2\int_{L(1+r)}^{P_H} \underbrace{v''[P - L(1+r)]f(P)dP}_{\text{<0 by assumption}} - \int_{L(1+r)}^{P_H} \underbrace{v'[P - L(1+r))]f(P)d(P)]}_{\text{>0 by assumption}} \tag{A.4}$$

It follows from (A.3) and (A.4) that $\frac{dL^D}{dr} = -\frac{\Omega_{Lr}}{\Omega_{LL}} < 0$

1.4: We need to show that $\frac{dL^D}{d\delta} < 0$, which can be derived as $\frac{dL^D}{d\delta} = -\frac{\Omega_{L\delta}}{\Omega_{LL}}$. The denominator is given by (A.3), so we just need to derive $\Omega_{L\delta}$:

$$\frac{d^2\Omega}{dLd\delta} = -(1+r)\int_{L(1+r)}^{P_H} \underbrace{v'[P - L(1+r))f(P)d(P)}_{\text{>0 by assumption}} \tag{A.5}$$

Since $\frac{d^2\Omega}{dLd\delta} < 0$, it follows that $\frac{dL^D}{d\delta} = -\frac{\Omega_{L\delta}}{\Omega_{LL}} < 0$

1.5 It is straightforward to see that $\frac{d^2\Omega}{dLdY} < 0$:

$$\frac{d^2\Omega}{dLdY} = \underbrace{u''[Y - 1 + L]}_{\text{<0 by assumption}} \tag{A.6}$$

Again, it follows from (A.3) and (A.5) that $\frac{dL^D}{dY} = -\frac{\Omega_{LY}}{\Omega_{LL}} < 0$

1.6 If P first-order stochastically dominates P', then by definition $E[v|P] \geq E[v|P']$. Since v is increasing, it follows that $E[v'|P] \geq E[v'|P']$. From (2.2), the left hand side can thus be written as $u'[Y - 1 + L^P] \geq u'[Y - 1 + L^{P'}]$. Since u is increasing in L, this implies that $L^P \geq L^{P'}$.

1.7 If P second-order stochastically dominates P', then by definition $E[v|P] \geq E[v|P']$. The rest of the proof is the same as above, showing that $L^P \geq L^{P'}$.

Rank Condition

For the structural demand equation, (2.3), to be identified, we need to check if the rank condition is satisfied. We begin with arranging the structural parameters in a matrix and also writing out the restrictions for (2.3):

$$A = \begin{pmatrix} 1 & -\gamma_1 \\ -\beta_0 & -\gamma_0 \\ -\beta_1 & 1 \\ -\beta_2 & -\gamma_2 \\ -\beta_3 & -\gamma_3 \\ 0 & -\gamma_4 \end{pmatrix} \qquad \Phi_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The one's in the first and second columns of A are the normalizations on the coefficients of L and r in (2.3) and (2.4), respectively. The zero in column 1 is the exclusion restriction on the coefficient of CLL in the demand equation, (2.3). The matrix Φ_1 collects the normalization and exclusion restrictions for (2.3). Multiplying Φ_1 and A gives:

$$\Phi_1 A = \begin{pmatrix} 1 & -\gamma_1 \\ 0 & -\gamma_4 \end{pmatrix}$$

The rank$(\Phi_1 A) = 2$ if $\gamma_4 \neq 0$. Thus, as long as the coefficient on CLL in (2.4) is non-zero, the demand equation, (2.3), is identified.

Reduced Form Parameters

Let the dimension of X and T be n by k and n by m, respectively. The structural equations, (2.3) and (2.4) can be expanded as follows:

$$L = \beta_0 + \beta_1 r + \beta_{21} x_1 + \ldots + \beta_{2k} x_k + \beta_{31} t_1 + \ldots + \beta_{3m} t_m + \epsilon \quad (2.3)$$

$$r = \gamma_0 + \gamma_1 L + \gamma_{21} x_1 + \ldots + \gamma_{2k} x_k + \gamma_{31} t_1 + \ldots + \gamma_{3m} t_m + \gamma_4 CLL + \nu \quad (2.4)$$

