Micro-level Return and Volatility Drivers in Boston's Single Family Home Market

by

Jay Valenta
B.ARCH., Architecture, 1993
Rensselaer Polytechnic Institute

Submitted to the Department of Urban Studies and Planning in Partial Fulfillment of the Requirements for the Degree of Master of Science in Real Estate Development

at the

Massachusetts Institute of Technology
September 2003

Signature of Author ______________________
Department of Urban Studies and Planning
August 4, 2003

Certified by ______________________
Henry O. Pollakowski
Visiting Scholar, Center for Real Estate, MIT

Accepted by ______
David Geltner
Chairman, Interdepartmental Degree Program in Real Estate Development

Micro-level Return and Volatility Drivers in Boston's Single Family Home Market

by

Jay Valenta
B.ARCH., Architecture, 1993
Rensselaer Polytechnic Institute

Submitted to the *Department of Urban Studies and Planning* on August 4, 2003 in Partial Fulfillment of the Requirements for the Degree of Master of Science in Real Estate Development

Abstract

Developers and investors commonly target neighborhoods close to the urban core and with low median incomes as potential growth markets. Investments in these areas however are often perceived by private sector capital as being high risk and low return. An understanding of the predictors of investment volatility and return is critical to investors and homeowners who wish to maximize investment returns and portfolio growth. Moreover, for mortgage lenders who are obligated to invest in a wide spectrum of communities, volatilities in house prices may affect the distribution of their collateral values, the probability of default, and the profitability of lending in certain areas.

This paper addresses the following questions: Does appreciation return and volatility in metropolitan house prices vary significantly among zip code areas? Can the variation in appreciation return and volatility among these areas be explained by additional data?

This paper uses appreciation and volatility statistics calculated from a repeat sale index of house prices in metropolitan Boston compiled biannually by Case Schiller Weiss (CSW) as well as data gathered from the 2000 U.S. Census and the 1997 Economic Census.

Thesis supervisor: Henry O. Pollakowski, Visiting Scholar, Center for Real Estate

Table of Contents

Introduction

Inter-metropolitan area appreciation rates

Appreciation rates can vary significantly among metropolitan markets in the United States. Preliminary studies by Case and Mayer (1986) showed that housing values appreciated in Boston 57.8% between 1983 and 1985, while Albany, Houston, Cincinnati, Miami, San Jose sustained nominal declines between 1981 and 1985.[1] In this study, Case et al. found the drivers of house prices were primarily construction cost and income growth.[2] Neither population nor employment growth nor demographics were significant enough to explain these trends.[3]

Abraham, et al. (1994) show that although review of national trends is meaningful, metropolitan area trends can vary from national trends, and metropolitan area trends may lead or lag national trends by up to two years.[4] This paper further suggests that return data aggregated to the MSA level can be used as consistent inputs to mean-variance optimization methods developed by Markowitz (1952) to provide diversification benefits to investors.[5]

In another study, Abraham, et al. (1994) found that certain groups of variables can vary in explanatory power between metropolitan areas. This paper found that variables including real income, real construction costs, changes in the real after-tax interest rate, and differences between the actual and equilibrium real house price levels predicted about one-half of the price movements in northeastern United States.[6]

---

[1] Case, 1986. p. 41.
[2] Case, 1986. p. 42.
[3] Case, 1986. p. 42.
[4] Abraham, Goetzmann, and Wachter, 1994. p. 191.
[5] Abraham, Goetzmann, and Wachter, 1994. p. 188.
[6] Abraham and Hendershott, 1994. p. 14.

According to the June 2003 report issued by the Office of Federal Housing Enterprise Oversight (OFHEO), Boston was ranked 55th among 220 MSAs for annual growth in house prices over the last 10 years.[7]

Intra-metropolitan, or sub-market, appreciation rates

Delineation of macro level sub-markets has added specificity to recent research. Bourassa, et al. (2002) found that established neighborhood or other urban boundaries probably define suitable sub-markets.[8] As modern efforts of data collection, such as those begun by Case and others reach fruition, more thorough analysis can be completed.

A preliminary study by Case, et al. (1996) using data from the metropolitan Boston area showed that towns with greatest appreciation rates were clustered around Boston and on the South Shore. Towns west of Boston, but still in the metropolitan area had the lowest rates of appreciation.[9] This current study continues the work of Case et. al. using an expanded data set, and finds similar conclusions.

