

# **The Investment Performance of Housing and “Hedonic” Spatial Equilibrium**

By

Tracey Seslen  
Marshall School of Business  
University of Southern California

William C. Wheaton  
Department of Economics  
MIT

Henry O. Pollakowski  
Center for Real Estate  
MIT

DRAFT: May, 2005

The authors are indebted to CSW/FISERV, the Warren Group, and Dataquick, Inc. for the provision of data. The Authors remain fully responsible for all conclusions and analysis drawn from this research.

## ABSTRACT

This research unites the two major strands of work that exist to date in the literature on Housing Markets. The first is the notion of spatial equilibrium wherein consumers inhabiting different units are thought to be at a constant utility level. As a consequence prices “compensate” for differential “hedonic” housing attributes. The second is the application of life-cycle analysis to the determination of the “full” cost of owning housing as a financial asset. Linking the two we hypothesize that it is the “risk-adjusted-annual-cost-of-ownership” which should compensate owners for the differential consumption flows that come from various houses. To test whether this is the case we develop a unique data set for 4 US metropolitan areas that ascertains the appreciation and risk from owning housing at the ZIP code level. We then combine this with transaction based data on Price levels – at the same level of geographic detail. We find that in ZIP codes with higher historic appreciation, price levels are indeed higher, but we suspect that this may represent misspecification through an identity. When we test over a shorter period for whether prices anticipate future appreciation – we get very poor results. We do not observe consistent, expected results across all specifications for any of the 4 MSAs. The results for risk are similarly disappointing. In only 1 market does risk seem to be priced reasonably. These results reinforce the doubts raised by others over whether the housing market is “efficiently” priced. They also suggest that three decades of “hedonic” research could possibly be biased by omitting the investment characteristics of housing assets from the estimated equations.

## **I: INTRODUCTION**

This paper connects two central ideas in the long literature on Housing markets. The first idea, Ricardian Rent, is hundreds of years old. Ricardo [1817] hypothesized that in equilibrium, land would absorb the advantages of location and its price would hence exactly compensate for any and all attributes that either consumers or producers would “value”. Following a significant expansion by Alonso [1967] and then Rosen [1974], there have appeared hundreds of papers using the idea of spatial equilibrium to implicitly value travel time, public goods and environmental externalities (e.g. Smith [1995], Bartik [1987]), not to mention a host of housing attributes (e.g. Case et. al [1989], Palmquist [1984], Brown and Rosen [1982]) . Recently there has arisen some criticism over the empirical specification of “hedonic” models (e.g. Epple [1987], Ekeland et. al [2004]) , but the central theoretical premise of prices acting to “compensate” remains quite central to much applied research.

The second strand in the housing literature is inter-temporal and deals with the consumption of housing services over time and the investment returns contained therein. Starting with Kearn [1979], Schwab [1982], Dougherty and Van Order [1982] economists realized that the un-taxed nominal capital gains earned by owning housing could bestow a major advantage to home ownership (Hamilton and Schwab [1985]). Poterba [1984] put this idea into a rational expectations equilibrium framework and demonstrated that unanticipated shocks to housing demand would generate anticipated patterns of future price appreciation that would generate further increases in housing demand. After this it was well recognized that the “user cost” of owning housing incorporated expected future capital gains as well as implicit rent – even with trading frictions (Grossman and Laroque [1990]).

A union of these two literatures began with a paper by Capozza and Helsley [1990] in which the authors showed that in a Ricardian equilibrium for a growing city, prices would have to not only compensate for location attributes but also for the anticipated changing valuation of those attributes over time due to growth. Dipasquale and Wheaton [1996] expounded upon this and showed how with anticipated growth, certain locations within a city would have both high price levels and high growth in

Ricardian rent, while other areas would exhibit the reverse. In equilibrium, a “user cost” measure of Rent-minus-appreciation would be exactly the same across locations.

The present paper takes this union one further step arguing that it should be Rent-minus-appreciation-plus risk that is equilibrated across locations. Empirically, the implication is that all hedonic equations should include the two investment dimensions of housing – expected appreciation and risk – in addition to observable physical and location characteristics. To the extent that such investment behavior is correlated with measured attributes, the omission biases results.

To test these ideas empirically for the first time we undertake a several-part study. We begin by obtaining repeat sale price indices at the ZIP-code level for 4 MSAs. These indices span roughly 25 years (from 1979 to 2003) and reveal several conclusions. First, virtually all ZIP codes do in fact closely follow the cyclic movements of the broader MSA. Secondly, these cyclic movements are quite predictable as other authors have argued (e.g. Case et. al [1989]). Thirdly, there are significant differences across ZIP codes (within an MSA) in longer term appreciation and risk. Finally, there is the expected positive simple correlation between risk and historic return across ZIP levels – although it is far from complete.

We next produce hedonic price equations for a single year within but near the end of those spanned by the indices (1998). We do this with thousands of transactions in each MSA and each transaction is linked to a ZIP code. Incorporating measured risk and appreciation measures into the hedonic equations we find the following:

1) Historic appreciation is reflected in 1998 prices in roughly the magnitude that would be suggested by theory. This occurs in all 4 of our MSAs.

2). However, several measures of *ex ante* appreciation (from 1998-2003) are not reflected in 1998 prices

3) Risk (which can only be measured historically) is correctly reflected in price in only 1 of 4 MSAs.

4). Not surprisingly, risk and return are correlated particularly with some location variables and hence their omission changes the estimated impacts of those variables. Alternatively, some of the impact of location or housing variables on house prices comes not from consumer’s valuing the flow of services they produce, but rather from the fact

that certain locations or types of housing have higher or lower expected long term appreciation.

Our paper proceeds in accordance with the discussion above. In Section II we review some theoretical literature and develop a simple model to illustrate how the historic investment behavior of housing markets can be expected to impact current or future consumption decisions. Section III then describes the repeat sale price indices obtained, their behavior over time, and offers several different measures of risk and appreciation. We present these metrics in several different ways and show that they are highly autocorrelated over time. We also show that risk and appreciation are positively related historically across ZIP codes and that there are strong location patterns to the investment performance of housing

In Section IV, we describe the data used to construct our hedonic equations in each MSA. We use a number of specifications with and without investment variables, but all appreciation measures are historically based over the full time period. We compare results across MSAs and find that the point estimates for historic appreciation are reasonable according to several theoretical perspectives.

In Section V we argue that the encouraging results of the previous section could be due to an identity, and remove this possible specification error by determining if anticipated future appreciation is reflected in prices. Our results in this case are quite poor and raise the prospect that ex ante investment performance is not well priced. We draw some conclusions which reinforce the literature on housing market inefficiency and suggest further research.

## **II: HOW SHOULD RISK AND APPRECIATION BE PRICED IN HOUSING?**

In this section we develop a simple model of housing consumption to illustrate how the expected (and possibly historic) appreciation and risk characteristics of housing markets determine the consumption decision. In short run equilibrium, with supply largely fixed, the impact on desired consumption will be directly translated into housing prices. In long run equilibrium, some of these impacts will be tempered by supply response, but with less than perfect supply elasticity, they will still hold qualitatively.

We begin with a 2 period model with two forms of consumption -- housing (a durable,  $h$ ) and a numeraire (non-durable,  $c$ ). In the first period, the individual has total wealth  $W_0$ , and must decide how much to allocate of this in the next period between riskless asset  $b$  and housing wealth  $ph$ . In the second period the return on  $b$  is  $R$  which is known with certainty. The appreciation on housing is  $H$  is known with *uncertainty* and comes from the probability distribution of the change in  $p$  over the second period. Efficient capital markets allow the individual to finance its numeraire consumption in the second period with the total return on wealth:  $phH + Rb$ .

In this framework, the individual has preferences represented by the CARA utility function with no intermediate consumption. Returns on housing,  $H$ , are normally distributed with mean  $\mu_H$  and variance  $\sigma_H^2$ . The individual maximizes:

$$U(c, h) = -e^{-ac - ah} \quad (1)$$

subject to

$$W_0 = ph + b$$

and

$$c = phH + Rb$$

Solving the maximization

$$\max E[-e^{-ac - ah}] = \max E[-e^{-a(phH + Rb) - ah}],$$

we get optimal housing expenditure

$$ph^* = \frac{\mu_H - R + \frac{\alpha}{a}}{a\sigma_H^2} \quad (2)$$

Thus optimal housing expenditure will increase when the appreciation on housing is expected to be greater, will decrease as other investments have higher returns and will decrease with the greater the risk associated with homeownership. It is purely a function of the moments of housing returns, and wealth effects do not come into play.

In a short run equilibrium in the housing market,  $h^*$  is fixed and so price must move positively with expected future return and negatively with both the opportunity cost

of investing as well as expected future housing risk. In fact the expression above could be used to derive point estimates of what the relationship should look like. As the time horizon lengthens, the housing stock can adjust and so the expression above becomes less binding. The qualitative relationship between prices, return and risk, however, still holds.

In the literature on housing consumption (e.g. Poterba [1984]), there is often the expression “user cost” to represent the annual cost of a dollar of housing expenditure (here  $ph$ ). The literature generally expresses this in a linear form as something like:  $R - \mu_H + a\sigma_H^2$ . With the CARA utility function, we more directly derive the expression below.

$$user\ cost = \frac{a\sigma_H^2}{\mu_H - R + \frac{\alpha}{a}} \quad (3)$$

Housing consumption should be inverse to the product of the price level  $p$  and the user cost expression above or alternatively inverse to the price level of housing ( $p$ ), the variance in housing returns  $\sigma_H^2$  and positively linearly related to the difference between expected housing appreciation  $\mu_H$  and the opportunity cost of investing ( $R$ ).

When the housing market is differentiated by a series of fixed location or housing attributes  $X$ , we need to generalize the notion of equilibrium. In the case of a product differentiated equilibrium, the total cost of owning a package of housing must compensate for the utility flow of services from that package (measured with the variables  $X$  and the indirect utility function  $U$ ). This will involve an equilibrium expression such as below:

$$p \frac{a\sigma_H^2}{\mu_H - R + \frac{\alpha}{a}} = U(X) \quad (4)$$

or alternatively an expanded hedonic equation:

$$p = \frac{(\mu_H - R + \frac{\alpha}{a})}{a\sigma_H^2} U(X) \quad (5)$$

In all of this discussion, the housing returns in the user cost equation are expected future returns (in the second period). In our (and others') empirical research, it has become clear that the moments of the probability distribution of housing returns  $\mu_H$  and  $\sigma_H^2$  may be time dependent on past movements in  $H$ . The positive observed autocorrelation between returns in adjacent periods certainly suggests that the process is at least AR(1).

$$H_t = \lambda + \gamma H_{t-1} + \varepsilon_t \quad (6)$$

In this case, the variance of  $H_t$  remains unchanged, since it is a function of (non-stochastic) past returns and a stochastic error component. This will be  $\sigma_H^2$ . The unconditional and conditional expected values of  $H_t$  however are quite different.

$$E[H_t] = \lambda / (1 - \gamma) = \mu_H \quad (7)$$

$$E[H_t | H_{t-1}] = \lambda + \gamma H_{t-1} = \mu'_H$$

If consumers are fully informed, then they realize the positive autocorrelation in the market and current period consumption will be quite closely connected to historic returns as long as  $1 > \gamma > 0$ . Thus when prices are on the upswing desired consumption will be as well, pushing prices further upward. This is not "irrational" and many authors (e.g. Wheaton [1999]) show that positively autocorrelated prices are an intrinsic feature of housing or real estate markets where the supply of durable capital can take considerable time to bring on line. The result is that with empirical work, user cost measures may be generated using recent or historic movements of the housing returns data without necessarily violating economic rationality.



### **III: THE INVESTMENT PERFORMANCE OF ZIP-LEVEL HOUSING MARKETS.**

Empirical evidence on the risk, appreciation and predictability of housing prices at the metropolitan level has been well-documented, starting with Case and Shiller [1989]. Using transactions and other administrative data from four major metropolitan areas, their study showed strong positive autocorrelation in returns over short intervals; an increase in housing prices in the current period predicted an increase  $\frac{1}{4}$  to  $\frac{1}{2}$  as large in the following period. Similar results from a larger sample of metropolitan areas were estimated in Capozza, Hendershott, and Mack [2004] and Seslen [2004]. No study, as of yet, has extended this type of analysis to the sub-metropolitan level.

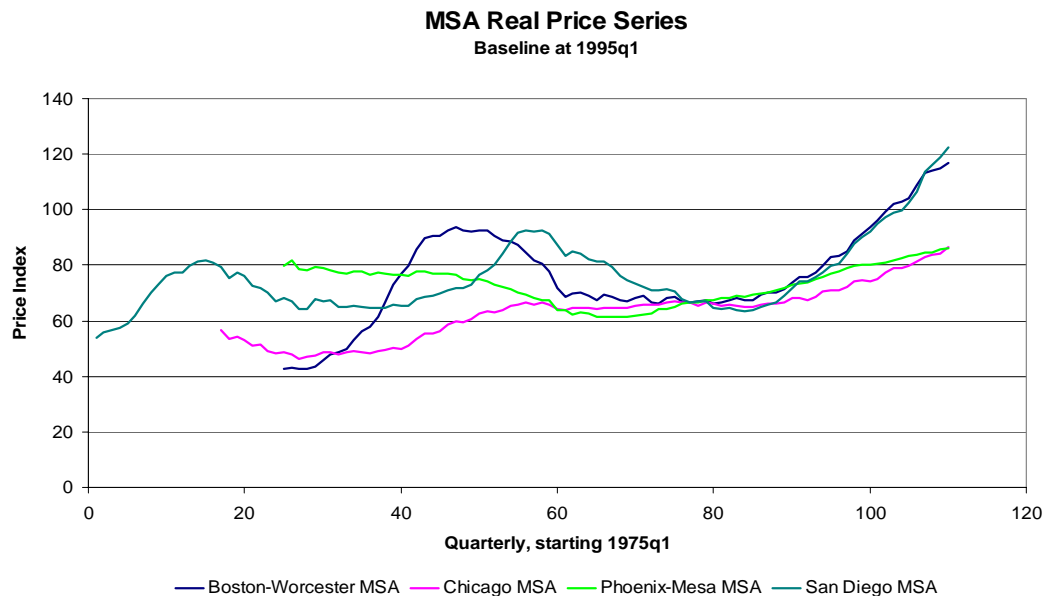
In this section, we use weighted repeat-sales housing price indices provided to us by Case Shiller Weiss/FISERV at the ZIP code level to examine a series of questions. These include whether ZIP level housing prices behave closely with their MSA aggregate index or whether there are wide differences. We next present a set of time series metrics to characterize each ZIP series. These include simple risk and appreciation measures as well as the parameters of a univariate model (estimated separately for each ZIP). The data cover four MSAs: Boston, Chicago, Phoenix, and San Diego. In choosing these areas, we have attempted to create a sample representing a diverse set of demographic, geographic, and housing market-related conditions.

The Boston metropolitan area covers 249 ZIP codes from 1982 through 2002, Chicago comprises 152 ZIP codes from 1987 through 2002; Phoenix includes 164 ZIP codes spanning 1988 to 2002, and San Diego covers 86 ZIP codes starting in 1975. In the final version of our empirical model, the data were kept or dropped based on two conditions: proximity to the center of the MSA, and availability of other data needed to carry out the later stages of our analysis. For Boston, we confined our sample to those ZIP codes within the I-495 beltway; outside that area, it could be argued that Boston is not the primary center of economic “pull” on housing prices, and we did not want that possibility to cloud our results. In the end, over half of the observations for Boston were dropped due to distance from the Boston city center. In Phoenix and San Diego, less than five ZIP codes were dropped as a result of being geographical “outliers”. The final sample resulted in 110 observations for Boston, 76 for Chicago, 88 for Phoenix, and 44

for San Diego. For purposes of discussion, price levels are deflated using the (urban) consumer price index. The risk and appreciation measures used in estimation are all calculated in current dollars.

As can be seen in Figure 1, housing prices across the four MSAs were behaving quite differently from one another over the last two decades. While Boston and San Diego were experiencing significant boom-bust cycles, Chicago and Phoenix were progressing through substantially less volatile paths.

**Figure 1:**

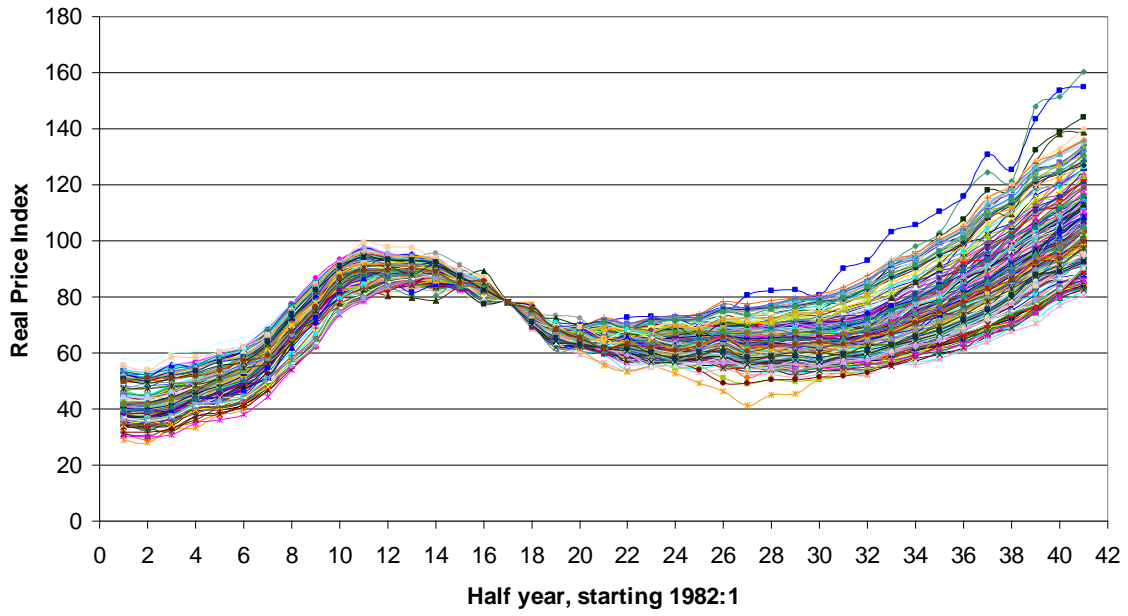


### Price Movements at the ZIP Level

In Figures 2 through 5, we have graphed the price series for every ZIP in each of the four MSAs again in constant dollars. Within Boston and San Diego, the ZIP-level series follow one another quite closely, while in Chicago and Phoenix, they exhibit greater divergence from one another, particularly towards the end of the sample period. From first glances, it would appear that “a rising tide raises all boats.”