Using the assumption that under equilibrium, leverage demand equals leverage supply, we can substitute (2.4) for r in (2.3) and obtain:

$$
\begin{aligned}
L = \beta_0 + \beta_1[\gamma_0 + \gamma_1 L + \gamma_{21}x_1 + \ldots + \gamma_{2k}x_k + \gamma_{31}t_1 + \ldots + \gamma_{3m}t_m + \gamma_4 CLL + \nu] \\
+ \beta_{21}x_1 + \ldots + \beta_{2k}x_k + \beta_{31}t_1 + \ldots + \beta_{3m}t_m + \epsilon
\end{aligned}
$$

which gives the expanded reduced form equation (2.5) as:

$$
\begin{aligned}
L &= \frac{\beta_0 + \beta_1\gamma_0}{1-\beta_1\gamma_1} + \left(\frac{\beta_1\gamma_{21} + \beta_{21}}{1-\beta_1\gamma_1}\right)x_1 + \ldots + \left(\frac{\beta_1\gamma_{2k} + \beta_{2k}}{1-\beta_1\gamma_1}\right)x_k \\
&+ \left(\frac{\beta_1\gamma_{31} + \beta_{31}}{1-\beta_1\gamma_1}\right)t_1 + \ldots + \left(\frac{\beta_1\gamma_{3m} + \beta_{3m}}{1-\beta_1\gamma_1}\right)t_m + \beta_1\gamma_4 CLL + \frac{\beta_1\nu + \epsilon}{1-\beta_1\gamma_1} \qquad (2.5)
\end{aligned}
$$

The coefficients above represent the π_i's in the text and in particular, the coefficient on CLL is $\beta_1\gamma_4$. Similarly, upon substituting (2.3) for L in (2.4), we obtain:

$$
\begin{aligned}
r = \gamma_0 + \gamma_1[\beta_0 + \beta_1 r + \beta_{21}x_1 + \ldots + \beta_{2k}x_k + \beta_{31}t_1 + \ldots + \beta_{3m}t_m + \epsilon] \\
+ \gamma_{21}x_1 + \ldots + \gamma_{2k}x_k + \gamma_{31}t_1 + \ldots + \gamma_{3m}t_m + \gamma_4 CLL + \nu
\end{aligned}
$$

which gives the reduced form equation (2.6) as:

$$
\begin{aligned}
r &= \frac{\gamma_0 + \gamma_1\beta_0}{1-\beta_1\gamma_1} + \left(\frac{\gamma_1\beta_{21} + \gamma_{21}}{1-\beta_1\gamma_1}\right)x_1 + \ldots + \left(\frac{\gamma_1\beta_{2k} + \gamma_{2k}}{1-\beta_1\gamma_1}\right)x_k \\
&+ \left(\frac{\gamma_1\beta_{31} + \gamma_{31}}{1-\beta_1\gamma_1}\right)t_1 + \ldots + \left(\frac{\gamma_1\beta_{3m} + \gamma_{3m}}{1-\beta_1\gamma_1}\right)t_m + \gamma_4 CLL + \frac{\gamma_1\epsilon + \nu}{1-\beta_1\gamma_1} \qquad (2.6)
\end{aligned}
$$

As before, the coefficients above represent the z_i's in the text and in particular, the coefficient on CLL is uncontaminated, i.e., it is the structural coefficient γ_4. Furthermore, the ratio of the coefficient on CLL in (2.5) to that in (2.6), gives the coefficient $\beta_1 = \frac{\beta_1\gamma_4}{\gamma_4}$.

A.2 Data Appendix

A.2.1 Data Cleaning

We dropped loan-to-value ratios greater than 110%. For purchase mortgages, where loan-to-value was missing, we calculated it as the ratio of the unpaid loan balance at origination to the purchase amount. After making this calculation we dropped observations with missing loan-to-values or note rates. We also could not use the observations that had missing borrower age, income or gender. We calculated the Debt Service Coverage Ratio (or Debt to Income Ratio) as the ratio of the Borrower's Total Monthly (non housing) Debt Expenses to their Total Monthly Income. DSCRs greater than 0.65 we dropped. We dropped monthly incomes below $1000 and dropped credit scores above 850 and below 350. For credit score, we use credit scores at mortgage origination and where that information is not available, we instead use the credit score at the acquisition of the mortgage by Fannie Mae. To the extent that there is a large delay between origination and acquisition, these scores could potentially be different and introduce a measurement error in the credit score. Robustness tests using only the credit score at origination reveal that there is no substantial difference in the point estimates. We also drop observations that may have represented companies instead of a borrower. Finally, for the purposes of this paper, we restrict our sample to Fixed Rate Mortgages only.