Additional factors must be considered when reviewing macro-level house price appreciation data. Although repeat sale indexes are 'quality controlled,' growth rates of house prices can vary with the quality of the home. Smith and Tesarek (1991) found that 'higher quality' homes in Houston appreciated faster during the boom, and fell further during the bust than did 'lower quality' homes.[10] Mayer (1993) found a pattern of increased volatility for high-priced homes in four cities during the 1970's and 1980's.[11] Finally, Case & Schiller (1994) show the same result as Mayer for Los Angeles, and an opposite result for Boston where lower-

[7] Office of Federal Housing Enterprise Oversight, First Quarter 2003 Report, p. 15.
[8] Bourassa, Hoeslib, and Pengd, 2002.
[9] Case and Mayer, 1996. p. 396.
[10] Smith and Tesarek, 1991.
[11] Mayer, 1993.

tier properties appreciated most during the boom and fell farther during the bust.[12]

Data

The Case Schiller Weiss Repeat Sale Index

This paper uses a repeat sale house price index compiled by Case Schiller Weiss (CSW) by zip code for metropolitan Boston.

Case Schiller Weiss has confirmed that it has performed a smoothing process called spatial autocorrelation upon certain zip codes in the house price index. This process uses transactions from adjacent zip codes to reduce random noise in the return series of zip code areas in which a small number of repeat transactions are recorded. This process also increases correlations between price and volatility among adjacent zip codes.

Repeat sales indexes generally have some other sources of bias and inefficiency. Properties included in a repeat sale index are considered constant quality, however these properties have depreciated between transactions. Case, et al. (1992) have confirmed that repeat sale indexes bias return series downward due to confounding effect of increasing age[13].

Moreover, DiPasquale et al. (1993) have found that improved properties (excluded from a repeat sale index) will typically be more expensive properties, and unimproved properties (included in a repeat sale index) will typically consist of less expensive properties. DiPasquale et al., have also found that recent movers (a subset of properties included in a repeat sales index, recorded by AHS

[12] Case and Mayer, 1996. p. 388.
[13] Case, Pollakowski, and Wachter, 1992.

as transactions occurring within the last 12 months) occupy better, newer units with more rooms, bathrooms and garages.[14]

This bias would not be troubling if unchanged properties could be assumed to represent the population of all transacting properties. Case et al. however have confirmed large differences in subsets of properties transacted only once, unimproved properties transacted more than once, and improved properties transacted more than once.[15]

Repeat sale indexes offer some advantages over hedonic indexes constructed from property attributes. As mentioned, repeat sales indexes are quality controlled. Therefore, repeat sale indexes avoid the problem of selecting possibly biased or incorrect attributes necessary to construct a hedonic index.[16]

Dependant Variables

Volatility is recorded here as the standard deviation of the inflation adjusted biannual return (STDEV.) Standard deviation implies a simple measure of risk. The standard deviation measure used in this paper "is risk in the *realized* returns [since these returns] are stochastic, random realizations over time."[17] According to David Geltner, Professor of Real Estate Finance at Massachusetts Institute of Technology, risk is more accurately an ex ante statistic, and is represented by the range or deviation of the possible future return outcomes around the prior expectation of the mean.[18]

However, the degree to which *ex post* return statistics predict *ex ante* risk in real estate is debated. Hendershott et al. (2002) found historical return distributions of real estate assets to be a viable predictor of future return distributions, or risk

[14] DiPasquale and Somerville, 1993. p. 200.
[15] Case, Pollakowski, and Wachter, 1992.
[16] Case, Pollakowski, and Wachter, 1992. p. 4.
[17] David Geltner, Professor of Real Estate Finance, Director of the Center for Real Estate, MIT
[18] Geltner and Miller, 2001. p. 195.

in the ex ante return.[19] Although this method of predicting future return distributions may yield an upper bound of risk, Hendershott et al. claim that mean reversion and auto-correlation in asset returns (characteristics that tend to increase predictability and decrease risk in real estate investments) do not uniquely affect real estate assets, and therefore, we should not disregard the value of historic return distributions of real estate assets in predicting future risk in these assets.[20]