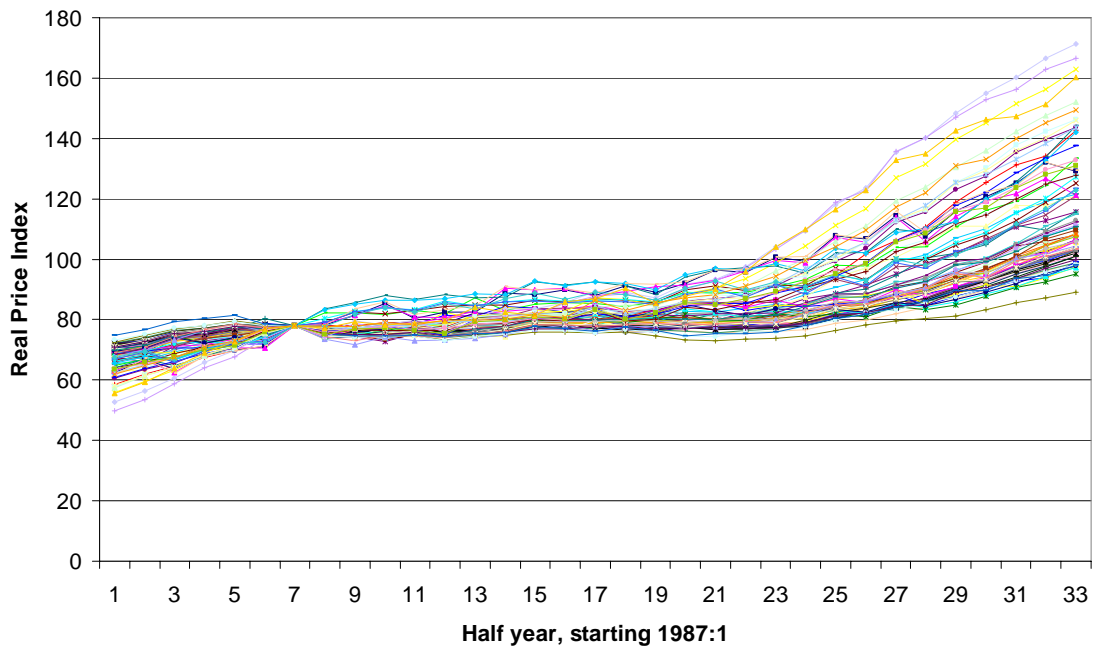
**Figure 2:**

**Price index evolution, 110 Boston zip codes**



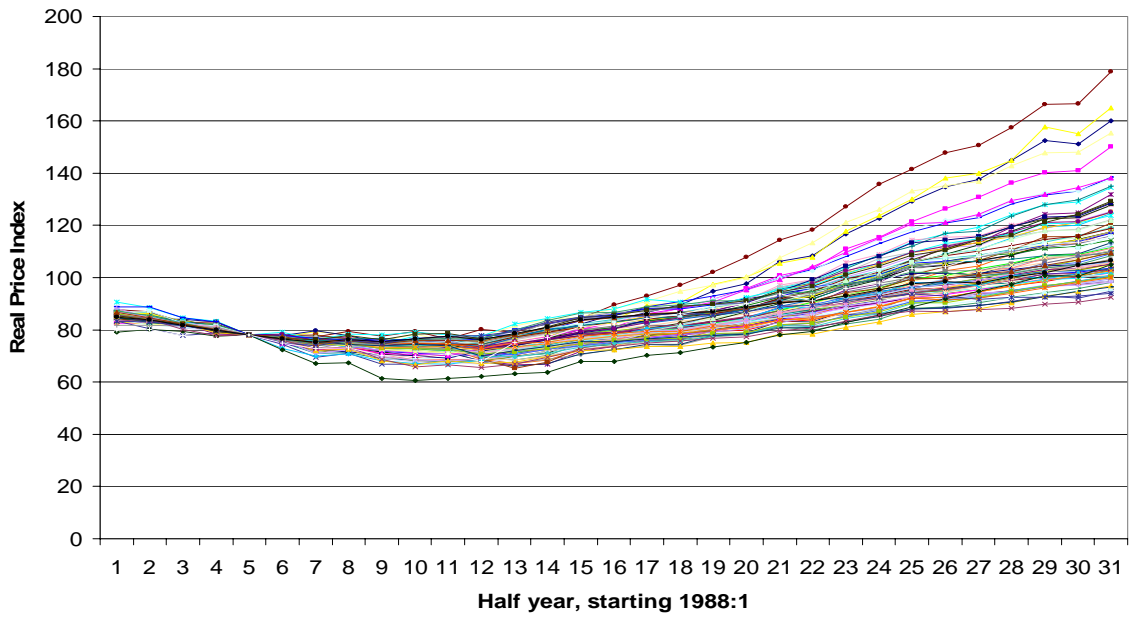
**Figure 3:**

**Price index evolution, 76 Chicago zip codes**



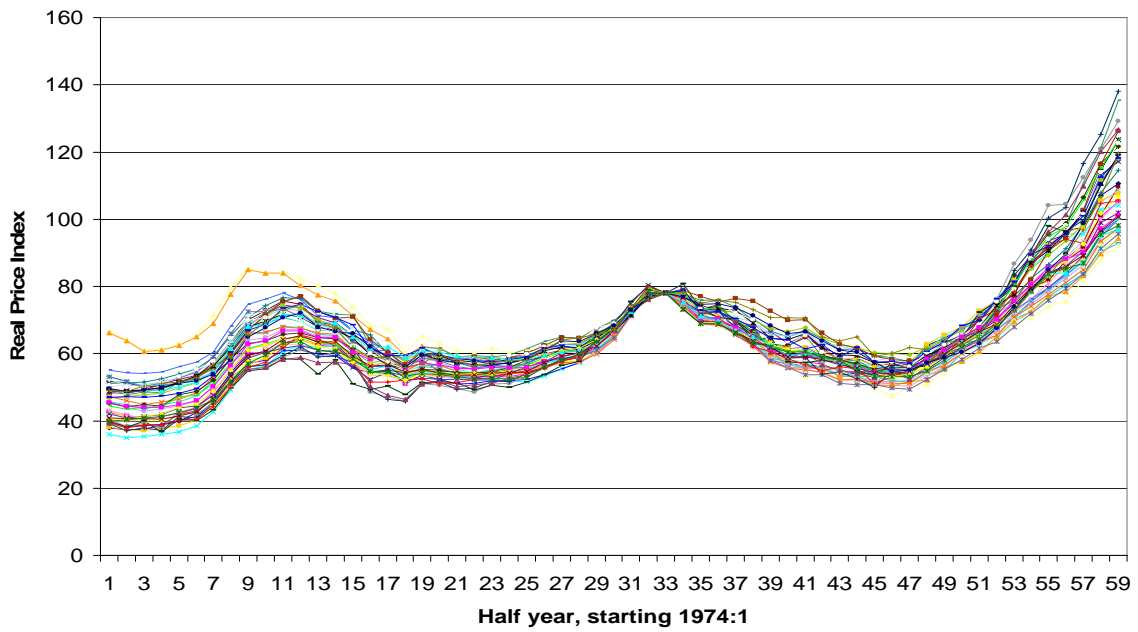
**Figure 4:**

**Price index evolution, 88 Phoenix zip codes**



**Figure 5:**

**Price index evolution, 44 San Diego zip codes**



It is possible to more formally characterize the distribution of ZIP series for each MSA. In Appendix 1, we provide the ZIP level quantile distribution of (current dollar) risk and appreciation for each market. Table 1 below summarizes those results and shows that for most markets, 90% of the ZIP codes have raw risk and appreciation (in current dollars) that spans two or three percentage points or alternatively lies with 1.5% either side of the MSA mean.

**Table 1: distribution of ZIP statistics (current \$):**

<u>MSA</u>	<u>MSA Return</u>	<u>90% Return</u>	<u>90% Risk</u>
Boston	8.5	7.5 – 9.5	8.5 – 11.5
Chicago	6.5	5.0 – 9.5	2.5 – 5.0
Phoenix	5.0	3.8 – 7.0	2.5 – 5.0
S. Diego	7.6	6.5 – 8.5	7.7 – 10.0

In one sense the range breadth in the distribution of return and risk across ZIPs is a measure of the degree to which each MSA is “spatially integrated”. If all ZIP codes are close substitutes and demand very elastic across areas then presumably ZIP codes would behave quite closely. Conversely, if there is lower spatial demand elasticity and ZIP codes are quite differentiated then the variation in behavior would be greater. By this standard, Boston and San Diego are the more “integrated” while Chicago and Phoenix are less “integrated”. In all these markets there is a significant positive correlation between risk and return – although this will be discussed in more detail later.

As an alternative to using simple descriptive statistics to examine the behavior of ZIP level price series we also estimate a basic autoregressive model for each series – and then examine the spread in the parameter distribution of this model across the ZIPs. The rationale for this is based on the idea that if some component of price variation is *expected*, individuals do not need to receive compensation for it (in the form of a “risk premium”) because the variation does not make them worse off. In a predictable market where participants observe the autoregressive and mean reverting behavior, appreciation should be the underlying trend in the series and true “risk” is the component of price variation that is *unexpected*. If there were little *expected* variation in housing prices (which seems unlikely

here given Figures 2-5), then the use of either measure would lead to similar results and conclusions.

The model we use is adapted from Capozza, Hendershott, and Mack [2004]<sup>1</sup>, and has three parameters of interest: one representing the autocorrelation of housing prices over time, one representing the degree of mean reversion in housing prices, and the last being the structural (or equilibrium) trend around which housing prices oscillate. The root MSE of the model will be of interest as a measure of the degree of unexplainable variation in housing prices and of course the trend will represent underlying return.

In the CHM model, the change in housing prices from today to tomorrow is a function of the change in housing prices from yesterday to today, and the deviation of housing prices from an equilibrium price level,  $P^*$ , which can change over time, and is estimated in a first-stage regression of price on various MSA-level economic and demographic characteristics. Due to the lack of time series economic data at the ZIP code level, and in name of simplicity, we eliminate the first stage, and assume that each ZIP's prices deviate around a constant trend. This gives the following specification:

$$\ln P_{t+2} - \ln P_t = \beta_1 (\ln P_{t+1} - \ln P_{t-1}) + \beta_2 (\ln P^* - \ln P_t) + \varepsilon \quad (8)$$

$$\text{where: } P^* = e^{a+bTrend}$$

In our model, housing price changes are measured over an interval of one year; however, the housing price index itself is measured every six months. So we end up with an overlapping of intervals. With this, the above equation translates to:

$$\ln P_{t+2} - \ln P_t = \gamma_0 + \gamma_1 (\ln P_{t+1} - \ln P_{t-1}) + \gamma_2 \ln P_t + \gamma_3 Trend + \varepsilon, \quad (9)$$

$$\text{where: } \gamma_0 = \beta_2 a, \gamma_2 = -\beta_2, \text{ and } \gamma_3 = \beta_2 b.$$

The parameter  $\gamma_1$  (or equivalently,  $-\beta_1$ ) measures autocorrelation in the housing series and  $\gamma_2$ , mean reversion. Estimation of the latter will allow the backing out of  $a$  and  $b$ , where  $b$  is the underlying ZIP trend.

In our time series model, the parameter space allows for four possible outcomes with regard to housing market cyclic behavior. With mean reversion ( $\beta_2 < 0$ ) and

autocorrelation ( $\beta_1$ ) greater than one, prices diverge in an oscillating fashion. With mean reversion, and autocorrelation less than one, prices converge with oscillations. With no mean reversion ( $\beta_2 \geq 0$ ) and autocorrelation greater than one, prices diverge with no oscillations. Finally, with no mean reversion and autocorrelation less than one (or with mean reversion and very small values of  $\beta_1$ ), prices converge with no oscillations.<sup>2</sup>

Given the large number of ZIP codes for which we estimate our model, we again present our results in Appendix 2 with a series of graphs that represent the quantile distributions of our measures of interest. Statistical significance is not reported graphically, but will be discussed in the text where noteworthy. Each graph includes the uniform distribution line as a point of reference. For convenience the distribution of parameters across the ZIPs for each market are summarized in Table 2 below, where we give the range of values within which 90% of the ZIPs fall.

**Table 2: distribution of ZIP Model parameters:**

<b><u>MSA</u></b>	<b><u>90%(<math>\gamma_1</math>)</u></b>	<b><u>90% (<math>-\gamma_2</math>)</u></b>	<b><u>90% Trend(b)</u></b>	<b><u>90% Error</u></b>
Boston	0.80 – 0.89	0.13 – 0.17	4.2 – 6.6	0.021 - 0.060
Chicago	0.25 - 0.87	0.12 - 0.35	4.5 – 8.5	0.013 - 0.033
Phoenix	0.30 - 0.70	0.10 – 0.30	5.5 – 10.0	0.010 - 0.029
S. Diego	0.76 - 0.88	0.11 - 0.15	4.5 – 6.5	0.032 - 0.060

For the Boston MSA, nearly 90% of the 110 ZIP codes have autocorrelation betas that are quite high - above 0.8. The values of the mean reversion beta are also quite uniform, and indicate convergent, oscillating housing price series for the entire MSA. All coefficients for autocorrelation and mean reversion are significant at the 5% level or better. The structural trend parameter is fairly uniform over the range of values, and is positive for all ZIPs. The fact that the trend is so small in comparison to the average Boston return is explained by the high autocorrelation coefficient. The Boston market gets its appreciation from great “momentum” following largely random shocks – resulting in a high root mean squared error. The fit of the time series model to the Boston data is still good, with a minimum R-squared of 0.68. Over 90% of the observations fall between 0.8 and the maximum of 0.96.<sup>3</sup>

For the Chicago MSA, returns exhibit far less autocorrelation, but still are fully within the convergent range. The parameters also are more broadly distributed across a wider range of values than Boston. Mean reversion is also more broadly distributed than Boston, and fewer than half of the observations are significant at the 10% level. Consistent with the graph in Figure 1, in which we observe no repeated boom-bust cycles, these results point to a pattern of convergence with little oscillations. This is also seen in the much larger trend coefficient for Chicago than Boston – despite a much lower overall average return. In Chicago appreciation comes from steady smooth growth rather than the lingering impact of shocks. This yields a lower root mean squared error. As might be expected from previously reported results, the fit of the model to the Chicago data is significantly worse than Boston. The distribution of the R-squared is fairly uniform across the range of values, 0.30 to 0.88.

The Phoenix MSA exhibits many of the traits of Chicago. The autocorrelations are similarly low, but still well in the convergent range and the trend distribution tends to be quite high. Thus Phoenix is a market in which ZIP level prices tend to smoothly grow. This pattern yields lower average root mean squared error. With regard to mean reversion, we observe a distribution that is somewhat similar to that of Chicago, but within a more compact range. The fit of the Phoenix sample is stronger than Chicago and just slightly worse than the Boston sample.

San Diego is a market that is similar in many respects to Boston. Autocorrelation is very high and quite uniform, while the trends are again quite small considering the strong appreciation that most areas in this market have experienced. Thus like Boston, San Diego has seen much of its growth come from the “long-lived” impacts of a few random shocks rather than a steady trend and this gives a high root mean squared error. Mean reversion in the San Diego MSA is very similar to Boston in terms of magnitude, dispersion, and statistical significance. The distribution of fit is similarly shaped, and almost identical to that of Boston.

Given that housing price behavior is ultimately determined by the *combination* of autocorrelation and mean reversion, we must look at the relationship between the two to complete our analysis – which we do with the figures in Appendix 3. Once again, we observe strong similarities within the Boston-San Diego and Chicago-Phoenix pairs. Across all four MSAs, there is a tendency toward lower levels of mean reversion ( $\beta_2$  or  $-\gamma_1$  closer to



zero) in ZIP codes with higher levels of autocorrelation. In Boston and San Diego, the observations are very tightly bunched, and the relationship is fairly subtle. For Chicago and Phoenix, the observations are highly dispersed and the relationship is extremely pronounced. In these two markets, we see quite a few ZIP codes that would fall into the parameter region where there is convergence with no oscillations, while we see no such observations in the other two MSAs.

In Section II we argued that a full equilibrium requires that the overall risk adjusted cost of owning a home compensate for the utility flow of services. Thus we would expect to find that “rent” (price level times opportunity cost of capital) minus expected appreciation plus the value of risk must compensate for service flow. In this equilibrium, this is equivalent to having risk equal the service flow minus rent plus expected appreciation. Thus holding service flow and prices fixed (if that is possible), we expect to find a positive (partial) relationship between risk and appreciation. and in Appendix 4 we examine if there are in fact such a positive relationships across the ZIP codes of our 4 MSAs. We do this with a series of scatterplots and accompanying regressions

The first panel in Appendix 4 shows the relationship between “raw” appreciation and risk – the average, one-year log difference in the price index and the standard deviation of that difference over the entire time-series interval. In the bottom panel, we substitute estimated appreciation and risk (the equilibrium slope coefficient and root MSE of the time series model, respectively) for the raw measure.

The strongest positive relationship between risk and appreciation can be found with the raw measures in the Chicago data, where we observe a correlation of almost 0.80. Chicago is followed by Phoenix at 0.30 and San Diego at 0.27. The worst of the four is Boston, in which the data exhibit a wide range of returns within a very narrow band of risk. Looking at our alternative measures of risk and return, we generally observe a lower degree of correlation than with the raw measure. Only Boston shows an improvement using the estimated values. Thus with respect to these historic performance, it does seem that ZIPs behave somewhat as an inter-temporal spatial equilibrium would demand.

#### **IV: INCORPORATING HISTORIC INVESTMENT PERFORMANCE INTO HEDONIC EQUATIONS.**

The data employed to estimate the hedonic regressions came from a variety of sources. The transactions data and housing unit characteristics were obtained from two real estate information clearing houses: The Warren Group and Dataquick, Inc. The former provided data for the Boston area, while the latter provided data for the other three MSAs. Data were initially limited to owner-occupied, single-family detached units. Further filtering led to the discarding of units with essential data missing, or values believed to be data reporting or recording errors. Homes with sale prices below \$20,000 (Boston) and \$10,000 (Chicago, Phoenix, and San Diego) were also discarded as possible indications of non-arms-length sales, possible data reporting/recording errors, or otherwise being non-representative of “normal” housing prices in our MSAs. All data were from 1998. For the Boston MSA, housing characteristics included the number of bedrooms, the number of bathrooms, interior square footage, lot size, and the year in which the house was built. Observations were further discarded if the number of bedrooms or bathrooms was less than one. The lot size was bounded between 0.02 and 10 acres. After filtering, we were left with 19,848 observations. For the Chicago MSA, housing characteristics were limited to number of bathrooms, interior square footage, and lot size. Observations were discarded if the number of bathrooms was less than one and no upper limit was placed on lot size. The final dataset contained 12,799 observations. Transaction data for the Phoenix MSA included the number of bathrooms, the number of *total* rooms, interior square footage, lot size, year built, and whether the house had a garage or pool. In this dataset, houses with “three-quarter” baths (containing a toilet and shower) were grouped together with those containing the next highest whole number of bathrooms. As with the Chicago MSA, no upper bound was placed on the lot size. The transactions data for San Diego contained the number of bedrooms, number of bathrooms, interior square footage, and whether the house had a garage or pool. Lot size was missing from over 40% of the observations, and therefore was ultimately omitted from the regression analysis. The final dataset contained 36,347 observations. In addition to those filters listed for Boston, square footage was bounded from below at a value of 300. The final dataset contained 15,068 observations.