A.2.2 Construction of MSA Price Level and Volatility

The MSA price levels for the year 2000 were constructed by running a hedonic regression of the log house value on housing characteristics and MSA dummies, using the 2000 PUMS 5% sample. The regression results are shown in Table A.2.1. The coefficients from this regression were weighted by sample means and proportions of housing characteristics in each MSA. The predicted log house value was then the sum of the MSA specific dummy and the hedonic sum of characteristics in each market. The log value was exponentiated to arrive at the 2000 MSA price level. These price levels are mapped in Figure D.1 and show a very reasonable distribution across the US. The final step in calculating the MSA price level for the years 1986 to 2010 was achieved by extrapolating the 2000 price level using the MSA repeat sale price indices published by FHFA.

The detrended HPI volatility variable was calculated by first running a regression of Ln MSA House Price Level (constructed as detailed above) on a continuous time variable plus MSA dummies. The log residuals from this regression were then used to calculate the 2-year lagged, standard deviation used in the text.

Table A.2.1: Price Level Regression - PUMS 5 percent sample

	Log House Value
rooms==2	0.20
	(0.03)
rooms==3	0.35
	(0.03)
rooms==4	0.35
	(0.03)
rooms==5	0.50
	(0.03)
rooms==6	0.65
	(0.03)
rooms==7	0.81
	(0.03)
rooms==8	0.97
	(0.03)
rooms==9	1.24
	(0.03)
builtyr==2-5 years	-0.05
	(0.00)
builtyr==6-10 years	-0.12
	(0.00)
builtyr==11-20 years	-0.23
	(0.00)

Table A.2.1: Price Level Regression - PUMS 5 percent sample

	Log House Value
builtyr==21-30 years	-0.36
	(0.00)
builtyr==31-40 years	-0.42
	(0.00)
builtyr==41-50 years	-0.48
	(0.00)
builtyr==51-60 years	-0.56
	(0.00)
builtyr==61+ years	-0.57
	(0.00)
unitsstr==1-family house, attached	-0.20
	(0.00)
kitchen==Yes	0.34
	(0.01)
msa/cbsa dummies	yes
Constant	10.36
	(0.03)
R^2	0.50
N	2,318,561