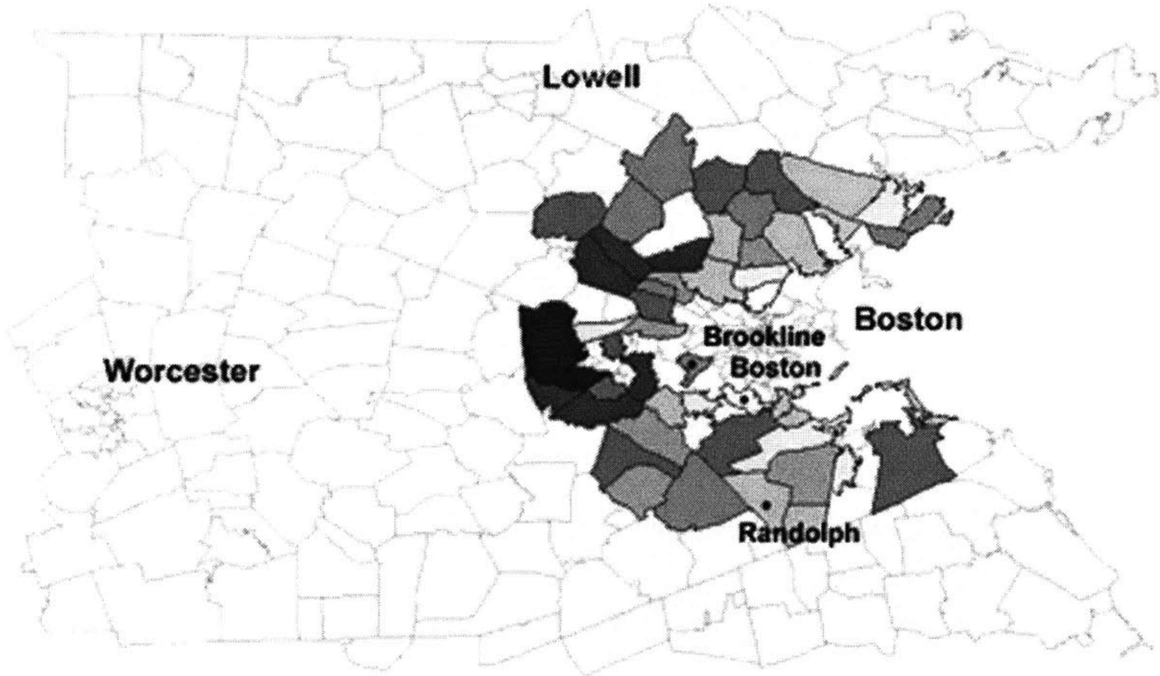
The return measure reported in this paper is the geometric mean return (GEOMEAN) based upon inflation adjusted biannual (six month) periodic returns. This statistic measures capital appreciation relevant to the growth objective of investors.[21] According to Professor Geltner, the geometric return is a better measure than arithmetic average return of the long-run historical price growth actually achieved (ending value related to beginning value.)[22] Furthermore, the geometric mean is independent of the volatility, while the arithmetic mean is influenced by the volatility.[23]

Explanatory Variables

This paper attempts to use physical and demographic characteristics of each zip code area to predict geometric mean return and volatility in the house price index. These variables are explained below.

Median household income for each zip code is measured here in 1999 dollars.[24] The chart below shows range of income from highest (dark) to lowest (light.)

[19] Hendershott and Hendershott, 2002. p. 41.
[20] Hendershott and Hendershott, 2002. p. 38.
[21] Appreciation return, or capital return [$g = (V_t - V_{t-1}) / V_{t-1}$] and does not include income return, or current yield.
[22] Geltner and Miller, 2001. Chapter 21.
[23] Geltner and Miller, 2001. Chapter 21.
[24] U.S. Census 2000, Summary File 1.



Median Household Income

Population (POP) is measured as gross population including renters and owners of housing. Percent change in population was not considered in this paper, however in a preliminary study, Case (1986) found population growth among key structural determinants of housing return and volatility.[25]

Total employment (TEMPL) is the number of employees whose place of work is located in the specific zip code area.[26] Case, et al. (1996) found that employment growth by sector can affect house price appreciation if the demand for proximity to that amenity changes over time, or if the income level of workers (residents) employed in specific sectors changes over time.[27]

Distance to the central business district is considered here as radius distances from city hall located in downtown Boston, zip code area 02108. The sample

[25] Case, 1986. p. 41.
[26] 1997 Economic Census.
[27] Case and Mayer, 1996. p. 403.

used in the final regression (Model III) includes zip code areas within a fifteen-mile radius of this point.

Comparative analysis was performed for samples in which zip code areas were limited to ten and twenty miles from Boston respectively (Model II was used for this analysis.) Samples limited to zip code areas located close to Boston exhibit greater standard errors for standard deviation (STDEV) and geometric mean return (GEOMEAN) when compared to samples that included zip code areas located further from Boston.