Among the location attributes used in our hedonic specification, population density, median income, rate of homeownership and percentage nonwhite were obtained from the 2000 Decennial Census gazetteer and summary files. With the exception of median income, the location attributes were calculated using various other data from the census files, i.e. total population, total land area, total housing units, number of owner-occupied units, and total nonwhite population. The distance from the city center was generated using Mapquest internet mapping software.<sup>4</sup>

For the Boston area only, we also included a set of variables measuring educational quality and crime, two location factors which one would expect to be strongly capitalized into housing values. Educational quality is measured by 1998 Massachusetts Comprehensive Assessment System (MCAS) combined scores.<sup>5</sup> The MCAS is a standardized test that is administered in a variety of grades as a means of measuring public school performance. All students in the participating grades must take the exam, regardless of disability status or level of English proficiency. Scores are reported for “all students” and “regular students” (non-disabled and English proficient). In our analysis, we include the district-level scores for regular students in Grade 10 only. MCAS data were obtained from the Massachusetts Department of Education website.<sup>6</sup>

Data on crime were broken down into two categories, property crime and violent crime. Data for the city of Boston for 1998, broken down by neighborhood, was obtained from the Boston Police Office of Media Relations.<sup>7</sup> Data on all other towns came from the Massachusetts Crime Reporting Unit website.<sup>8</sup> Crime rates are presented as a per capita measure. Crimes incurred on college campuses are recorded separately from the towns in which they are located, and are not included in our measure. Data on crime and educational quality were not included in the analysis of the other three MSAs.<sup>9</sup>

The final variables included in the hedonic model measure risk and appreciation. In our analysis, we use two different sets of measures, each based off of the CSW/FISERV dataset described earlier. In the first instance, we use raw measures of risk and average historic appreciation – the log difference in the housing price index over a one-year interval, averaged (by ZIP code) over the entire duration of the time series, and the standard deviation of that average. In the second instance, return is measured as the slope coefficient on the structural trend from our time series model, while “risk” is the root mean squared error of

the model. To carry out our analysis, we run six different hedonic specifications for each of the four MSAs: 1) prices against housing characteristics alone, 2) prices against housing characteristics with the addition of price appreciation and risk, 3) prices against housing and trend and root mean squared error, 4) prices against housing characteristics and location variables, 5) prices against housing characteristics, location variables with price appreciation and risk, and 6), prices against housing characteristics, location variables and trend and root mean squared error. All equations regress the log of prices against a linear list of right hand side variables (Cropper [1988], Case et. al [1991])

The results of these hedonic equations are presented in full in Appendix 5. For each city there is a table of summary statistics, followed by a table of the combined regression results. Our primary concern is with the additional role that the various risk and return measures add to the equations so to that effect we present Tables 3a, and 3b. We then discuss the results for each city in turn.

**Table 3a: Historic Appreciation Coefficients<sup>1</sup>**

<b><u>MSA</u></b>	<b><u>Average, only Housing</u></b>	<b><u>Average plus Location</u></b>	<b><u>Trend, only Housing</u></b>	<b><u>Trend, plus Location</u></b>
Boston	25.4	11.6	33.2	14.9
Chicago	12.4	12.4	10.6	10.8
Phoenix	10.3	4.8	2.1	1.7
S. Diego	13.6	2.7	12.1	6.0

1. All coefficients significant at 5%.

**Table 3b: Historic Risk Coefficients<sup>1</sup>**

<b><u>MSA</u></b>	<b><u>Risk, only Housing</u></b>	<b><u>Risk, plus Location</u></b>	<b><u>RMSE, only Housing</u></b>	<b><u>RMSE, plus Location</u></b>
Boston	-22.9	-9.9	-12.1	-3.2
Chicago	-12.2	-8.7	-2.1	-6.5
Phoenix	-3.0	4.0	-5.7	3.6
S. Diego	3.6	10.2	8.4	8.5

1. All coefficients significant at 5%.

**Boston.** The basic equation with only housing characteristics has an  $R^2$  of 0.53. All coefficients except bedrooms are significant. This result is often found when total interior square feet is controlled for. Homes with many small rooms are not as valued. When the investment variables are added, the  $R^2$  jumps quite dramatically to 0.66 in the case where raw risk and appreciation are used or 0.64 when trend and root mean squared error are the metrics. Both variables are very significant, and have the correct sign.

When the location variables are added into the hedonic equation, the  $R^2$  rises to 0.70. It is important to note that the equation with location variables and no investment performance metrics has about the same  $R^2$  as the equation with just the investment metrics. In other words there is a high degree of collinearity between the investment metrics and the location variables.

**Chicago.** The basic equation for Chicago, containing only housing characteristics has a somewhat lower  $R^2$  than Boston at 0.44. Unlike Boston, adding in our investment metrics increases the explanatory power of the equation only modestly, to an  $R^2$  of around 0.47. The appreciation variable has the hypothesized positive sign, and is smaller in magnitude to Boston when measured either as return or trend.

When the location variables are added the investment metrics the  $R^2$  increases from 0.47 to 0.61 (Appendix Table 5.3b columns 2 and 5). The appreciation and trend metrics now are much closer in point estimate to the Boston results, and the risk metrics remain significantly negative in sign.

**Phoenix.** Phoenix, with a richer array of structural variables, has a “base”  $R^2$  of 0.68. When we add in raw risk and average appreciation the  $R^2$  increases only to 0.69, and barely at all with trend and root mean squared error. However, both raw appreciation and risk are correctly signed and significant statistically. The magnitude of the coefficients however is much smaller than Boston and Chicago.

Once location variables are added to the equation the  $R^2$  increases only modestly, from 0.68 to 0.73 (Appendix Table 5.4b columns 2 and 5). When in turn the investment metrics are added  $R^2$  barely increases at all, to just under 0.74. With the location variables included both risk measures become positive in sign as well as being significant statistically.

**San Diego.** The San Diego results are similar to those in Boston. San Diego has a “base”  $R^2$  of 0.56. When the investment metrics are added the  $R^2$  increases to 0.59. Both

average appreciation and trend have significant positive coefficients – between the results for Boston and Phoenix. Risk, however, has a positive (and significant) coefficient throughout all of the results.

With the addition of the location variables the  $R^2$  increases to 0.68 (Table 5.5b columns 2 and 5). The appreciation and trend variables continue to have positive coefficients, but reduced in magnitude while both raw risk and root mean squared error have larger positive coefficients.

In summary, the hedonic equations produce quite reasonable results without the investment variables – results that seem typical with other studies [ ]. The inclusion of historic appreciation, measured either raw or with the trend variable is always highly significant and with a large positive coefficient. The risk metric, however, has very mixed results. Only in Boston and Chicago does this variable have the expected significant and negative impact (measured either raw or as root mean squared error). In many of the other markets, its impact is positive on price and in many cases this effect is significant. At this time we have no explanation for the absence of risk-pricing in the markets outside of Boston and Chicago.

The magnitude of the coefficients also should be judged against the theory of Section II. The point estimates in Table 3a suggest in exchange for a 1% increase in annual appreciation (for 25 years – or effectively forever) owners are willing to pay between 10% and 20% more for a unit (with the same flow of utility-based services). If there are no liquidity constraints we can think of this 1% yearly increase in appreciation as an income flow which must be discounted. In perpetuity we should be willing to pay  $1/\text{discount rate}$  for that. By this reasoning a coefficient of 10 to 20 is just in the correct ballpark.

We can also examine equation (4) in more detail. There if we take the percentage derivative of  $p$  with respect to a unit change in  $\mu_H$  we get  $1/(\mu_H - R + \frac{\alpha}{a})$ . In real terms housing appreciates a bit more than the real interest rate  $R$ , but this expression actually is dominated by the CARA coefficient ratio  $\frac{\alpha}{a}$ . This ratio is the same as the ratio of housing to “other” expenditure and might have a value on average of say .2. By this formulation we would get a coefficient in a log price regression that is smaller – around 5.0.

When we turn to the risk coefficients, things obviously become much more complicated. When we examine equation (4) we would get a percentage derivative of price with respect to the “value” of risk (the product:  $a\sigma_H^2$ ) that is minus 1 over that value. From an investment perspective, in liquid financial markets, the “value” of the risk in housing should be the difference between the total return to housing and the risk free return (R). With 3% real appreciation and say a 6% rent payment and 2% real R, we would say that the product should be between 5% and 10%. One over that would give a regression coefficient of between -10 and -20. Interestingly that is almost exactly the case in Boston, but not in the other cities, where the coefficient most often has the wrong sign. Further investigation of the data may be needed to resolve this issue.

A second objective of the paper is to ascertain if the inclusion of the investment performance metrics changes any other coefficients that are included in such equations – in particular for variables that represent location attributes. Here the most pronounced results occur in Boston where it was possible to collect some data on public services that overlapped nicely with ZIP codes. The inclusion of the investment metrics had a very inconsistent impact on the importance of crime and no impact on the valuation of school quality. For the other MSA and attributes we turn to Table 4.

**Table 4: Impact of Historic Return/Risk on Location Coefficients <sup>1</sup>**

<u>Variable</u>	<u>Boston</u> <u>WO / W</u>	<u>Chicago</u> <u>WO / W</u>	<u>Phoenix</u> <u>WO / W</u>	<u>San Diego</u> <u>WO / W</u>
Non White	-0.40 / -0.20	-0.52 / -0.37	-0.73 / -0.67	-0.59 / -0.60
Distance	-1.6 / -0.61	0.24* / 1.2	-2.2 / 0.24*	-1.0 / -0.61
Distance <sup>2</sup>	1.1 / 0.24	-3.6 / -4.8	4.3 / -0.04*	1.4 / 0.40
Own Rate	-0.83 / -0.51	-1.2 / -1.0	-0.73 / -0.67	-1.1 / -0.81
Ln med. inc.	0.81 / 0.58	1.2 / 1.1	0.43 / 0.41	0.81 / 0.66

1. All coefficients significant at 5% *except* those with an \*.

In Boston, Chicago, and Phoenix the inclusion of the investment metrics modestly reduces the negative impact of a ZIP’s racial makeup. Likewise, in all four MSAs, it reduces the negative impact of the ownership rate. Most significant is the impact of the investment variables on distance to the city center. This variable is reduced and/or reverses

sign in 3 of the 4 MSAs. Finally, the inclusion of the investment variables reduces the impact of median household income on prices across all four MSAs. These changes suggest that estimates of the pure utility value of commuting, racial exclusion, neighborhood ownership, and neighborhood income may have been over estimated. This raises some doubt about why people pay so much for housing in “better” areas. Is it the actual flow of services or the perception that purchasing in these areas is a better investment which drives prices?

## **V: EX ANTE INVESTMENT PERFORMANCE AND HOUSE PRICES**

There is a significant problem with using historic appreciation in a housing price equation in which price is measured near the end of the period over which appreciation is calculated. By construction, ZIP areas that differ in prices randomly at the beginning of the period will clearly have a positive correlation with intervening appreciation by construction. In fact the only justification for using historic appreciation is that it is a good predictor of future appreciation which is after all what informed consumer/investors care about. While Case and Shiller show this historic link exists at various shorter term frequencies in the housing market it is not clear that it holds over decades.

A common test in Finance is to judge if asset prices contain any information about the future (Campbell and Shiller, 1998). In the context of this paper, then, we might ask if price levels – controlling for attributes – are good predictors of actual or forward forecasts of price appreciation. Since attributes effectively determine housing “rent”, this is equivalent to asking if price/rent ratios have any predictive power. With our data sample, then we can test if prices in 1998 forecast higher appreciation for the next 5 years. To do this we use two measures of subsequent price appreciation. The first is the actual average yearly return from 1998 through 2003, and the second is the average yearly return from 1998 through 2003 forecasted using the coefficients generated from our time-series model (described in Section III, but estimated only with data through 1998). The results of including these measures of appreciation are shown in Tables 5a and 5b below. In both of these tables we also include the results for risk – measuring risk the only way we can – that is historically. (Full hedonic results can be found in Appendix 5, Tables 5.Xc)



**Table 5a: Actual Future Appreciation (1998-2003) and Historic Risk Coefficients<sup>1</sup>**

<u>MSA</u>	<u>Actual, only Housing</u>	<u>Actual, plus Location</u>	<u>Risk, only Housing</u>	<u>Risk, plus Location</u>
Boston	12.7	-0.23*	-32.6	-9.3
Chicago	9.1	6.3	-6.6	2.3
Phoenix	0.75	-0.43	1.2	5.0
S. Diego	-5.5	-3.4	13.5	12.6

1. All coefficients significant at 5% *except* those with an \*.

**Table 5b: Forecast Future Appreciation (1998-2003) and Historic Risk Coefficients<sup>1</sup>**

<u>MSA</u>	<u>Forecast, only Housing</u>	<u>Forecast, plus Location</u>	<u>Risk, only Housing</u>	<u>Risk, plus Location</u>
Boston	13.0	3.8	-8.1	-5.7
Chicago	-15.4	-1.3*	20.8	1.2
Phoenix	3.3	1.6	1.3	5.8
S. Diego	-3.9	-1.1	9.5	11.7

1. All coefficients significant at 5% *except* those with an \*.

The results in these tables are quite disappointing. A quick look back at Figures 2 through 5 shows that the period from 1998-2003 saw both a great deal of housing price appreciation, and if anything a widening gap between the appreciation of individual ZIP areas. Despite this the predictive power of appreciation is very poor. In the first two columns of Tables 5a and 5b, 4 out of 8 appreciation coefficients are either insignificant or of the incorrect sign! It is very questionable if future appreciation (either actual or forecast) was anticipated by price (to rent) levels in 1998. This is very much worse than the results we obtained using historic appreciation in Table 3a, and suggests that our concern over misspecification was correct.

In the last two columns in each table, we see that the risk variable is again of the wrong sign in the majority of cases – and with larger coefficients. These results also are worse than those in Table 3b. These results cast considerable doubt over the previous results using historic appreciation. Equally important they imply that in a panel within metropolitan

areas it is not necessarily true that higher prices (controlling for attributes or “rent”) are indicative of enhanced *future* appreciation and lower risk. The market is not efficient.

In terms of further research, there are clear priorities. This starts with adding on the last two years worth of housing appreciation (2003-2005) into the sample so that we can use 7 rather than 5 years of subsequent appreciation in the test of the forward predictive power of prices. We also need to try and unravel the puzzle around consistent positive signs for the risk variables. Finally, we need to expand the number of location variables to possibly include environmental ones, and to obtain the crime and school data for the remaining MSAs. Quite possibly the absence of these important variables is altering our results. For the moment, however, we conclude that while the housing market is quite predictable, across locations, this predictability is not efficiently priced into current price levels.

### **Bibliography**

- Alonso, W. Location and Land Use, (Cambridge, MA: Harvard University Press, 1964).
- Bartik, Timothy. “The Estimation of Demand Parameters in Hedonic Price Models, *Journal of Political Economy*, 95 (1987) 172-183.
- Brown, James, and Harvey Rosen. “On the Estimation of Structural Hedonic Price Models”, *Econometrica*, 50 (1982) 765-768.
- Campbell, John, and Shiller, R.J. "Valuation Ratios and the Long Run Stock Market Outlook" *Journal of Portfolio Management*, (Winter, 1998) 11-26.
- Capozza, Dennis and Bob Helsley. "The Stochastic City", *Journal of Urban Economics*, 28 (1990) 187-203.
- Capozza, Dennis, Hendershott, Patric, and Charlotte Mack. “An Anatomy of Price Dynamics in Illiquid Markets: Analysis and Evidence from Local Housing Markets,” *Real Estate Economics*, 32 (2004) 1-21.
- Case, Bradford, Henry O. Pollakowski, and Susan M. Wachter. “On Choosing among Housing Price Index Methodologies,” *AREUEA Journal*, 19 (1991), 286-307.
- Case, K.E., and Shiller, R.J. "The Efficiency of the Market for Single Family Homes," *American Economic Review*, 79 (1989) 125-137.

Cropper, Maureen, Leland Deck and Kenneth McConnell. "On the Choice of Functional Form for Hedonic Price Functions." *Review of Economics and Statistics*. 70 (1988) 688-675.

DiPasquale, D. and Wheaton, W.C. Urban Economics and Real Estate Markets, Prentice Hall, New Jersey (1996).

Dougherty, A., and R. Van Order. "Inflation, Housing Costs and the Consumer Price Index", *American Economic Review*, 72 (1982).

Ekeland, Ivar, James Heckman, and Lars Nesheim. "Identification and Estimation of Hedonic Models", *Journal of Political Economy*, 112 (2004 forthcoming).

Epple, Dennis. "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products", *Journal of Political Economy*, 95 (1987) 59-80.

Grossman, Sanford, and Guy Laroque. "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods", *Econometrica*, 58 (1990) 25-51.

Hamilton, B. and R. Schwab, "Expected Price Appreciation in Urban Housing Markets," *Journal of Urban Economics*, 23 (1985).

Kearl, J.R. "Inflation, Mortgages and Housing", *Journal of Political Economy*, 87 (1979).

Palmquist, Raymond. "Estimating the Demand for the Characteristics of Housing", *Review of Economics and Statistics*, 66 (1984) 394-404.

Poterba, J. M. "Tax Subsidies to Owner-Occupied Housing: An Asset-Market Approach," *Quarterly Journal of Economics*, 99 (1984) 729-752.

Ricardo, D. On the Principles of Political Economy and Taxation, 1817.

Rosen, Sherwin. "Hedonic Prices and Implicit Markets; Product Differentiation in Pure Competition", *Journal of Political Economy*, 82 (1974) 35-55.

Schwab, R.M. "Inflationary Expectations and the Demand for Housing", *American Economic Review*, 72 (1982).