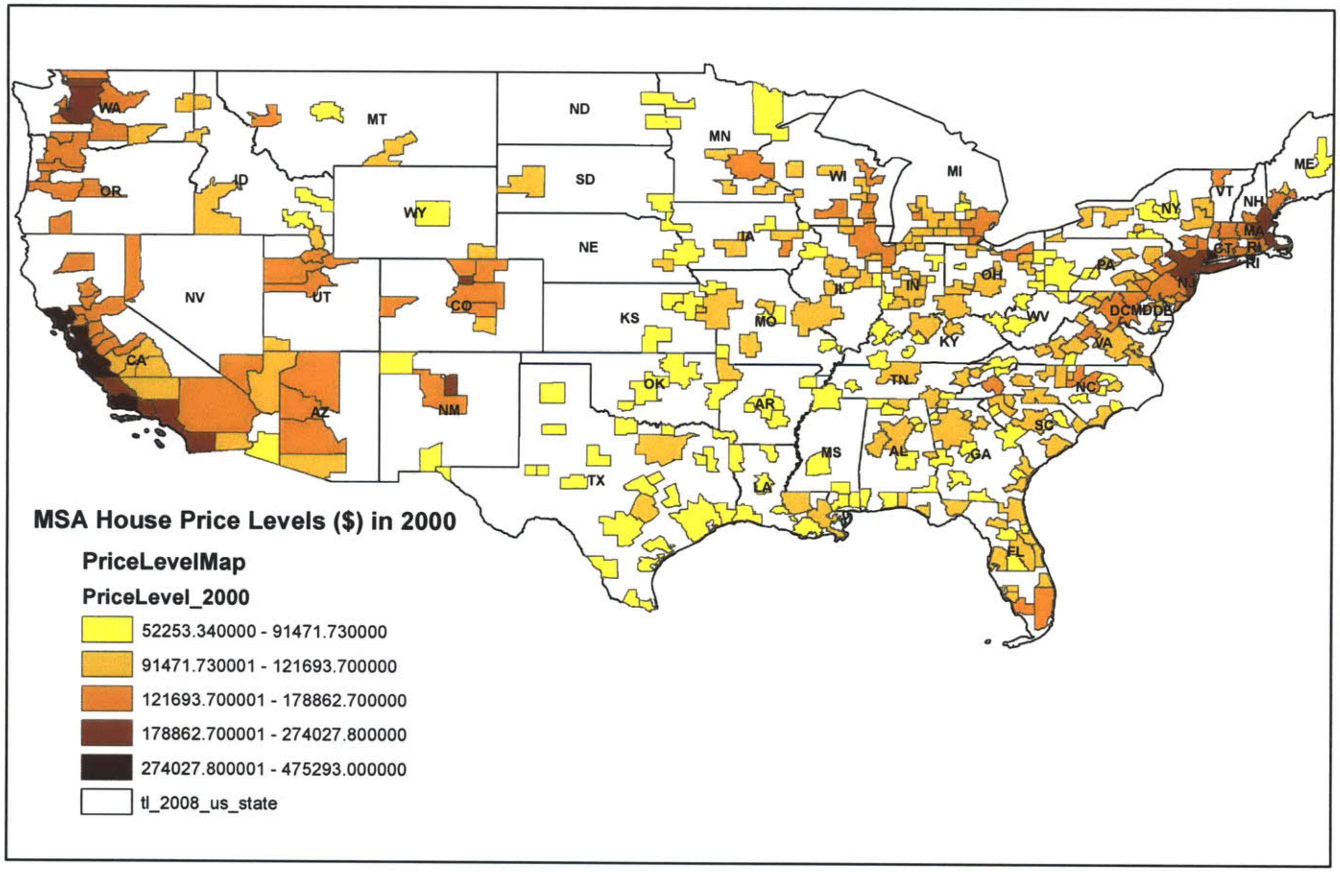


Figure A.2.1: MSA 2000 Price Levels

Appendix B

Appendix to Chapter 4

B.1 The Moore-Penrose Pseudoinverse or the Generalized Inverse

The Moore-Penrose pseudoinverse is a general way of solving the following system of linear equations:

$$y = Xb, \quad y \in R^n, \quad b \in R^k, \quad X \in R^{nxk}$$

It can be shown that there is a general solution to these equations of the form:

$$b = X^{\dagger} y \tag{B.1}$$

The $X^{\dagger}$ matrix is the unique Moore-Penrose pseudoinverse of X that satisfies the following properties:

1. $XX^{\dagger}X = X$ ($XX^{\dagger}$ is not necessarily the identity matrix)

2. $X^{\dagger}XX^{\dagger} = X^{\dagger}$

3. $(X^{\dagger}X)^{\intercal} = XX^{\dagger}$ ($XX^{\dagger}$ is Hermitian)

4. $(X^{\dagger}X)^{\intercal} = X^{\dagger}X$ ($X^{\dagger}X$ is also Hermitian)

The solution given by equation (B.1) is a minimum norm least squares solution. When X is of full rank (i.e., rank is at most $min(n, k)$), the generalized inverse can be calculated as follows:

- Case 1: When $n = k$ (same number of equations as unknowns): $X^{\dagger} = X^{-1}$
- Case 2: When $n < k$ (fewer equations than unknowns): $X^{\dagger} = X^{\intercal}(XX^{\intercal})^{-1}$
- Case 3: When $n > k$ (more equations than unknowns): $X^{\dagger} = (X^{\intercal}X)^{-1}X^{\intercal}$

In the application for deriving higher frequency indexes from staggered lower frequency indexes, Case 2 provides the relevant calculation. Furthermore, it should be noted that when the rank of X is less than k, no unbiased linear estimator, b, exists. However, for such a case, the generalized inverse provides a minimum bias estimation.[1] For the basic references on the Moore-Penrose pseudoinverse see the references by Penrose (1955, 1956), Chipman (1964), and Albert (1972) in the bibliography.