Samples limited to zip code areas located close to Boston show less explanatory power (Adj. R^2) for geometric mean return (GEOMEAN) when compared to samples that included zip code areas located further from Boston. Samples limited to zip code areas located close to Boston show greater explanatory power (Adj. R^2) for standard deviation (STDEV) when compared to samples that included zip code areas located further from Boston.

Employment per resident (EPOP) is total employment (TEMPL) divided by total population (POP.) This variable controls for correlation between total population and total number of employees for a specific zip code area.

Density (DENS) is total population (POP) divided by land area (AREA.) Density is included in the analysis as a proxy for land availability. Land is factor of supply affecting house prices, and here it is assumed that more dense zip codes likely have less land available for development.

In order to select variables for Model III, high correlations between density and distance variables were reviewed as possible source of colinearity in results. To review this issue, a model using a full set of variables was compared to models excluding density and distance variables separately. Results of these models showed consistency in signs and magnitudes of coefficients, standard errors, and t-stats between all combinations of variables.

Land area (AREA) is included here as a measure of market share. Zip code areas with a large land area have a relatively large market share.

The number single-family detached houses (STOCK) in each zip code was obtained from the Census 2000 Summary File, "Units in structure, One, Detached." This variable is included in the analysis to identify zip code areas that might show extreme volatily because of a small number of repeat transactions. This analysis is elaborated in Appendix B.