Seslen, Tracey N. "Housing Price Dynamics and Household Mobility Decisions," mimeo, USC Marshall School of Business (2004).

Smith, V. Kerry, and Ju-Chin Huang. "Can Markets Value Air Quality? A Meta-Analysis of Hedonic Property Value Models." *Journal of Political Economy*, 103 (1995) 209-227.

Wheaton, William. "Real Estate Cycles: Some Fundamentals." *Real Estate Economics*, 27 (1999) 122-222.

---

<sup>1</sup> Heretofore referred to as the CHM model

<sup>2</sup> See CHM [2004] Figure 1.

<sup>3</sup> In the interest of space, the graphs of the R-squared distributions have been omitted. They are available upon request.

<sup>4</sup> Distance was measured based on the most efficient driving route between the center of the ZIP code and the center of the city proper.

<sup>5</sup> This is the sum of scores from English/Language Arts, Mathematics, and Science and Technology

<sup>6</sup> <http://www.doe.mass.edu/mcas/results.html>

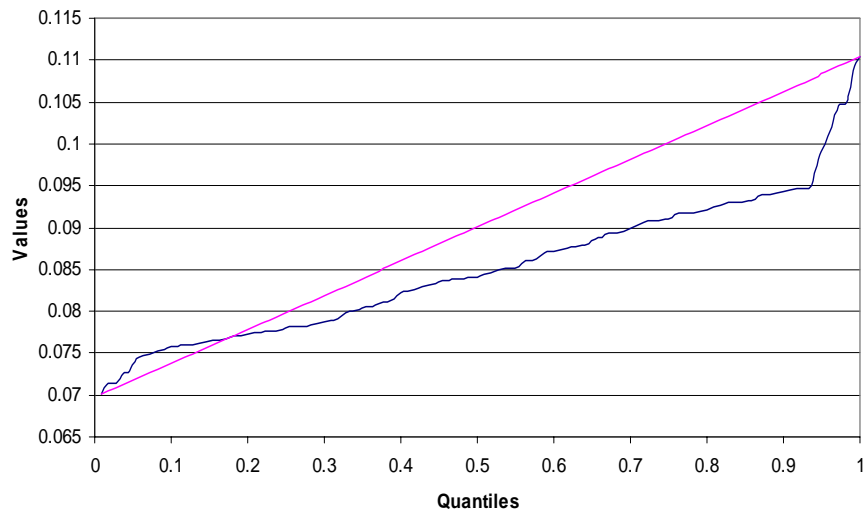
<sup>7</sup> The neighborhood data conform very closely, if not perfectly, to ZIP code boundaries. Assigning neighborhood-specific crime values to the various ZIP codes within the Boston city limits proved very important, due to the overall city size and strong variation in crime rates across locations

<sup>8</sup> <http://www.ucrstats.com/>

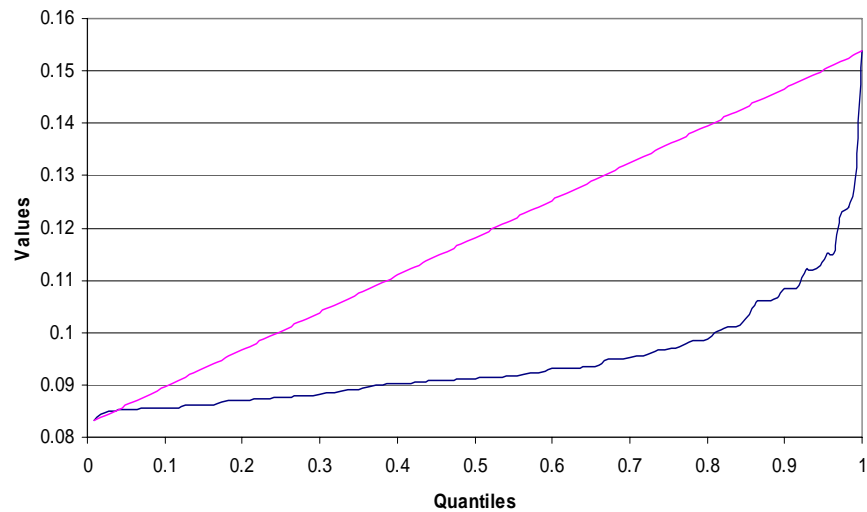
<sup>9</sup> Boston was a very convenient “test case” for the explanatory power and “proxy value” of education and crime statistics, since 1) every town has its own school district and police force, and 2) the ZIP codes contained within those towns *never* cross town borders. In the other three MSAs, ZIP codes often contain more than one school district or police jurisdiction, making it very difficult to pinpoint the district/jurisdiction governing the particular housing unit in the sample.

# APPENDIX 1

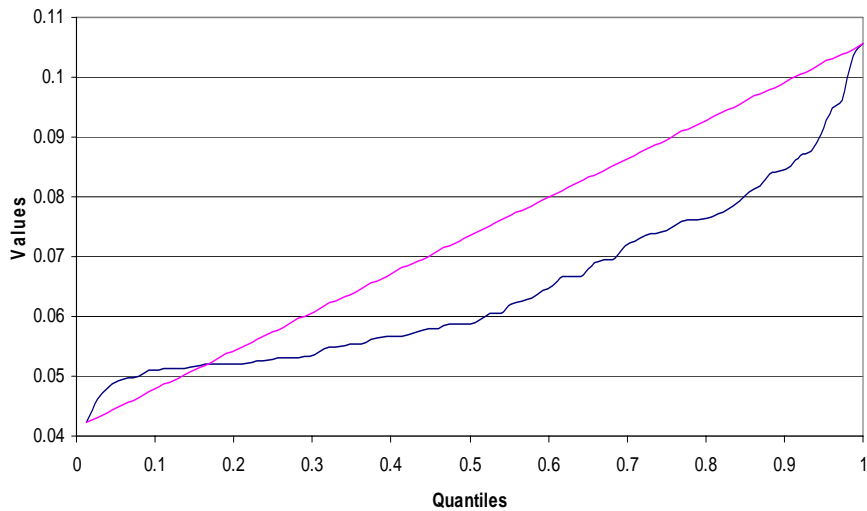
Boston sample: quantiles of raw 1-year return



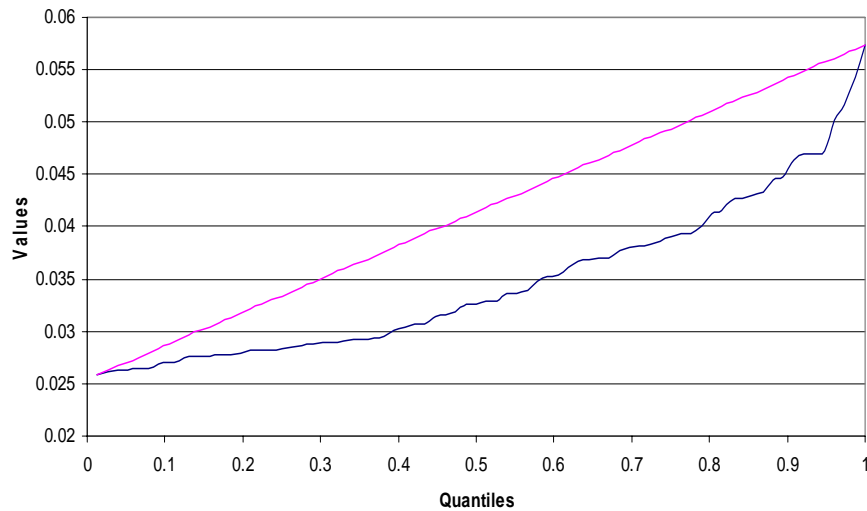
Boston sample: quantiles of raw 1-year risk



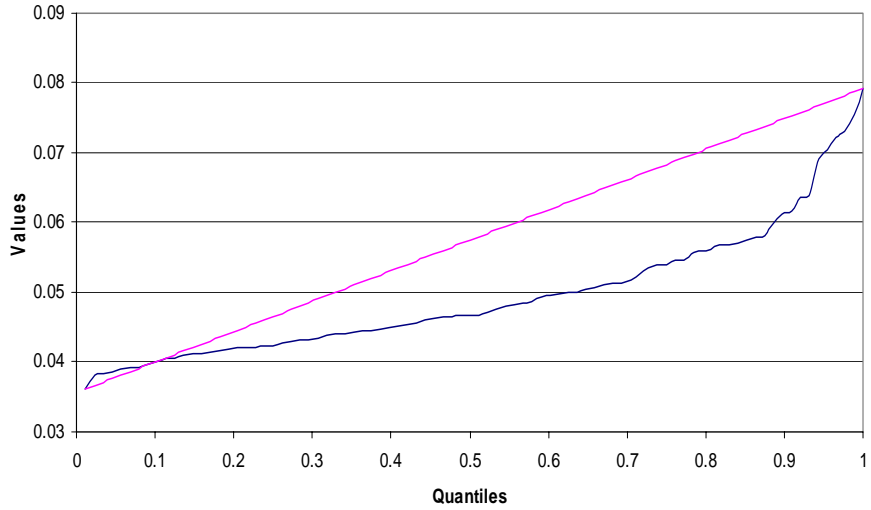
Chicago sample: quantiles of raw 1-year return



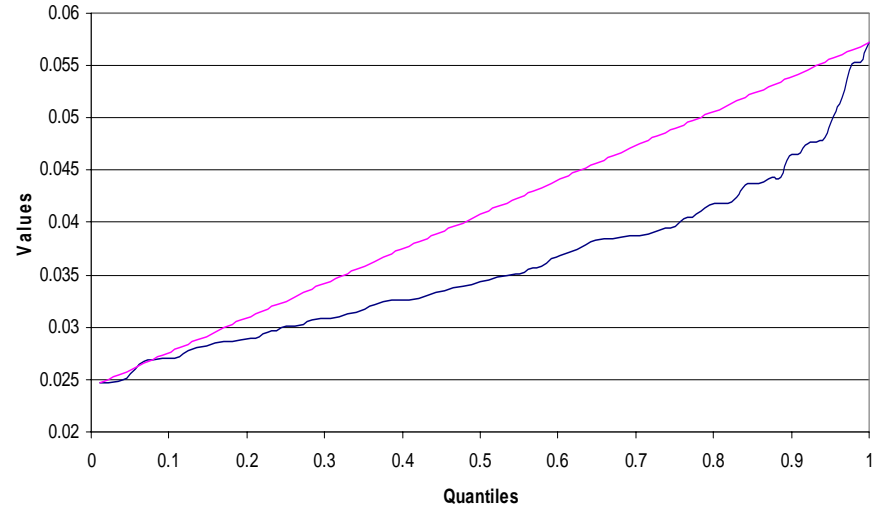
Chicago sample: quantiles of raw 1-year risk



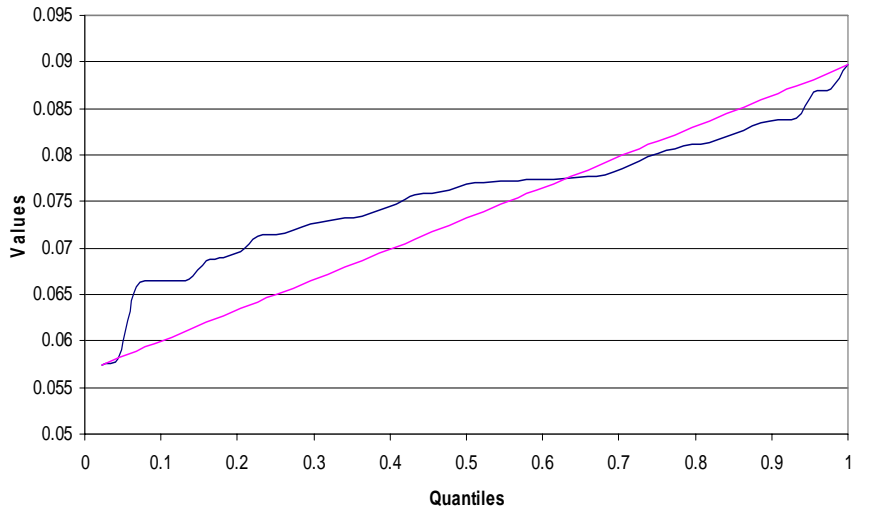
Phoenix sample: quantiles of raw 1-year return



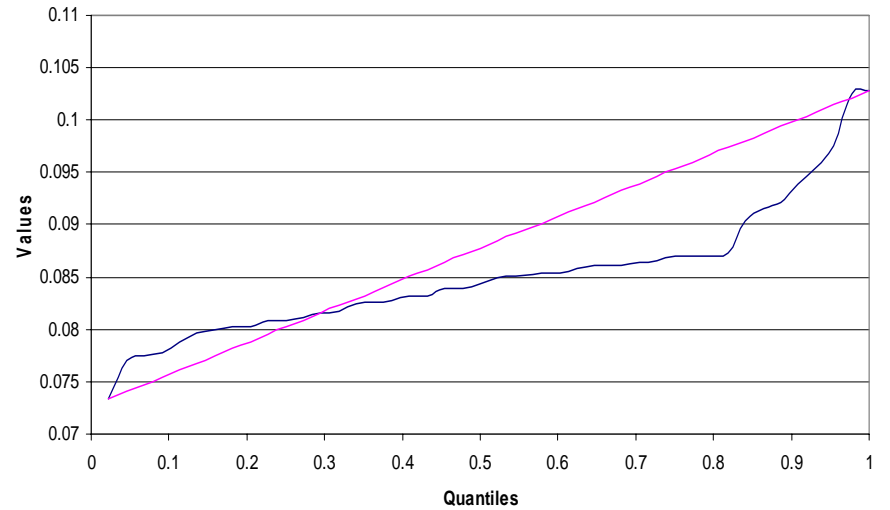
Phoenix sample: quantiles of raw 1-year risk



San Diego sample: quantiles of raw 1-year return

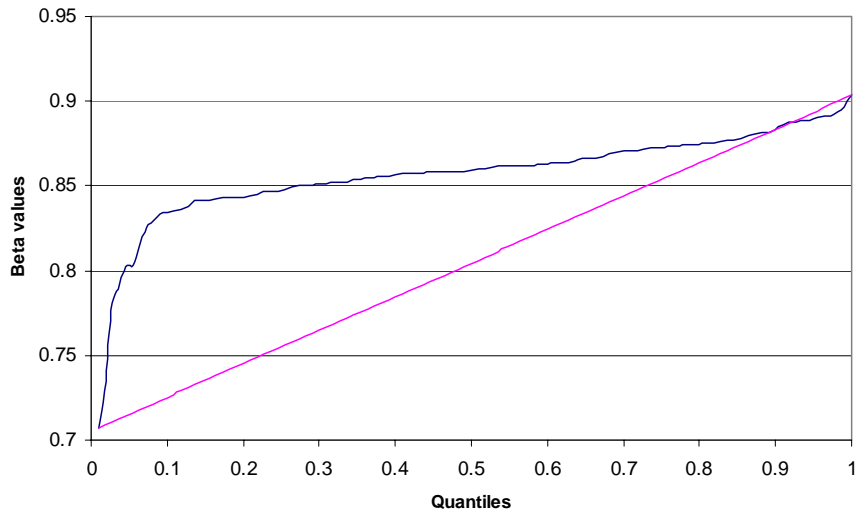


San Diego sample: quantiles of raw 1-year risk

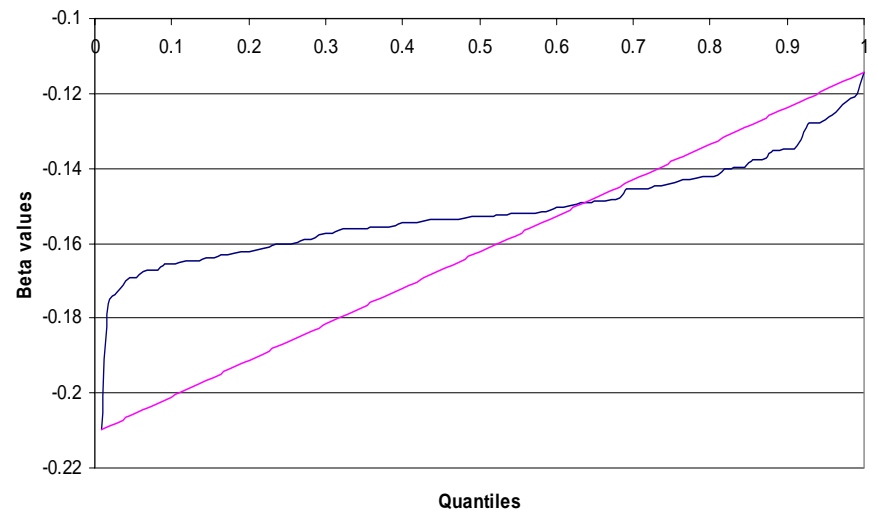


# APPENDIX 2

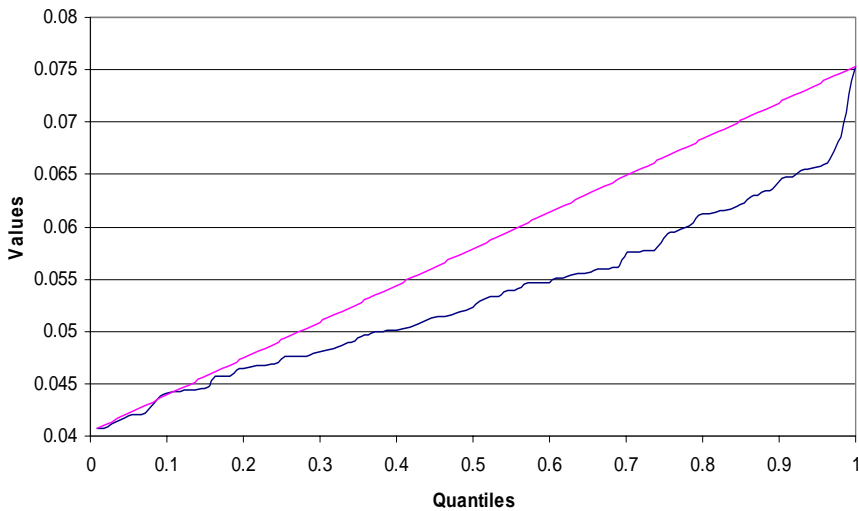
**Boston sample: quantiles of autocorrelation beta**