B.2 A Note on the Bias in the Generalized Inverse Estimator (GIE)

Here we consider the case relevant to our present purposes, i.e. where $X^{\dagger} = X^{\intercal}(XX^{\intercal})^{-1}$. Therefore, in our application, the solution (or estimation) of the second-stage regression (equation (B.1)) can be re-written as:

$$b = X^{\dagger} = X^{\intercal}(XX^{\intercal})^{-1}y$$

Considering that the true value of the predicted variable (y) is by definition: Xb_{true}, therefore the expected value of b is:

$$E[b|X] = X^{\intercal}(XX^{\intercal})^{-1}Xb_{true}$$

[1]Properties of the generalized inverse can be found in Penrose (1954) and equation (2) first appeared in Penrose (1956). Proofs of Cases 1, 2 and 3 can be found in Albert (1972) and a proof of minimum biasedness is given in Chipman (1964).

Let $R = X^{\intercal}(XX^{\intercal})^{-1}X$ be the "resolution" matrix, which would have otherwise been the k by k identity (I) matrix if X had been of full column rank. In our case, the resolution matrix is instead a symmetric matrix describing how the generalized inverse solution "smears" out the b_{true} into a recovered vector b. The bias in the generalized inverse solution is

$$E[b|X] - b_{true} = Rb_{true} - b_{true} = (R - I)b_{true}$$

We can formulate a bound on the norm of the bias:

$$\|E[b|X] - b_{true}\| \leq \|R - I\| \|b_{true}\|$$

Computing $\|R - I\|$ can give us an idea of how much bias has been introduced by the generalized inverse solution. However, the bound is not very useful since we typically have no knowledge of $\|b_{true}\|$.

In practice, we can use the resolution matrix, R, for two purposes. First, we can examine the diagonal elements of R. Diagonal elements that are close to one correspond to coefficients for which we can expect good resolution. Conversely, if any of the diagonal elements are small, then the corresponding coefficients will be poorly resolved.

For the particular data matrix used in this study, i.e X is Toeplitz, the diagonal elements of R approach one very fast. For instance, for the annual to quarterly conversion, a 24 by 27 matrix (24 observations, 27 quarterly return estimates), the diagonal elements of R have a value of 0.89. For a 50 by 53 matrix, the diagonal elements have a value of 0.94. By induction, as the number of periods to be estimated (T) go to infinity, and the percentage difference between T and T-3 becomes negligible, the diagonal elements of R approach a value of 1. Hence, the bias goes to zero as the system gets closer to being effectively identified.

Bibliography

[1] Acharya, V., M. Richardson, S. V. Nieuwerburgh, and L. White. 2011. *Guaranteed to Fail: Fannie Mae, Freddie Mac and the Debacle of Mortgage Finance.* Princeton University Press.

[2] Adelino, M., A. Schoar, and F. Severino. 2011. Credit Supply and House Prices: Evidence from Mortgage Market Segmentation. MIT Sloan Working Paper.

[3] Albert, A. 1972. *Regression and the Moore-Penrose Pseudoinverse.* Academic Press, New York.

[4] Aster, R.C, Borchers, B. and Thurber, C. H. 2005. *Parameter Estimation and Inverse Problems.* Elsevier Academic Press

[5] Benartzi, S. and R. Thaler. 1995. Myopic Loss Aversion and the Equity Premium Puzzle. *Quarterly Journal of Economics* 110: 73 - 92.

[6] Black, R. and J. Diaz III. 1996. The Use of Information versus Asking Price in the Real Property Negotiation Process. *Journal of Property Research* 13: 287 - 297.

[7] Brueckner, J. K. 1994. The demand for mortgage debt: some basic results. *Journal of Housing Economics* 3(4): 251262.

[8] Brueckner, J. K. 2000. Mortgage default with asymmetric information. *The Journal of Real Estate Finance and Economics* 20(3): 251274.

[9] Bryan T. and P.Colwell. 1982. Housing Price Indices. In: C.F.Sirmans eds. *Research in Real Estate*, vol.2. Greenwich, CT: JAI Press.

[10] Bucchianeri,G. and J.Minson. 2011. Listing Behaviors and Housing Market Outcomes. Working Paper, Wharton School of the University of Pennsylvania.