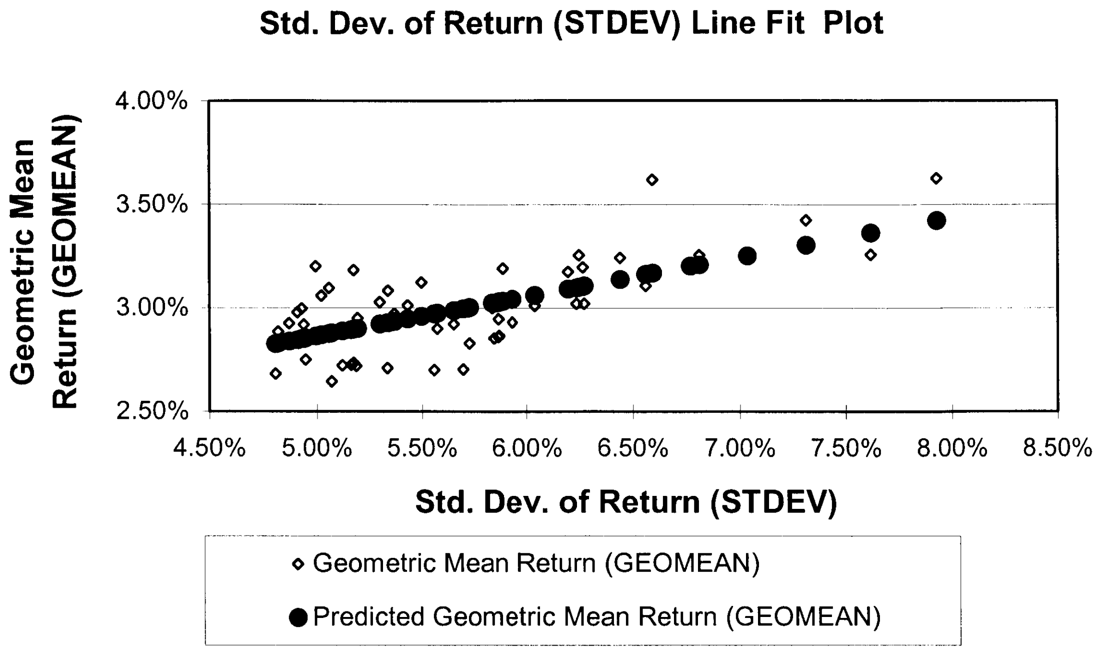
Models

Model I

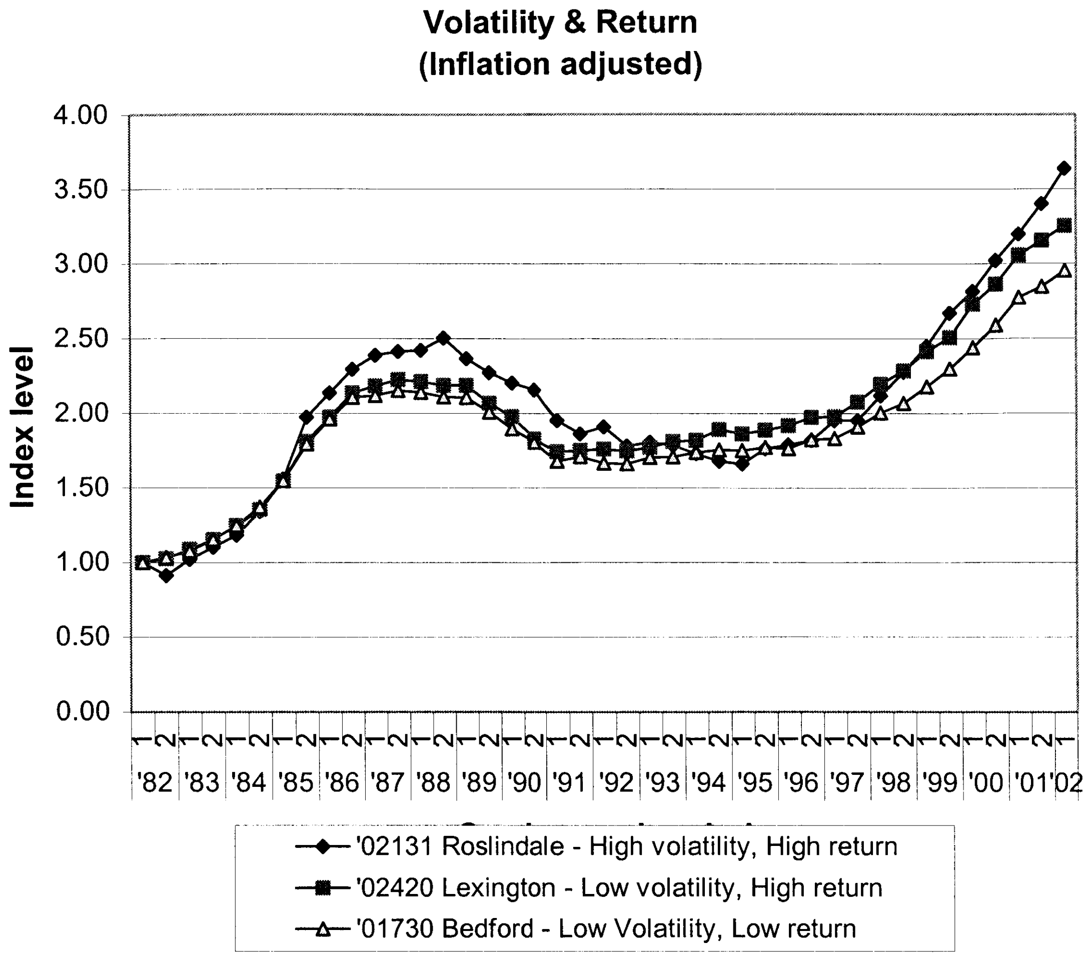
This model uses the following equation to test the relationship of geometric mean return (GEOMEAN) and standard deviation (STDEV):

$$[(\text{GEOMEAN}) = \alpha + x_1(\text{STDEV})]$$

Comparative statistics indicate that geometric mean return and standard deviation in this sample have a positive correlation of 72%. That is, towns with higher standard deviation (volatility) exhibit higher geometric mean return (see graph below.)