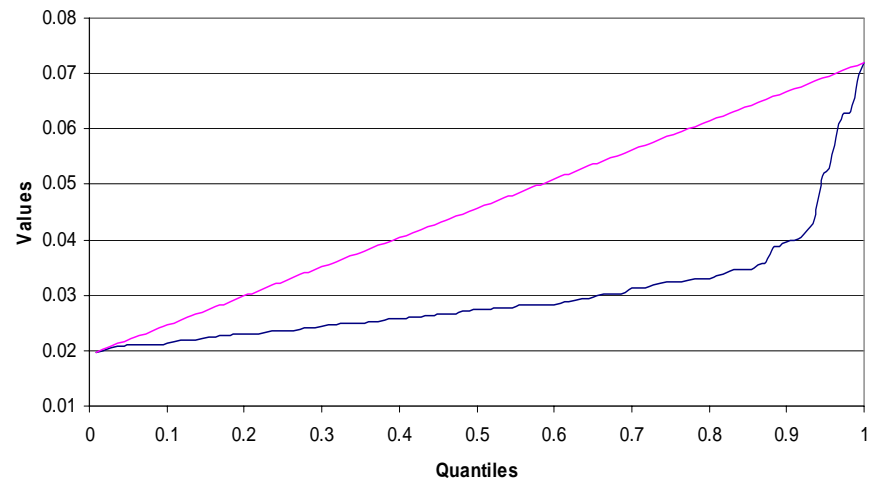
**Boston sample: quantiles of mean reversion beta**



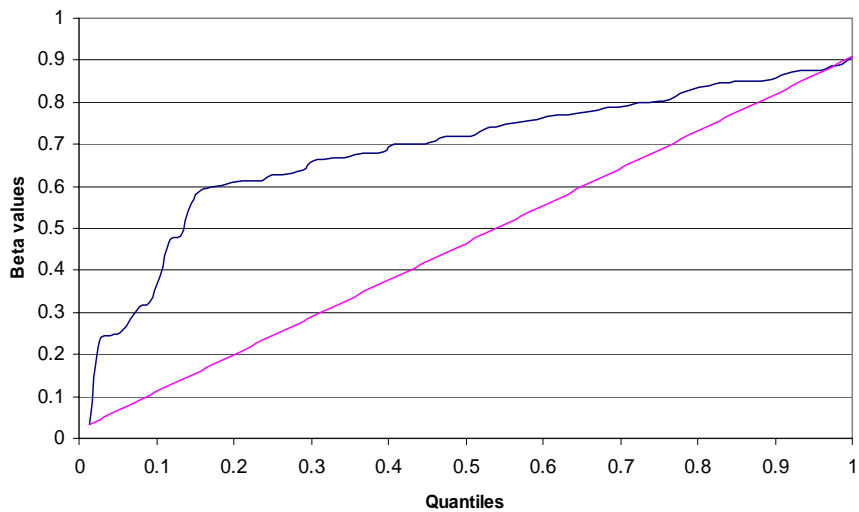
**Boston sample: quantiles of structural trend**



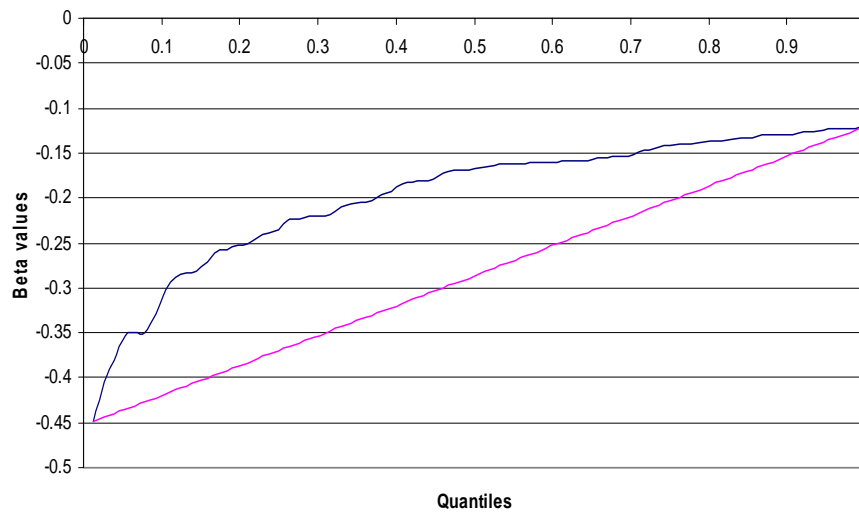
**Boston sample: quantiles of unanticipated volatility  
(root mean squared error of univariate model)**



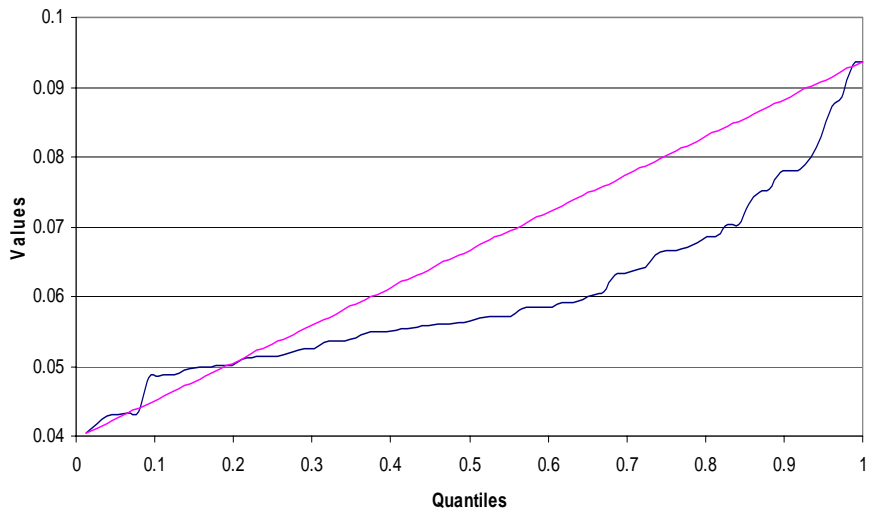
Chicago sample: quantiles of autocorrelation beta



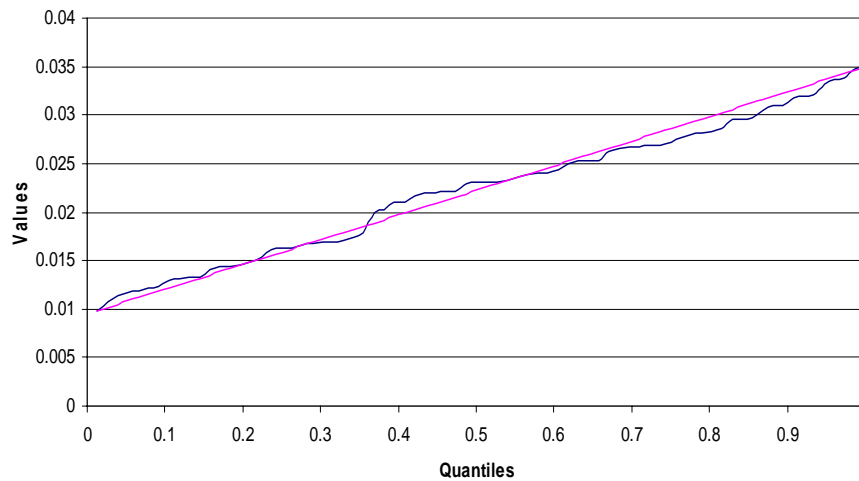
Chicago sample: quantiles of mean reversion beta



Chicago sample: quantiles of structural trend

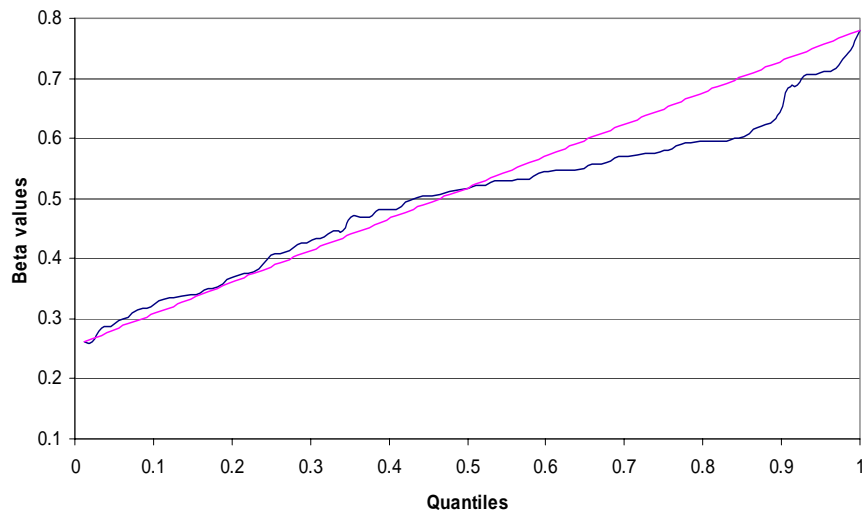


Chicago sample: quantiles of unanticipated volatility  
(root mean squared error of univariate model)

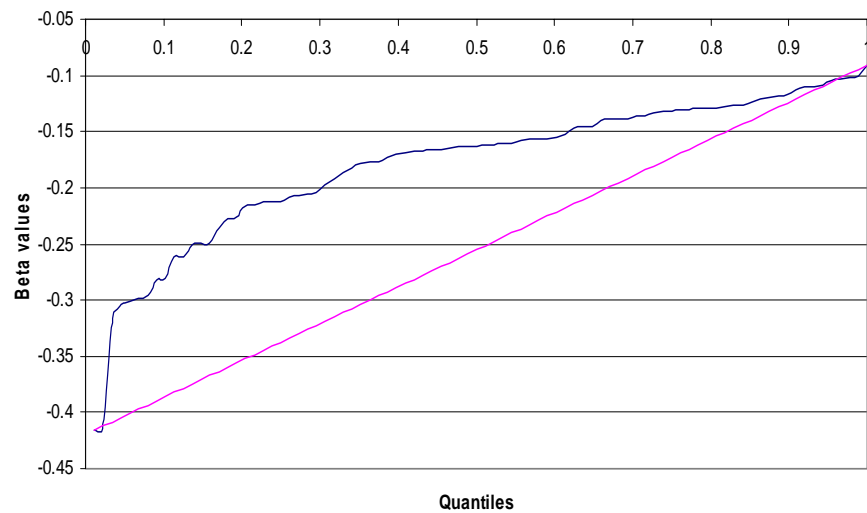




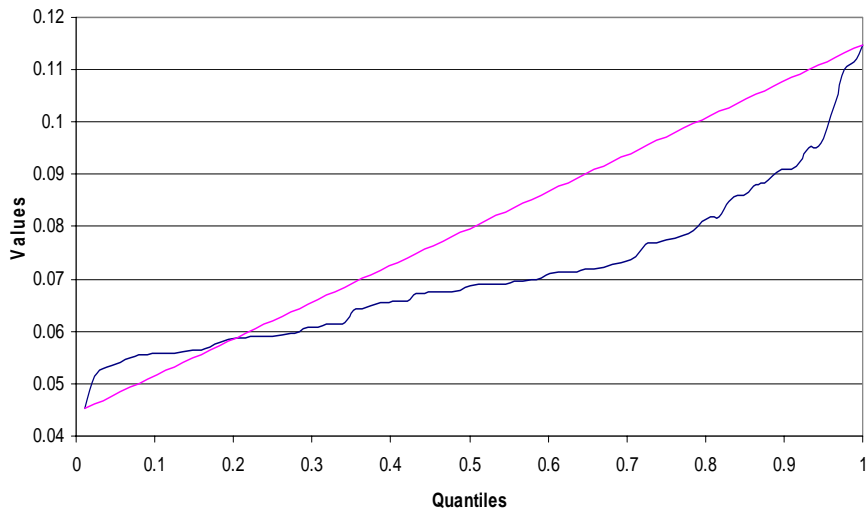
Phoenix sample: quantiles of autocorrelation beta



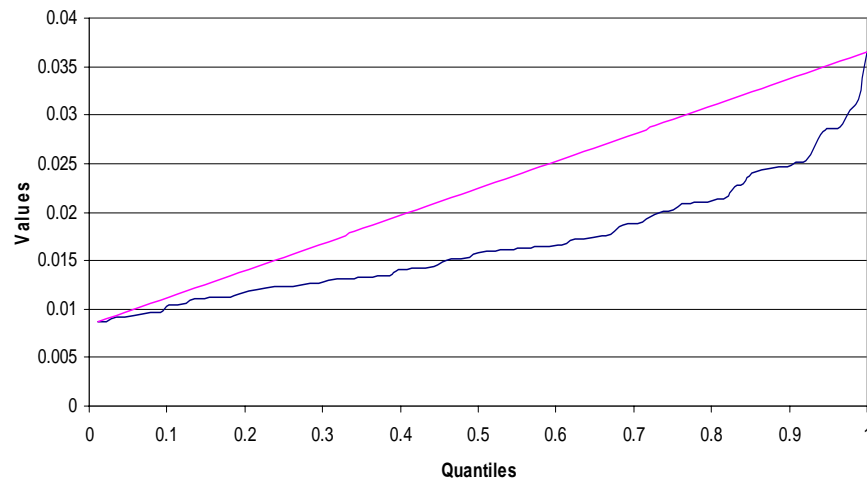
Phoenix sample: quantiles of mean reversion beta



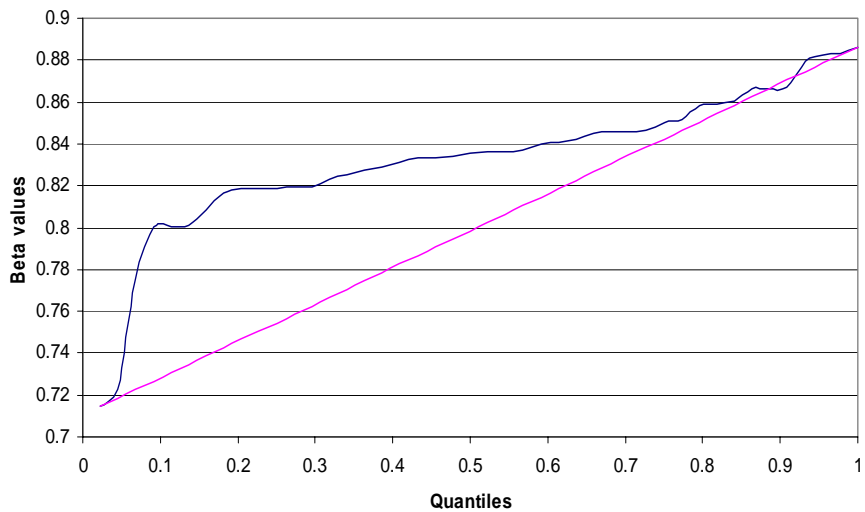
Phoenix sample: quantiles of structural trend



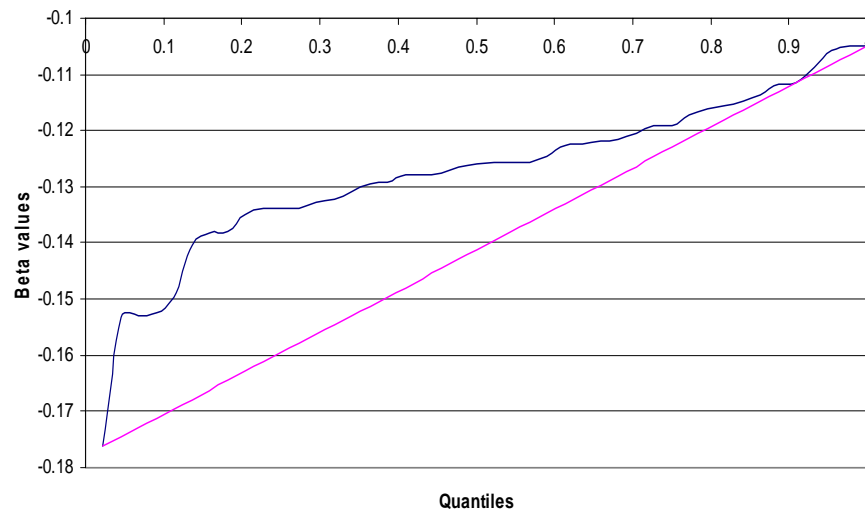
Phoenix sample: quantiles of unanticipated volatility  
(root mean squared error of univariate model)



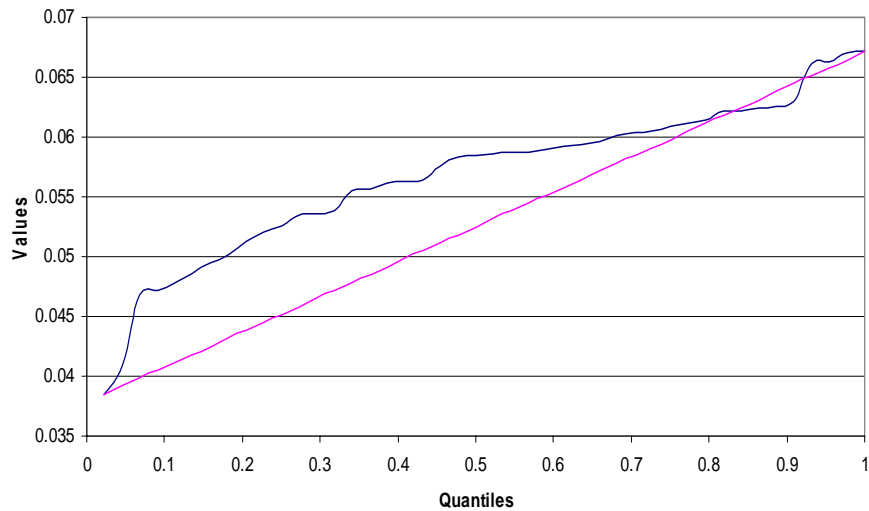
San Diego sample: quantiles of autocorrelation beta



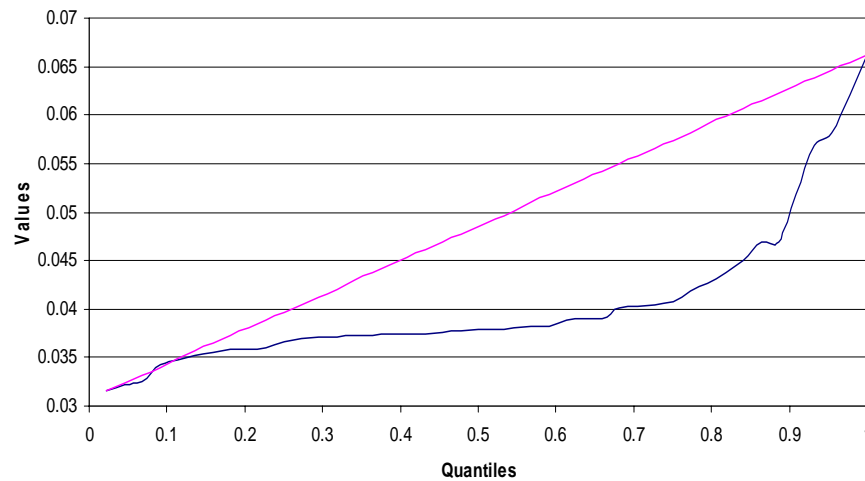
San Diego sample: quantiles of mean reversion beta