[11] Case, K. and R. Shiller. 1987. Prices of Single Family Homes Since 1970: New Indexes for Four Cities. *New England Economic Review*: 45-56.

[12] Cheng, P., Z. Lin, and Y. Liu. 2011. Do Women Pay More for Mortgages? *The Journal of Real Estate Finance and Economics* 43(4): 423440.

[13] Chipman, J. 1964. On Least Squares with Insufficient Observations. *Journal of the American Statistical Association* 59, No. 208, 1078-1111.

[14] Clapp, J. 2004. A Semiparametric Method for Estimating Local House Price Indices. *Real Estate Economics* 32: 127-160.

[15] Crane, A. and J. Hartzell. 2009. Is there a Disposition Effect in Corporate Investment Decisions? Evidence from Real Estate Investment Trusts. Working Paper, The University of Texas at Austin.

[16] Elliehausen, G. 2010. Mortgage contract choice in subprime mortgage markets. Federal Reserve Working Paper Series.

[17] Feng, L. and M.S. Seasholes. 2005. Do Investor Sophistication and Trading Experience Eliminate Behavioral Biases in Financial Markets? *Review of Finance* 9: 305 - 351

[18] Fisher,J., D.Gatzlaff, D.Geltner, & D.Haurin. 2003. Controlling for the Impact of Variable Liquidity in Commercial Real Estate Price Indices. *Real Estate Economics*, 31: 269 - 303.

[19] Fisher, J., D. Gatzlaff, D. Geltner and R. Haurin. 2004. An Analysis of the Determinants of Transaction Frequency of Institutional Commercial Real Estate Investment Property. *Real Estate Economics* 32: 239 - 264.

[20] Follain, J. 1990. Mortgage Choice. *Real Estate Economics* 18(2): 125144.

[21] Follain, J. and R. Dunsky. 1997. The demand for mortgage debt and the income tax. *Journal of Housing Research* 8: 155200.

[22] Francke, M. 2010. Repeat Sales Index for Thin Markets: A Structural Time Series Approach. *Journal of Real Estate Finance and Economics:*

[23] Geanakoplos, J. 2009. The leverage cycle. Number 24 in *NBER Macroeconomics Annual.*

[24] Geltner, D. 1997. Bias and Precision of Estimates of Housing Investment Risk Based on Repeat-Sales Indexes: A Simulation Analysis. *Journal of Real Estate Finance and Economics*, 14: 155-172.

[25] Geltner, D. 2000. Benchmarking Manager Performance Within the Private Real Estate Investment Industry. *Real Estate Finance* 17: 1-11.

[26] Geltner, G. and D. Ling. 2006. Considerations in the Design and Construction of Investment Real Estate Research Indices. *Journal of Real Estate Research*, 28: 411-444.

[27] Geltner, G, N. Miller, J. Clayton, and P. Eichholtz. 2007. *Commercial Real Estate Analysis and Investments* 2nd Edition, South-Western College Publishing, Cincinnati.

[28] Genesove, D. and C. Mayer. 2001. Loss Aversion and Seller Behavior: Evidence from the Housing Market. *Quarterly Journal of Economics* 166: 1233 - 1260.

[29] Goetzmann. W. 1992. The Accuracy of Real Estate Indices: Repeat-Sale Estimators. *Journal of Real Estate Finance and Economics* 5: 5-54.

[30] Goetzmann, W. N., L. Peng, and J. Yen. 2011. The Subprime Crisis and House Price Appreciation. *The Journal of Real Estate Finance and Economics*

[31] Haigh, M. and J. List. 2005. Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis? *Journal of Finance* 60: 523 - 534.

[32] Harrison, D., T. Noordewier, and A. Yavas. 2004. Do Riskier Borrowers Borrow More? *Real Estate Economics* 32(3): 385411.

[33] Hendershott, P., D. Pryce, and D. White. 2002. Household leverage and the deductibility of home mortgage interest: evidence from UK house purchasers. *Journal of Housing Research* 14(1): 4982.

[34] Hendershott, P. H. and W. C. LaFayette. 1997. Debt Usage and Mortgage Choice: The FHA-Conventional Decision. *Journal of Urban Economics* 41(2): 202217.