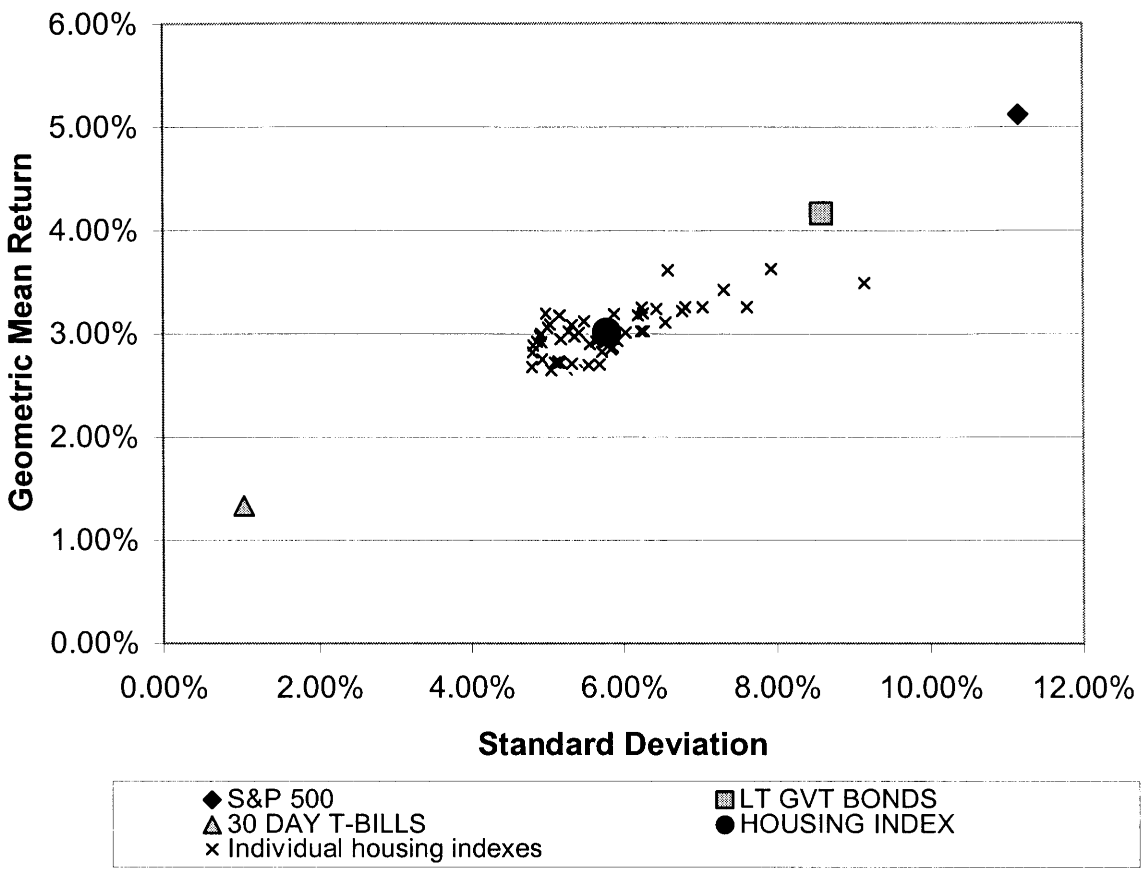


This relationship of standard deviation and geometric mean return is illustrated for selected zip code areas (see chart below.) In this chart, we see that Roslindale exhibits high volatility and high geometric mean return, while Bedford exhibits low volatility and low geometric mean return. Additionally, we see that Lexington exhibits relatively low volatility and relatively high return (we will see in Model III that the high median income in Lincoln sets this zip code area outside of the typical relationship of return and volatility.)



Given the close relationship of standard deviation and geometric mean, one must ask if these results imply an asset pricing model for real estate. This analysis finds that the arithmetic average of volatility and geometric mean return for the entire sample of zip code area housing indexes generally corresponds to the relationship of volatility and geometric mean return observed over the same time period in T-bills, long-term government bonds, and the Standard & Poor's stock market index (all biannual returns.) That is, the asset classes as a whole seem lie along a consistent security market line (see chart below.)

Major Asset Categories



At the macro level, however, this relationship does not seem to hold. The zip code area sample used in this paper exhibits a different relationship between standard deviation and geometric mean return within the sample as compared with asset classes considered as a whole. This seems to suggest that a different 'price of risk' or a different model than that which influences broader security prices governs the housing index in this sample.

Model II

This model uses the following equations to test the relationship of each dependant variable to explanatory variables including total employment (TEMPL,) population (POP,) median household income (INC,) and distance from the Boston's central business district (DIST):

$$[(STDEV) = \alpha + x_1(TEMPL) + x_2(POP) + x_3(INC) + x_4(DIST)]$$

In this equation, standard deviation (STDEV) has a negative relationship to distance from Boston (DIST.) In other words, volatility in the index is greater in zip code areas located close to Boston. Standard deviation (STDEV) has a negative relationship to median household income (INC.) That is, holding distance from Boston constant, volatility in the index is greater in zip code areas with lower median household incomes. One possible explanation of this result could be that wealthy homeowners are less likely to sell their house during a bust period thereby sustaining values through troughs in the market. Total employment (TEMPL) and population (POP) variables are not significant in this equation.

$$[(GEOMEAN) = \alpha + x_1(TEMPL) + x_2(POP) + x_3(INC) + x_4(DIST)]$$

In this equation, geometric mean return (GEOMEAN) has negative relationship to distance from Boston (DIST.) In other words, appreciation in index is greater in zip codes close to Boston. Total employment (TEMPL,) total population (POP,) and median household income (INC) variable are not significant in this equation.