San Diego sample: quantiles of structural trend

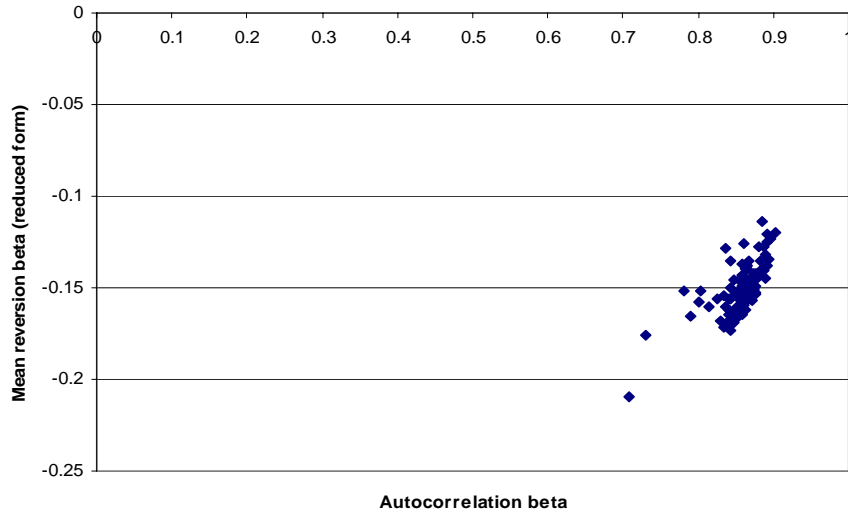


San Diego sample: quantiles of unanticipated volatility  
(root mean squared error of univariate model)

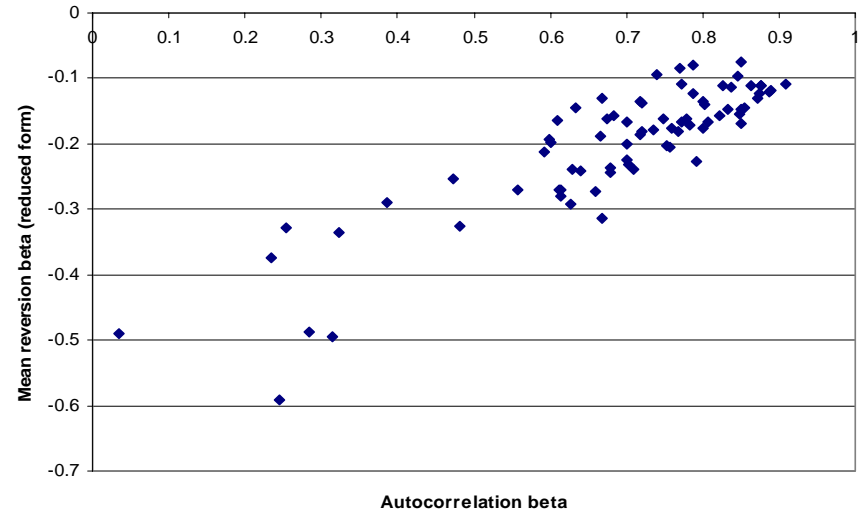


# APPENDIX 3

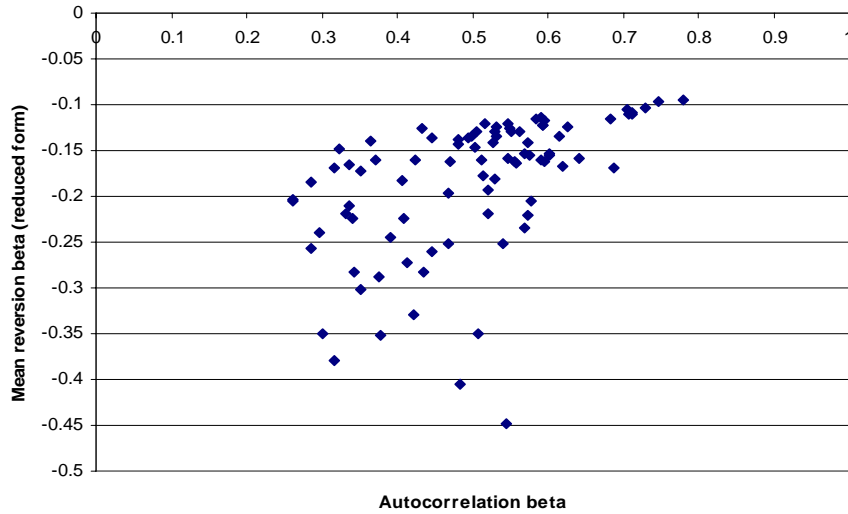
**Boston Sample: Autocorrelation vs. Mean Reversion**



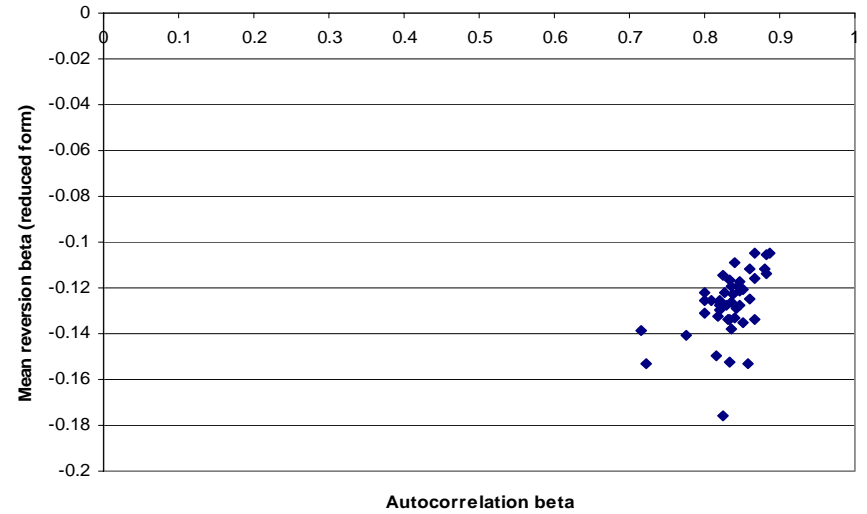
**Chicago sample: Autocorrelation vs. Mean Reversion**



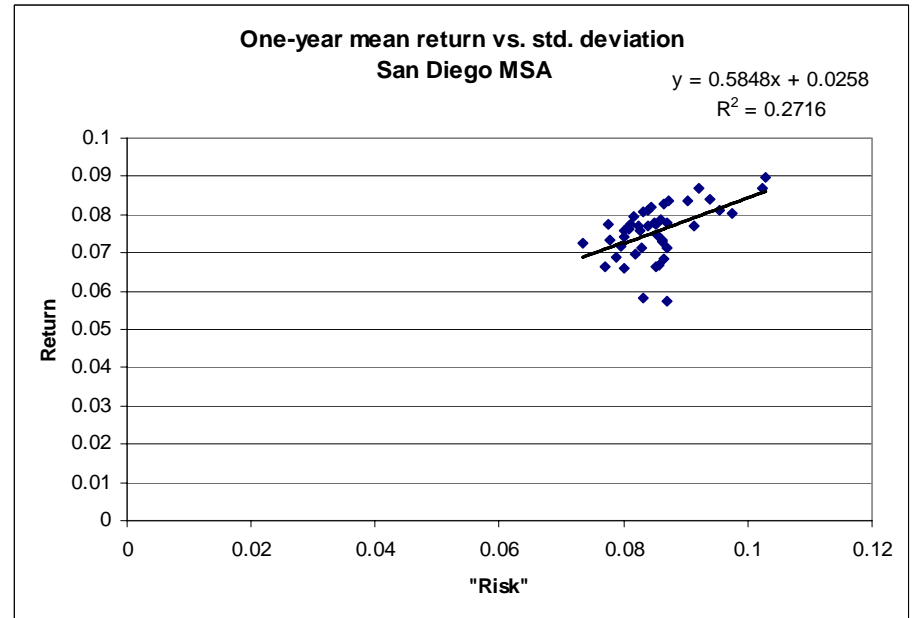
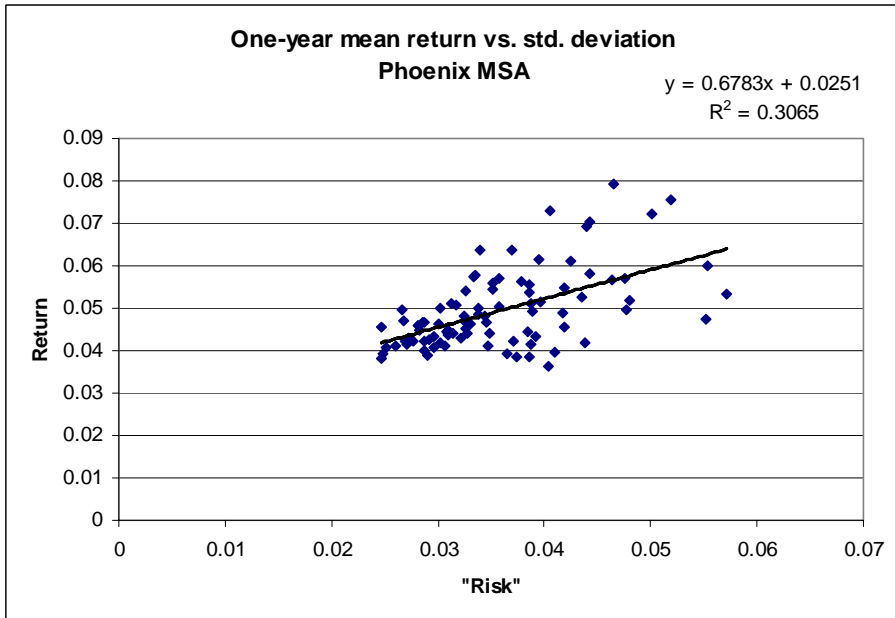
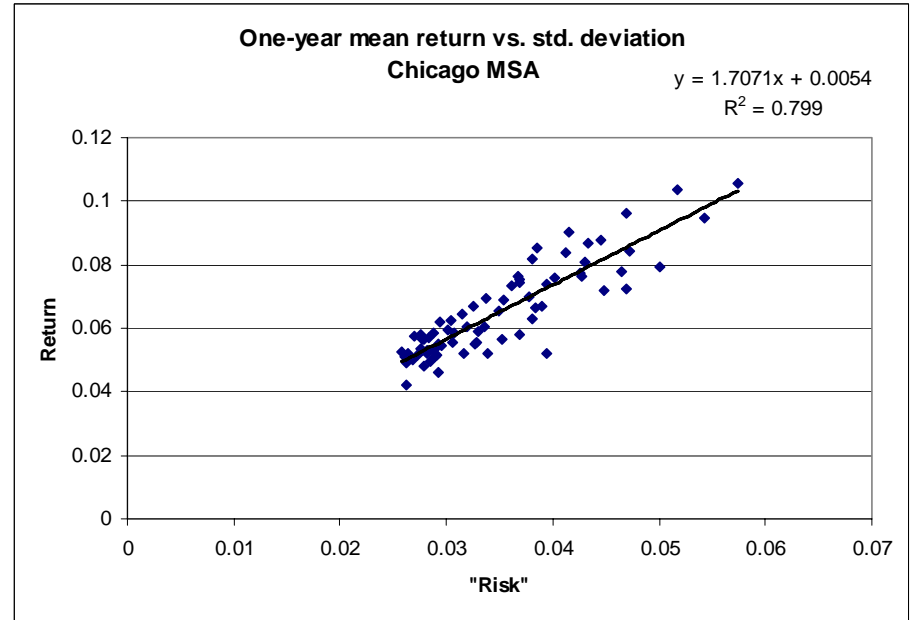
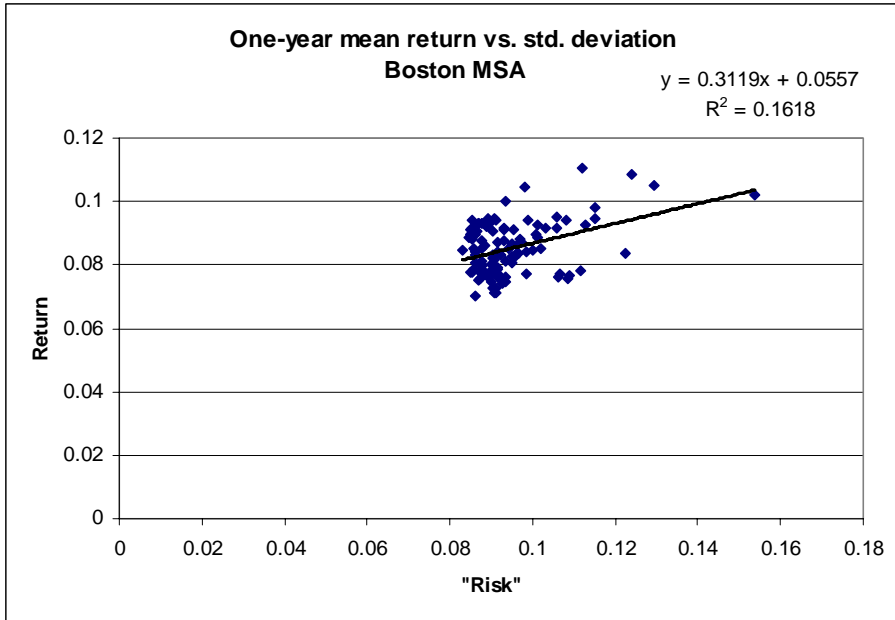
**Phoenix Sample: Autocorrelation vs. Mean Reversion**



**San Diego Sample: Autocorrelation vs. Mean Reversion**

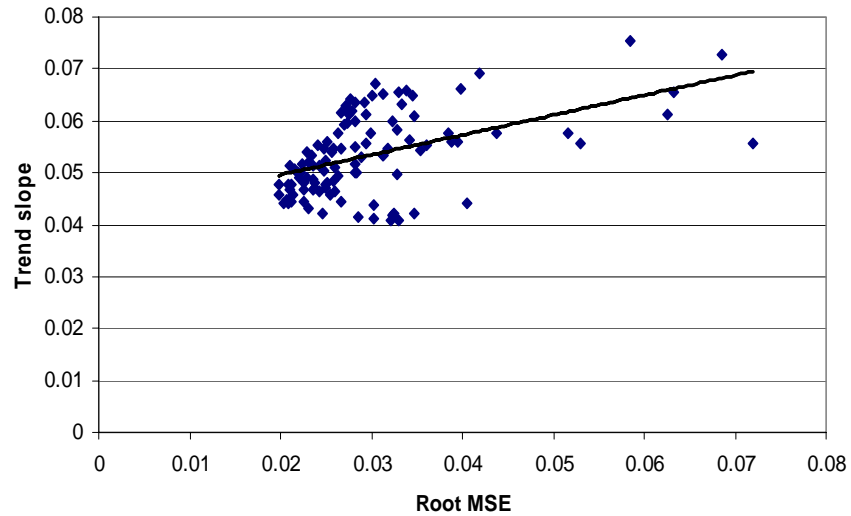


# APPENDIX 4



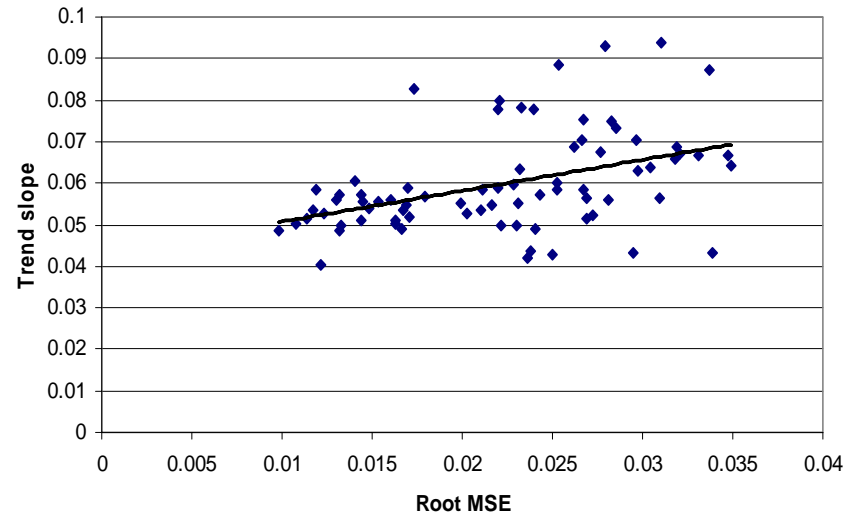
**Structural trend slope coefficient vs. root MSE,  
Boston MSA**

$$y = 0.3833x + 0.042$$
$$R^2 = 0.242$$



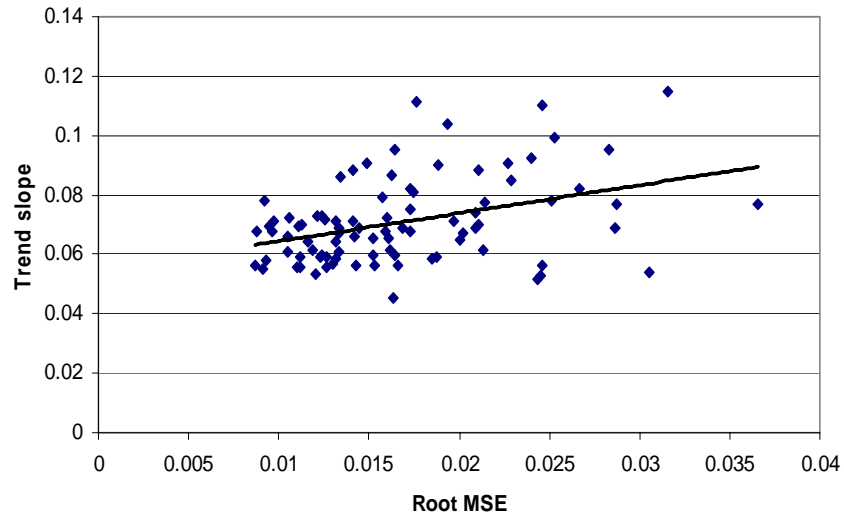
**Structural trend slope coefficient vs. root MSE  
Chicago MSA**

$$y = 0.7446x + 0.0433$$
$$R^2 = 0.1824$$



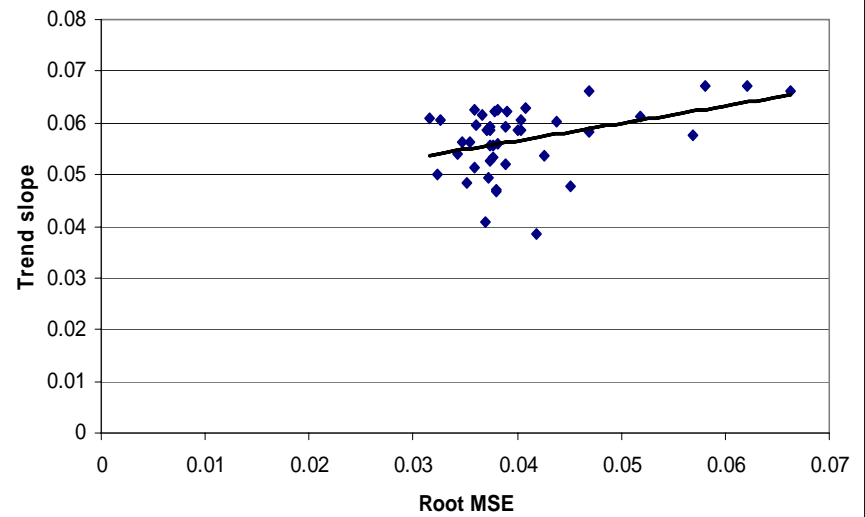
**Structural trend slope coefficient vs. root MSE  
Phoenix MSA**

$$y = 0.944x + 0.055$$
$$R^2 = 0.1509$$



**Structural trend slope coefficient vs. root MSE  
San Diego MSA**

$$y = 0.3357x + 0.0431$$
$$R^2 = 0.1559$$



**APPENDIX 5**  
**Table 5.1: Variable Definitions**

Boston:

lnsale = natural log of sale price

built60 = 1 if home was built between 1960 and 1979

built40 = 1 if home was built between 1940 and 1959

built00 = 1 if home was built between 1900 and 1939

pre1900 = 1 if home was built before 1900

(omitted category is built 1980 or later)

bed1 = 1 if one bedroom

bed2 = 1 if two bedrooms

bed3 = 1 if three bedrooms

bed4 = 1 if four bedrooms

(omitted category is >4 bedrooms)