[35] Jaffee, D. and F. Modigliani. 1969. A Theory and Test of Credit Rationing. *The American Economic Review* 59(5): 850872.

[36] Jaffee, D. and T. Russell. 1976. Imperfect information, uncertainty, and credit rationing. *The Quarterly Journal of Economics* 90(4): 651666.

[37] Jaffee, D. and J. Stiglitz. 1990. Credit Rationing. *Handbook of Monetary Economics* 2: 836888.

[38] Jones, L. 1993. The demand for home mortgage debt. *Journal of Urban Economics* 33: 1028.

[39] Kahneman, D. and A. Tversky. 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica* 47: 263 - 291.

[40] Lamont, O. and J. C. Stein. 1999. Leverage and House-Price Dynamics in U.S. Cities. *The RAND Journal of Economics* 30(3): 498514.

[41] Leece, D. 2006. Testing a theoretical model of mortgage demand on UK data. *Applied Economics* 38(17), 20372051.

[42] Ling, D. C. and G. A. McGill. 1998. Evidence on the Demand for Mortgage Debt by Owner-Occupants. *Journal of Urban Economics* 44(3): 391414.

[43] List, J. 2003. Does Market Experience Eliminate Market Anomalies? *Quarterly Journal of Economics* 118: 41 - 71.

[44] Locke, P. and S. C. Mann. 2000. Do Professional Traders Exhibit Loss Realization Aversion? Working Paper, Texas Christian University.

[45] McMillen, D and J.Dombrow. 2001. A Flexible Fourier Approach to Repeat Sales Price Indexes. *Real Estate Economics* 29: 207-225.

[46] Mian, A. and A. Sufi. 2009. The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis. *The Quarterly Journal of Economics* 124(4): 1449.

[47] Mian, A. and A. Sufi. 2011. House prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis. *American Economic Review* 101(1): 21322156.

[48] Northcraft, G. and M. Neale. 1987. Experts, Amateurs and Real Estate; An Anchoring-and-Adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes* 39: 84 - 97.

[49] Odean, T. 1998. Are Investors Reluctant to Realize their Losses? *Journal of Finance* 53: 1775 - 1798.

[50] Pace, R. K and J. P. LeSage. 2004. Spatial Statistics and Real Estate. Journal of Real Estate Finance and Economics, 29:147-148.

[51] Pavia, J.M. and Bernardi Cabrer. 2008. On distributing quarterly national growth among regions. *Environment and Planning A* 1 - 16.

[52] Penrose, R. 1955. A Generalized Inverse for Matrices. *Proceedings of the Cambridge Philosophical Society* 51, 406-413.

[53] Penrose, R. 1956. On best approximate solutions of linear matrix equations. *Proceedings of the Cambridge Philosophical Society* 51, 17-19.

[54] Pryke, M. and P. Du Gay. 2002. Property companies and the remaking of markets: stories from the 1990s. In: Guy, S. and J. Henneberry eds. *Development and Developers: Real Estate Issues.* Oxford, UK: 267 - 284.

[55] Reinhart, C. M. and K. S. Rogoff. 2008. This Time is Different: A Panoramic View of Eight Centuries of Financial Folly. NBER working paper 13882, 1123.

[56] Schwann, G. 1998. A Real Estate Price Index for Thin Markets. *Journal of Real Estate Finance and Economics* 16: 269-287.

[57] Shapira, Z. and I. Venezia. 2001. Patterns of Behavior of Professionally Managed and Independent Investors. *Journal of Banking and Finance* 25: 1573 – 1587.

[58] Thaler, R. 1985. Mental Accounting and Consumer Choice. *Marketing Science* 4: 199 - 214.

[59] Tversky, A. and D. Kahneman. 1991. Loss Aversion in Riskless Choice: A Reference-Dependent Model. *Quarterly Journal of Economics* 106: 1039 - 1061.

[60] Tversky, A. and D. Kahneman. 1974. Judgment Under Uncertainty: Heuristics and Biases. *Science* 185: 1124-1130.

[61] Yavas, A. and S. Yang. 1995. The Strategic Role of Listing Price in Marketing Real Estate: Theory and Evidence. *Real Estate Economics*23: 347 - 368.