In the following equation, total employment per resident (EPOP) is substituted for population (POP) variable:

$$[(STDEV) = \alpha + x_1(EPOP) + x_2(POP) + x_3(INC) + x_4(DIST)]$$

Total employment per resident (EPOP) was not found to be significant in this equation. This variable could be a proxy for tax base per resident. As a static measure however, this tax revenue is initially capitalized into individual house prices and would not be reflected in the growth of house prices, unless the amount or type of employment changed through time.

Model III

This model uses the following equations to test the relationship of each dependant variable to explanatory variables including total employment per resident (EPOP,) population (POP,) median household income (INC,) distance from the Boston's central business district (DIST,) density (DENS,) and land area (AREA):

$$[(\ln STDEV) = \alpha + x_1(EPOP) + x_2(POP) + x_3(INC) + x_4(DIST) + x_5(DENS) + x_6(AREA)]$$

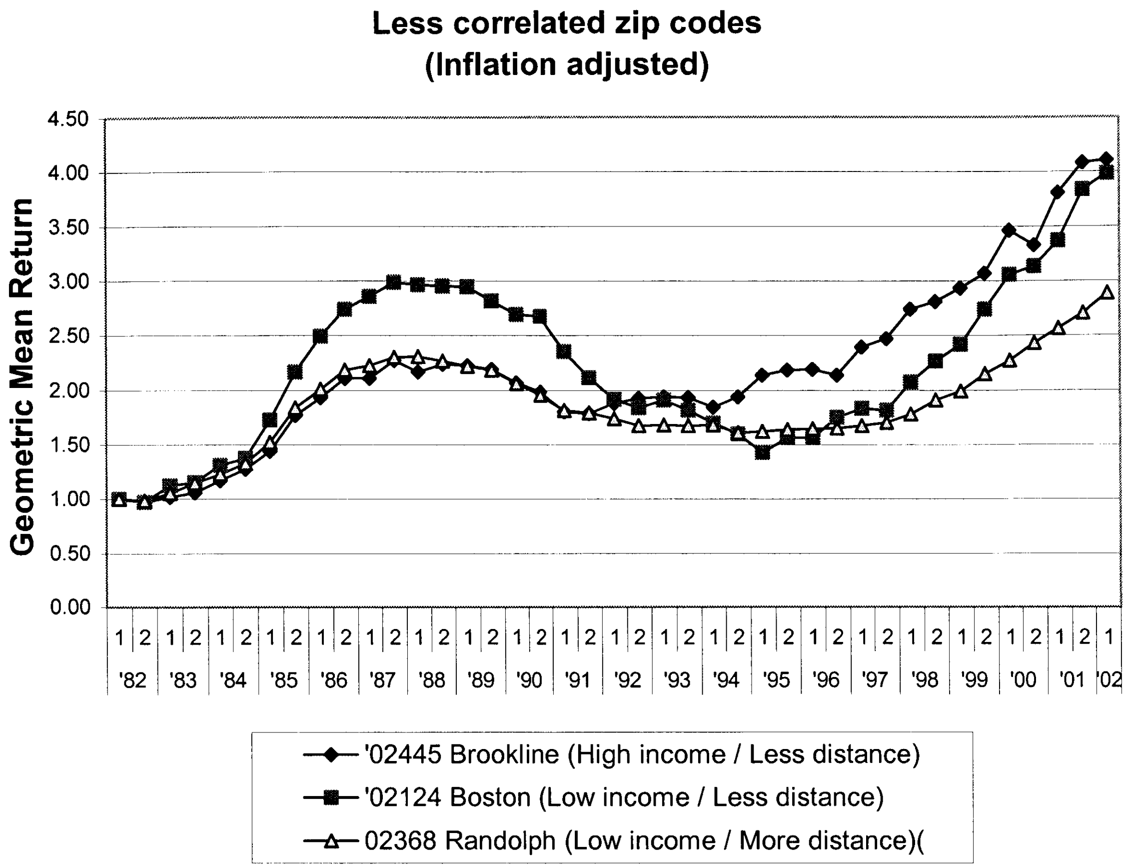
In this equation, the natural log of standard deviation (lnSTDEV) has a negative relationship to median income (INC.) As in Model II, holding distance constant, volatility in the index is greater in zip code areas with lower median income. The natural log of standard deviation (lnSTDEV) has positive relationship to density (DENS.) Volatility in the index is greater in zip code areas with higher density. One possible explanation could be that greater lengths of time required for permitting and constructing housing in more dense zip code areas could lead supply of new housing to overshoot demand, resulting in more volatility in the index. The natural log of standard deviation (lnSTDEV) has positive relationship to land area (AREA.) That is, volatility in index is greater in zip code areas with more land area.