(all obs with <1 bedroom dropped)

bath1 = 1 if one bathroom

bath15 = 1 if 1.5 bathrooms

bath2 = 1 if 2 bathrooms

bath25 = 1 if 2.5 bathrooms

(omitted category is >2.5 bathrooms)

(all obs with <1 bathroom omitted)

sqft is the number of interior square feet

lotsize is the size of the lot on which the house is built in ACRES.

regscs is the combined MCAS score for "regular" students only (no "special" students)

vcrs is the violent crime rate (per capita, 1998)

pcrs is the property crime rate (per capita, 1998)

distance is the distance from the center of the zip to the center of the city / 100

density is the population density (residents per square mile, based

on 2000 census data, divided by 10000)

ownrate is the percent of owner-occupied homes in the zip (based on 2000 census data)

lnmedinc is the natural log of the median household income in the zip (based on 2000 census data)

return\_n is the average 1 year (overlapping) nominal return based on the CSW zip level house price index

volat\_n is the std. deviation of the average 1 year nominal return

strtr\_n is the structural trend from the time series regressions

rmse\_n is the root MSE from the same time series regressions

Chicago (where different from Boston):

built80 = 1 if house built after 1980

built70 = 1 if house built between 1970 and 1979

built60 = 1 if house built between 1960 and 1969

built50 = 1 if house built between 1950 and 1959

(omitted category: built before 1950)

bathrooms -- omitted category is >3 bathrooms

Phoenix (where different from Boston):

built80 = 1 if house built between 1980 and 1989

built70 = 1 if house built between 1970 and 1979

built60 = 1 if house built between 1960 and 1969

builtp60 = 1 if house built before 1960

(omitted category: built in 1990 or later)

There is no bedroom measure here; rather we have TOTAL ROOMS

rooms5 = 1 if the house has 5 rooms

rooms6 = 1 if the house has 6 rooms

rooms7 = 1 if the house has 7 rooms

rooms8 = 1 if the house has 8 or more rooms.

garage = 1 if the house has a garage

pool = 1 if the house has a pool

San Diego (where different from Boston):

builtXX (same as Phoenix)

Omitted category for bedrooms is >4 bedrooms

**Table 5.2a: Descriptive Statistics: Boston**

Variable	Obs	Mean	Std. Dev.	Min	Max
lnsale	19848	12.3152	.5591425	9.903487	15.03929
built60	19848	.2106006	.4077454	0	1
built40	19848	.255794	.4363176	0	1
built00	19848	.2508565	.433517	0	1
pre1900	19848	.0687223	.2529877	0	1
bed1	19848	.0075071	.0863196	0	1
bed2	19848	.1252519	.3310127	0	1
bed3	19848	.4939037	.4999754	0	1
bed4	19848	.3074869	.4614644	0	1
bath1	19848	.2450121	.4301052	0	1
bath15	19848	.2382104	.4259992	0	1
bath2	19848	.1629887	.3693647	0	1
bath25	19848	.2474809	.4315593	0	1
sqft	19848	1917.354	888.3383	375	14241
lotsize	19848	.5185134	.6836794	.0210514	9.956543
regsc0	19848	701.7816	24.18696	646	744
vcrate	19848	.0024855	.0032777	0	.0129326
pcrate	19848	.0189665	.0115578	0	.0479798
distance	19848	.2083409	.0988717	.0284	.4597
density	19848	3.104961	2.984181	.2630942	23.67669
ownrate	19848	.7103232	.1500517	.2705509	.9563218
nonwhite	19848	.1142403	.1172665	.0077124	.9431201
lnmedinc	19848	11.1274	.3018406	10.20152	11.94418
return_n	19848	.0837736	.0070835	.0700997	.1103469
volat_n	19848	.0935405	.008674	.0830847	.1537745
strtr_n	19848	.0521837	.0067157	.0407364	.0753547

Table 5.2b: Comparison of coefficients, with and without risk and return. Dependent variable: ln(sale price)

Boston	(1) Housing attributes only	(2) Housing attributes and raw risk/return	(3) Housing attributes with estimated risk/return	(4) Housing attributes and locational char.	(5) Full model with raw risk/return	(6) Full model with estimated risk/return
Built 1960-1980	0.0945*** (0.0081)	0.0054 (0.0072)	0.0147 (0.0074)	-0.0214*** (0.0073)	-0.0242*** (0.0071)	-0.02515*** (0.0072)
Built 1940-1960	0.1287*** (0.0092)	-0.0495*** (0.0084)	-0.0395*** (0.0085)	-0.0793*** (0.0092)	-0.0961*** (0.0084)	-0.0995*** (0.0084)
Built 1900-1940	0.0731** (0.0101)	-0.0752*** (0.0903)	-0.0896*** (0.0924)	-0.1148*** (0.0092)	-0.1205*** (0.0089)	-0.1260*** (0.0089)
Built pre-1900	0.0198*** (0.0143)	-0.1023*** (0.0128)	-0.1084*** (0.0130)	0.1619*** (0.0127)	0.1599*** (0.0124)	0.1634*** (0.0124)
1 bedroom	-0.1176*** (0.0342)	-0.1266*** (0.0326)	-0.1224*** (0.0325)	-0.1786*** (0.0331)	-0.1494*** (0.0328)	-0.1561*** (0.0323)
2 bedrooms	-0.0527*** (0.0179)	-0.0654*** (0.0152)	-0.0584*** (0.0154)	-0.0891*** (0.0152)	-0.0746*** (0.0148)	-0.0742*** (0.0148)
3 bedrooms	-0.0120 (0.0158)	-0.0192 (0.0133)	-0.0068 (0.0134)	-0.0350** (0.0133)	-0.0239* (0.0128)	-0.0199 (0.0128)
4 bedrooms	-0.0049 (0.0153)	-0.0006 (0.0129)	0.0168 (0.0131)	-0.0183 (0.0130)	-0.0115 (0.0125)	-0.0047 (0.0126)
1 bathroom	-0.5418*** (0.0151)	-0.3170*** (0.0129)	-0.3496*** (0.0131)	-0.2892*** (0.0131)	-0.2508*** (0.0126)	-0.2530*** (0.0126)
1.5 bathrooms	-0.3567*** (0.0141)	-0.2091*** (0.0119)	-0.2290*** (0.0122)	-0.1991*** (0.0120)	-0.1688*** (0.0115)	-0.1695*** (0.0115)
2 bathrooms	-0.3838*** (0.0145)	-0.2356*** (0.0124)	-0.2576*** (0.0127)	-0.2317*** (0.0123)	-0.1995*** (0.0119)	-0.2034*** (0.0119)
2.5 bathrooms	0.1468*** (0.0123)	-0.0827*** (0.0102)	-0.0815*** (0.0105)	-0.0907*** (0.0104)	-0.0733*** (0.0099)	-0.0713*** (0.0100)
interior square feet	0.0004*** (0.00001)	0.0003*** (0.00001)	0.0003*** (0.00001)	0.0003*** (0.00001)	0.0003*** (0.00001)	0.0003*** (0.00001)
sq. feet squared	-1.56E-08*** (2.35E-09)	-1.52E-08*** (2.43E-09)	-1.39E-08*** (2.59E-09)	-1.17E-08*** (2.48E-09)	-1.31E-08*** (2.54E-09)	-1.22E-08*** (2.63E-09)
lot size in acres	0.1361*** (0.0104)	0.1644*** (0.0098)	0.2162*** (0.0105)	0.1258*** (0.0142)	0.1331*** (0.0096)	0.1378*** (0.0139)
lot size squared	-0.0168*** (0.0020)	-0.0193*** (0.0019)	-0.0260*** (0.0021)	-0.0147*** (0.0044)	-0.0154*** (0.0017)	-0.0161*** (0.0043)
distance to city center/100				-1.637*** (0.1233)	-0.6086*** (0.1566)	-0.4220*** (0.1439)
distance squared				1.094*** (0.2179)	0.2430 (0.2466)	-0.0431** (0.2339)
Population density/10000				0.0003 (0.0020)	0.0093*** (0.0021)	0.0025 (0.0020)
homeownership rate				-0.8259*** (0.0366)	-0.5073*** (0.0402)	-0.6052*** (0.0378)
percent nonwhite				-0.4078*** (0.0302)	-0.2017*** (0.0333)	-0.2749*** (0.0361)
log of median income				0.8076*** (0.0230)	0.5830*** (0.0253)	0.6175*** (0.0248)
MCAS score (regular students only)				0.0026*** (0.0002)	0.0025*** (0.0002)	0.0027*** (0.0002)
violent crime rate				1.439 (1.174)	2.492*** (1.207)	-0.2287 (1.219)
property crime rate				-0.0258 (0.3247)	0.2424 (0.3295)	0.7137** (0.3303)
1-year return		25.37*** (0.4177)			11.64*** (0.8006)	
1-year risk		-22.91*** (0.4078)			-9.871*** (0.6290)	
structural trend			33.21*** (0.4579)			14.88*** (0.6715)
root MSE			-12.09*** (0.4363)			-3.161*** (0.6728)
R-squared	0.5266	0.6598	0.6394	0.6886	0.6967	0.6968
N	19848	19848	19848	19848	19848	19848

Robust std. errors in parentheses. \*, \*\*, and \*\*\* represent statistical significance at the 10%, 5%, and 1% levels, respectively.



Table 5.2c: Comparison of coefficients, historic risk and actual or predicted return '98-'03. Dependent variable: ln(sale price)

Boston	(1) Housing attributes historic risk actual return '98-'03	(2) Full model with historic risk actual return '98-'03	(3) Housing attributes historic risk predicted return	(4) Full model with historic risk predicted return
distance to city center/100		-2.002*** (0.1267)		-1.776*** (0.1253)
distance squared		1.811*** (0.2216)		1.478*** (0.2220)
Population density/10000		0.0115*** (0.0021)		0.0083*** (0.0021)
homeownership rate		-0.5726*** (0.0401)		-0.5821*** (0.0401)
percent nonwhite		-0.1376*** (0.0339)		-0.1576 (0.0327)
log of median income		0.6639*** (0.0246)		0.6431*** (0.0246)
MCAS score (regular students only)		0.0024*** (0.0002)		0.0026*** (0.0002)
violent crime rate		4.955*** (1.266)		2.691*** (1.208)
property crime rate		-0.2351 (0.3310)		0.4902 (0.3379)
1-year average return	12.68*** (0.524)	-0.2294 (0.6403)	13.00*** (0.4250)	3.873*** (0.4141)
1-year historic risk	-32.55*** (0.881)	-9.266*** (0.8662)	-8.078*** (0.530)	-5.734*** (0.7557)
R-squared	0.6037	0.6932	0.6099	0.6945
N	19848	19848	19848	19848

Robust std. errors in parentheses. \*, \*\*, and \*\*\* represent statistical significance at the 10%, 5%, and 1% levels, respectively. The above regressions all include the housing attributes found in Table 5.2b. These coefficients have been suppressed for visual clarity.

**Table 5.3a: Descriptive Statistics: Chicago**

Variable	Obs	Mean	Std. Dev.	Min	Max
lnsale	12799	12.03402	.5270517	9.21034	18.61541
built80	12799	.0682866	.2522469	0	1
built70	12799	.107274	.3094734	0	1
built60	12799	.1761856	.3809929	0	1
built50	12799	.2913509	.4544026	0	1
bath1	12799	.4124541	.4922953	0	1
bath15	12799	.237206	.4253863	0	1
bath2	12799	.1469646	.3540845	0	1
bath25	12799	.1645441	.3707831	0	1
bath3	12799	.019064	.1367553	0	1
sqft	12799	1489.091	654.2923	400	11860
lotsize	12799	.1671203	.1449357	.0249541	6.157805
distance	12799	.1865422	.0895033	.0505	.3758
density	12799	7.909259	5.703254	2.095443	31.86026
ownrate	12799	.702625	.1490591	.2000984	.8846387
nonwhite	12799	.2545424	.2004363	.0516317	.9917094
lnmedinc	12799	10.91572	.2494083	10.22915	11.58773
return_n	12799	.0680904	.0129333	.0497799	.1055857
volat_n	12799	.037281	.0064613	.0258375	.05736
strtr_n	12799	.0607937	.0115686	.041798	.0936214
rmse_n	12799	.0250315	.0058237	.0098268	.0349461

Table 5.3b: Comparison of coefficients, with and without risk and return. Dependent variable: ln(sale price)

Chicago	(1) Housing attributes only	(2) Housing attributes and raw risk/return	(3) Housing attributes with estimated risk/return	(4) Housing attributes and locational char.	(5) Full model with raw risk/return	(6) Full model with estimated risk/return
Built after 1980	-0.0643*** (0.0133)	0.0863*** (0.0154)	0.1164*** (0.0153)	-0.0147 (0.0150)	0.0764*** (0.0143)	0.0753*** (0.0148)
Built 1970-1980	-0.0686*** (0.0116)	0.0718*** (0.0135)	0.0935*** (0.0131)	-0.0491*** (0.0127)	0.0347*** (0.0121)	0.0234*** (0.0124)
Built 1960-1970	-0.0225** (0.0103)	0.0759*** (0.0115)	0.1027*** (0.0112)	-0.0128 (0.0109)	0.0326*** (0.0103)	0.0365*** (0.0105)
Built 1950-1960	0.0280*** (0.0094)	0.0872*** (0.0096)	0.0950*** (0.0093)	0.0091*** (0.0094)	0.0456*** (0.0088)	0.0350*** (0.0088)
1 bathroom	-0.3926*** (0.0424)	-0.3740*** (0.0414)	-0.4020*** (0.0430)	-0.3767*** (0.0409)	-0.3472*** (0.0399)	-0.3792*** (0.0413)
1.5 bathrooms	-0.2589*** (0.0418)	-0.2259*** (0.0408)	-0.2613*** (0.0425)	-0.2871*** (0.0407)	-0.2584*** (0.0395)	-0.2947*** (0.0410)
2 bathrooms	-0.2493*** (0.0418)	-0.2246*** (0.0409)	-0.2548*** (0.0425)	-0.2585*** (0.0406)	-0.2346*** (0.0395)	-0.2565*** (0.0409)
2.5 bathrooms	-0.1564*** (0.0402)	-0.1231*** (0.0394)	-0.1522*** (0.0411)	-0.1786*** (0.0392)	-0.1590*** (0.0382)	-0.1923*** (0.0396)
3 bathrooms	-0.1084** (0.0470)	-0.0920** (0.0459)	-0.1056** (0.0476)	-0.1342*** (0.0455)	-0.1269*** (0.0437)	-0.1407*** (0.0453)
interior square feet	0.0006*** (0.00002)	0.0006*** (0.00002)	0.0005*** (0.00002)	0.0005*** (0.00002)	0.0005*** (0.00002)	0.0005*** (0.00002)
sq. feet squared	-4.71E-08*** (4.58E-09)	-4.48E-08*** (4.74E-09)	-4.24E-08*** (4.58E-09)	-4.07E-08*** (4.31E-09)	-3.69E-08*** (4.15E-09)	-3.63E-08*** (4.16E-09)
lot size in acres	0.5373*** (0.0533)	0.8666*** (0.0745)	0.8771*** (0.0752)	0.3243*** (0.0496)	0.4653*** (0.0508)	0.4049*** (0.0530)
lot size squared	-0.0896*** (0.00255)	-0.1548*** (0.0378)	-0.1595*** (0.0405)	-0.0409*** (0.0148)	-0.0701*** (0.0185)	-0.0600*** (0.0196)
distance to city center/100				0.2412 (0.2376)	1.181*** (0.2607)	1.447*** (0.2480)
distance squared				-3.563*** (0.4902)	-4.080*** (0.4904)	-5.737*** (0.5134)
Population density/10000				0.0364** (0.0013)	-0.0052*** (0.0014)	-0.0093*** (0.0014)
homeownership rate				-1.229*** (0.0423)	-1.003*** (0.0471)	-0.9422*** (0.0467)
percent nonwhite				-0.5232*** (0.0239)	-0.3734*** (0.0254)	-0.3711*** (0.0258)
log of median income				1.228*** (0.0259)	1.097*** (0.0281)	1.037*** (0.0283)
1-year return		12.40*** (0.5932)			12.35*** (0.6650)	
1-year risk		-12.17*** (1.127)			-8.728*** (0.9402)	
structural trend			10.62*** (0.439)			10.79*** (0.5623)
root MSE			-2.107*** (0.6271)			-6.491*** (0.7639)
R-squared	0.4429	0.4656	0.4729	0.6014	0.6108	0.6143
N	12799	12799	12799	12799	12799	12799

Robust std. errors in parentheses. \*, \*\*, and \*\*\* represent statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5.3c: Comparison of coefficients, historic risk and actual or predicted return '98-'03. Dependent variable: ln(sale price)

Chicago	(1) Housing attributes historic risk actual return '98-'03	(2) Full model with historic risk actual return '98-'03	(3) Housing attributes historic risk predicted return	(4) Full model with historic risk predicted return
distance to city center/100		2.249*** (0.2680)		0.3485 (0.2487)
distance squared		-7.108*** (0.5249)		-3.858*** (0.5024)
Population density/10000		-0.0029** (0.0014)		0.0031** (0.0013)
homeownership rate		-0.9371*** (0.0467)		-1.223*** (0.0455)
percent nonwhite		-0.3341*** (0.02511)		-0.5198*** (0.0235)
log of median income		1.010*** (0.2893)		1.205*** (0.0281)
1-year average return	9.091*** (0.2576)	6.303*** (0.2704)	-15.356*** (0.712)	-1.334* (0.767)
1-year historic risk	-6.590*** (0.8446)	-2.327*** (0.8888)	20.797*** (1.023)	1.238 (1.117)
R-squared	0.4915	0.6148	0.4699	0.6015
N	12799	12799	12799	12799

Robust std. errors in parentheses. \*, \*\*, and \*\*\* represent statistical significance at the 10%, 5%, and 1% levels, respectively. The above regressions all include the housing attributes found in Table 5.3b. These coefficients have been suppressed for visual clarity.

**Table 5.4a: Descriptive Statistics: Phoenix**

Variable	Obs	Mean	Std. Dev.	Min	Max
lnsale	36347	11.79828	.5105783	9.21034	17.19278
built80	36347	.1755303	.3804253	0	1
built70	36347	.1625994	.3690048	0	1
built60	36347	.0628112	.2426265	0	1
builtp60	36347	.0957438	.2942436	0	1
bath1	36347	.0618208	.2408331	0	1
bath15	36347	.0017333	.0415973	0	1
bath2	36347	.3261892	.4688239	0	1
bath25	36347	.1411396	.3481703	0	1
bath3	36347	.0959089	.2944703	0	1
rooms5	36347	.2213663	.4151722	0	1
rooms6	36347	.2670922	.442447	0	1
rooms7	36347	.2042534	.4031606	0	1
rooms8	36347	.2591686	.4381843	0	1
sqft	36347	1887.79	711.7943	302	11347
lotsize	36347	.2225363	.2233044	.0252525	13.09401
pool	36347	.3199714	.4664715	0	1
garage	36347	.7959942	.4029788	0	1
distance	36347	.1941158	.0706613	.0073	.396
density	36347	4.269927	2.65339	.218084	19.23176
ownrate	36347	.6763427	.1342984	.2825996	.9554904
nonwhite	36347	.170428	.1215788	.0170152	.6032994
lnmedinc	36347	10.84074	.3041278	10.00428	11.4894
return_n	36347	.0459049	.007016	.0361343	.0792224
volat_n	36347	.0332599	.0062594	.0246634	.0571979
strtr_n	36347	.0647912	.0112705	.0452661	.1148396
rmse_n	36347	.0157675	.0052074	.0086453	.036557

Table 5.4b: Comparison of coefficients, with and without risk and return. Dependent variable: ln(sale price)

Phoenix	(1) Housing attributes only	(2) Housing attributes and raw risk/return	(3) Housing attributes with estimated risk/return	(4) Housing attributes and locational char.	(5) Full model with raw risk/return	(6) Full model with estimated risk/return
Built 1980-1990	-0.0810*** (0.0041)	-0.0890*** (0.0041)	-0.0916*** (0.0043)	-0.0847*** (0.0042)	-0.0776*** (0.0042)	-0.0776*** (0.0043)
Built 1970-1980	-0.1632*** (0.0057)	-0.1844*** (0.0057)	-0.1764*** (0.0061)	-0.1734*** (0.0062)	-0.1720*** (0.0061)	-0.1644*** (0.0063)
Built 1960-1970	-0.2016*** (0.0075)	-0.2548*** (0.0075)	-0.2118*** (0.0080)	-0.2224*** (0.0080)	-0.2357*** (0.0079)	-0.2170*** (0.0080)
Built before 1960	-0.1829*** (0.0083)	-0.3065*** (0.0091)	-0.1865*** (0.0097)	-0.2041*** (0.0102)	-0.2342*** (0.0102)	-0.2058*** (0.0102)
1 bathroom	-0.1896*** (0.0119)	-0.2004*** (0.0118)	-0.1834*** (0.0117)	-0.1629*** (0.0105)	-0.1693*** (0.0105)	-0.1675*** (0.0105)
1.5 bathrooms	-0.1425*** (0.0286)	-0.1615*** (0.0283)	-0.1386*** (0.0284)	-0.1436*** (0.0262)	-0.1583*** (0.0261)	-0.1531*** (0.0263)
2 bathrooms	-0.1113*** (0.0056)	-0.1053*** (0.0055)	-0.1065*** (0.0056)	-0.1004*** (0.0051)	-0.1000*** (0.0050)	-0.1045*** (0.0051)
2.5 bathrooms	-0.0420*** (0.0050)	-0.0354*** (0.0050)	-0.0392*** (0.0050)	-0.0422*** (0.0048)	-0.0426*** (0.0048)	-0.0451*** (0.0048)
3 bathrooms	-0.0010 (0.0063)	-0.0068 (0.0063)	-0.0015 (0.0064)	-0.0016 (0.0061)	-0.0024 (0.0060)	-0.0014 (0.0060)
5 or fewer total rooms	-0.0345*** (0.0098)	-0.0386*** (0.0100)	-0.0361*** (0.0098)	-0.0118 (0.0092)	-0.0120 (0.0092)	-0.0108 (0.0093)
6 total rooms	-0.0729*** (0.0125)	-0.0784*** (0.0126)	-0.0761*** (0.0125)	-0.0427*** (0.0114)	-0.0408*** (0.0114)	-0.0405*** (0.0115)
7 total rooms	-0.1302*** (0.0156)	-0.1332*** (0.0157)	-0.1339*** (0.0156)	-0.0868*** (0.0139)	-0.0858*** (0.0139)	-0.0848*** (0.0141)
8 total rooms	-0.1977*** (0.0166)	-0.2017*** (0.0166)	-0.2038*** (0.0166)	-0.1324*** (0.0151)	-0.1304*** (0.0150)	-0.1275*** (0.0152)
interior square feet	0.0007*** (0.00004)	0.0007*** (0.00004)	0.0008*** (0.00004)	0.0007*** (0.00004)	0.0007*** (0.00004)	0.0007*** (0.00004)
sq. feet squared	-5.73E-08*** (8.58E-09)	-5.70E-08*** (8.62E-09)	-5.73E-08*** (8.55E-09)	-4.79E-08*** (7.27E-09)	-4.79E-08*** (7.28E-09)	-4.80E-08*** (7.31E-09)
lot size in acres	0.2912*** (0.0247)	0.2378*** (0.0242)	0.2963*** (0.0248)	0.2566*** (0.0222)	0.2483*** (0.0219)	0.2528*** (0.0220)
lot size squared	-0.0238*** (0.0078)	-0.0220*** (0.0075)	-0.0241*** (0.0078)	-0.0204*** (0.0062)	-0.0194*** (0.0062)	-0.0200*** (0.0061)
pool	0.0782*** (0.0036)	0.0703*** (0.0036)	0.0772*** (0.0036)	0.0683*** (0.0034)	0.0613*** (0.0034)	0.0669*** (0.0034)
garage	0.0774*** (0.0055)	0.0831*** (0.0053)	0.0761*** (0.0054)	0.0452*** (0.0048)	0.0553*** (0.0048)	0.0462*** (0.0048)
distance to city center/100				-2.200*** (0.1821)	-0.2472 (0.2188)	-1.094*** (0.2269)
distance squared				4.392*** (0.4635)	-0.0395 (0.5525)	1.849*** (0.5715)
Population density/10000				-0.004*** (0.0007)	-0.0044*** (0.0007)	-0.0040*** (0.0007)
homeownership rate				-0.7284*** (0.0160)	-0.6724*** (0.0171)	-0.7387*** (0.0172)
percent nonwhite				-0.7873*** (0.0174)	-0.7439*** (0.0201)	-0.8233*** (0.0193)
log of median income				0.4330*** (0.0091)	0.4063*** (0.0093)	0.4357*** (0.0093)
1-year return		10.27*** (0.3125)			4.887*** (0.3759)	
1-year risk		-3.026*** (0.3258)			3.965*** (0.3170)	
structural trend			2.064*** (0.2182)			1.726*** (0.2391)
root MSE			-5.670*** (0.2946)			3.597*** (0.3471)
R-squared	0.6822	0.6937	0.6858	0.7337	0.7374	0.7349
N	36347	36347	36347	36347	36347	36347

Robust std. errors in parentheses. \*, \*\*, and \*\*\* represent statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5.4c: Comparison of coefficients, historic risk and actual or predicted return '98-'03. Dependent variable: ln(sale price)

Phoenix	(1) Housing attributes historic risk actual return '98-'03	(2) Full model with historic risk actual return '98-'03	(3) Housing attributes historic risk predicted return	(4) Full model with historic risk predicted return
distance to city center/100		-1.561*** (0.2085)		-0.5880*** (0.2007)
distance squared		2.673*** (0.5185)		0.3866 (0.5111)
Population density/10000		-0.0043*** (0.0007)		-0.0061*** (0.0007)
homeownership rate		-0.7272*** (0.0167)		-0.6372*** (0.0173)
percent nonwhite		-0.8852*** (0.0186)		-0.7632*** (0.0187)
log of median income		0.4233*** (0.0094)		0.4080*** (0.0092)
1-year return	.7475*** (0.1948)	-0.4306** (0.2041)	3.334*** (0.0925)	1.634*** (0.1004)
1-year historic risk	-1.167*** (0.3498)	4.971*** (0.3306)	1.331*** (0.3197)	5.811*** (0.3119)
R-squared	0.6824	0.7356	0.6965	0.7381
N	36347	36347	36347	36347

Robust std. errors in parentheses. \*, \*\*, and \*\*\* represent statistical significance at the 10%, 5%, and 1% levels, respectively. The above regressions all include the housing attributes found in Table 5.4b. These coefficients have been suppressed for visual clarity.

**Table 5.5a: Descriptive Statistics: San Diego**

Variable	Obs	Mean	Std. Dev.	Min	Max
lnsale	15068	12.32528	.5096997	9.21034	16.53798
built80	15068	.220069	.4143067	0	1
built70	15068	.2127024	.4092325	0	1
built60	15068	.1154101	.3195269	0	1
builtp60	15068	.1509822	.3580434	0	1
bed1	15068	.0039156	.0624541	0	1
bed2	15068	.1088399	.3114486	0	1
bed3	15068	.4357579	.4958722	0	1
bed4	15068	.3511415	.4773429	0	1
bath1	15068	.0884656	.2839803	0	1
bath15	15068	.0245553	.1547707	0	1
bath2	15068	.4009822	.4901137	0	1
bath25	15068	.2250465	.4176268	0	1
bath3	15068	.1784577	.3829102	0	1
sqft	15068	1952.194	777.1708	396	7099
garage	15068	.9386116	.2400493	0	1
pool	15068	.1732811	.3785027	0	1
distance	15068	.2499059	.1225991	.0316	.5352
density	15068	3.119691	2.692898	.113493	12.65931
ownrate	15068	.6170105	.1328326	.257829	.861325
nonwhite	15068	.2769033	.152345	.0862681	.7832828
lnmedinc	15068	10.90891	.2786167	10.15007	11.48277
return_n	15068	.074254	.0060522	.0574699	.0896487
volat_n	15068	.0842093	.0050614	.0733829	.1028171
strtr_n	15068	.0553981	.0059435	.0385058	.0671204
rmse_n	15068	.0389528	.0057099	.0316088	.06628



Table 5.5b: Comparison of coefficients, with and without risk and return. Dependent variable: ln(sale price)

San Diego	(1) Housing attributes only	(2) Housing attributes and raw risk/return	(3) Housing attributes with estimated risk/return	(4) Housing attributes and locational char.	(5) Full model with raw risk/return	(6) Full model with estimated risk/return
Built 1980-1990	-0.0116 (0.0081)	-0.0185** (0.0075)	-0.0249*** (0.0075)	-0.0214*** (0.0075)	-0.0072 (0.0073)	-0.0135* (0.0073)
Built 1970-1980	-0.0683*** (0.0092)	-0.0847*** (0.0086)	-0.0993*** (0.0087)	-0.1016*** (0.0087)	-0.0851*** (0.0084)	-0.0961*** (0.0084)
Built 1960-1970	-0.0463** (0.0101)	-0.0872*** (0.0105)	-0.1088** (0.0106)	-0.1305*** (0.0108)	-0.1003*** (0.0105)	-0.1104*** (0.0105)
Built before 1960	0.0275** (0.0143)	-0.0492*** (0.0124)	-0.0871** (0.0129)	-0.0907*** (0.0135)	-0.0689*** (0.0131)	-0.0798*** (0.0131)
1 bedroom	0.1518*** (0.0342)	0.3062*** (0.0619)	0.2972*** (0.0615)	0.1586*** (0.0552)	0.1797*** (0.0547)	0.1808*** (0.0538)
2 bedrooms	0.0817*** (0.0179)	0.1714*** (0.0163)	0.1659*** (0.0163)	0.0991*** (0.0160)	0.1056*** (0.0153)	0.1109*** (0.0155)
3 bedrooms	0.0453*** (0.0158)	0.1070*** (0.0121)	0.1030*** (0.0122)	0.5705*** (0.0121)	0.0736*** (0.0114)	0.0742*** (0.0117)
4 bedrooms	-0.0248** (0.0153)	-0.0512*** (0.0103)	-0.0476** (0.0104)	-0.0365*** (0.0103)	-0.0437*** (0.0097)	-0.0417*** (0.0099)
1 bathroom	-0.2061*** (0.0151)	-0.1332*** (0.0250)	-0.1384*** (0.0248)	-0.1718*** (0.0246)	-0.1397*** (0.0236)	-0.1418*** (0.0236)
1.5 bathrooms	-0.2071*** (0.0141)	-0.1155*** (0.0281)	-0.1170*** (0.0278)	-0.1465*** (0.0272)	-0.1078*** (0.0262)	-0.1117*** (0.0262)
2 bathrooms	-0.1203*** (0.0145)	-0.0417** (0.0196)	-0.0463** (0.0195)	-0.0946*** (0.0196)	-0.0493*** (0.0186)	-0.0607*** (0.0187)
2.5 bathrooms	-0.0695*** (0.0123)	-0.0515*** (0.0173)	-0.0572*** (0.0172)	-0.0907*** (0.0174)	-0.0522*** (0.0166)	-0.0623*** (0.0167)
3 bathrooms	-0.0822*** (0.0123)	-0.0388** (0.0156)	-0.0437*** (0.0156)	-0.0684*** (0.0157)	-0.0349** (0.0149)	-0.0435*** (0.0150)
interior square feet	0.0007*** (0.00001)	0.0007*** (0.00002)	0.0007*** (0.00002)	0.0006*** (0.00002)	0.0006*** (0.00002)	0.0006*** (0.00002)
sq. feet squared	-4.79E-08*** (2.42E-09)	-5.78E-08*** (4.42E-09)	-5.71E-08*** (4.39E-09)	-3.94E-08*** (4.29E-09)	-4.06E-08*** (4.18E-09)	-4.19E-08*** (4.16E-09)
pool	0.1458*** (0.0045)	0.0464*** (0.0077)	0.0454*** (0.0077)	0.0204*** (0.0078)	0.0369*** (0.0074)	0.0365*** (0.0075)
garage	0.0360*** (0.0141)	0.1301*** (0.0136)	0.1289*** (0.0135)	0.1247*** (0.0135)	0.1072*** (0.0128)	0.1135*** (0.0127)
distance to city center/100				-1.033 (0.1316)	-0.6084*** (0.1422)	-0.3208** (0.1463)
distance squared				1.397*** (0.2289)	0.3908 (0.2690)	0.4989* (0.2522)
Population density/10000				0.0108*** (0.0014)	-0.0012 (0.0015)	0.0022 (0.0013)
homeownership rate				-1.144*** (0.0293)	-0.8078*** (0.0362)	-0.9229*** (0.0313)
percent nonwhite				-0.5900*** (0.0189)	-0.5986*** (0.0186)	-0.4876*** (0.0191)
log of median income				0.8090*** (0.0164)	0.6663*** (0.0184)	0.7544*** (0.0163)
1-year return		13.58*** (0.5454)			2.691*** (0.6542)	
1-year risk		3.606*** (0.6313)			10.22*** (0.8934)	
structural trend			12.10*** (0.5244)			6.017*** (0.5955)
root MSE			8.400*** (0.6147)			8.500*** (0.6339)
R-squared	0.5646	0.5917	0.5922	0.6781	0.6826	0.6827
N	15068	15068	15068	15068	15068	15068

Robust std. errors in parentheses. \*, \*\*, and \*\*\* represent statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5.5c: Comparison of coefficients, historic risk and actual or predicted return '98-'03. Dependent variable: ln(sale price)

Phoenix	(1) Housing attributes historic risk actual return '98-'03	(2) Full model with historic risk actual return '98-'03	(3) Housing attributes historic risk predicted return	(4) Full model with historic risk predicted return
distance to city center/100		-1.247*** (0.1698)		-0.5889* (0.1421)
distance squared		0.9511*** (0.2840)		0.0824 (0.2587)
Population density/10000		0.0014 (0.0015)		-0.0001 (0.0015)
homeownership rate		-0.7285*** (0.0382)		-0.7868*** (0.0372)
percent nonwhite		-0.5848*** (0.0186)		-0.5971*** (0.0190)
log of median income		0.5449*** (0.0248)		0.6401*** (0.0203)
1-year real return	-5.475*** (0.2069)	-3.430*** (0.4097)	-3.930*** (0.2096)	-1.139*** (0.2508)
1-year real risk	13.49*** (0.6173)	12.55*** (0.8348)	9.523*** (0.6035)	11.74*** (0.8310)
R-squared	0.5919	0.6836	0.5796	0.6825
N	15068	15068	15068	15068

Robust std. errors in parentheses. \*, \*\*, and \*\*\* represent statistical significance at the 10%, 5%, and 1% levels, respectively. The above regressions all include the housing attributes found in Table 5.5b. These coefficients have been suppressed for visual clarity.