The employment per resident (EPOP) and distance to Boston's central business district (DIST) variables are not significant in this equation. The distance from Boston is almost significant, and this result may be the consequence of high correlations between distance and density variables.

$$[(\ln GEOMEAN) = \alpha + x_1(EPOP) + x_2(POP) + x_3(INC) + x_4(DIST) + x_5(DENS) + x_6(AREA)]$$

In this equation, the natural log of geometric mean return (lnGEOMEAN) has a positive relationship to median income (INC.) In other words, holding distance constant, appreciation in the index is greater in zip code areas with higher median household incomes. The natural log of geometric mean return (lnGEOMEAN) has a negative relationship to distance from Boston (DIST.) that is, appreciation in the index is greater in zip code areas located close to Boston. The natural log of geometric mean return (lnGEOMEAN) has a positive relationship to density (DENS.) Holding all other variables (including distance) constant, appreciation in index is greater in zip code areas with higher density. This result is not surprising since higher density could represent a supply constrained market where we would expect to see higher levels of appreciation.

One question that has come through working with these indexes is how closely house prices in zip individual codes move together. Or, in other words, does a 'rising tide raise all boats?' To answer this question, correlations between inflation adjusted index values were reviewed to reveal differences in price movements among respective zip codes. Selected zip codes with the least significant correlations are shown on the chart below.



This chart again illustrates the simple relationships described by Model III. Holding distance constant at a point close to Boston, zip code areas with higher median household incomes (Brookline) exhibit less overall volatility and greater geometric mean return, or total appreciation, than lower income areas (Boston.) Holding median incomes at a constant level (here Randolph and Boston both have relatively low median incomes,) zip code areas located further from Boston (Randolph) exhibit lower volatility and lower geometric mean return than zip code areas located close to Boston (Boston.)

One must now ask, what is the value of this information in making investment decisions. Perhaps the most striking result in this study is that the median household income (INC) variable (and only this variable) changes signs when used to predict geometric mean return and standard deviation. Theoretically, this relationship represents arbitrage opportunity since higher income areas have both higher returns and lower standard deviation. Furthermore, the inputs to this model change slowly over time if at all.

The extent to which investors (and homeowners) may profit from this information however, depends upon the efficiency of the market for single-family detached dwellings in which these assets trade. Indeed, there are many indications that this market is not an efficient one. Investors in single-family homes must purchase whole assets, and may not be able to capitalize on arbitrage mentioned above if the price of a home in a wealthy zip code is not within their budget. Investors in single-family home assets are typically residents, and residents may not be able to relocate to take advantage of this opportunity. Finally, investors (homeowners) may be willing to exchange growth in the value of their single-family home asset for other forms of utility such as proximity to work or recreation.

The extent to which investors can use this information to achieve higher returns also depends upon the continuity of this model. Since the dependant variables are very closely related to distance from the urban core, an increasing 'desire for centrality' over the study period may have resulted in higher returns in areas located close to Boston than those located further from Boston. Clearly, the sustainability of this particular model depends upon continued growth of metropolitan Boston as an employment, residential, and recreational hub.

Conclusion

Using a repeat sale index of single-family house prices compiled by Case Schiller Weiss, this analysis finds a clear relationship between standard deviation and geometric mean return. With relatively high correlation (72%,) zip code areas exhibiting higher volatility also exhibit higher geometric mean returns. Using aggregate measures, this analysis finds consistent relationships between volatility and return among this sample of house prices and other asset classes. When disaggregated however, there appears to be a different relationship between volatility and return among individual zip code areas and other asset classes.

Explanatory variables including distance, density, and income predict approximately 65% of volatility and geometric mean return for individual zip code areas. Zip code areas with high density and those located close to Boston exhibit higher return and higher volatility than their low density/more distant counterparts. Although distance and density variables were found to be independently significant, the close relationship of these variables precludes total isolation of either variable in the analysis.

Finally, zip code areas with higher median household income exhibit higher return and lower volatility than their lower income counterparts. Although this is perhaps the most striking result of this study, the extent to which investors or homeowners can increase returns using this information is unclear.

Appendix A: Summary Results

	Dependant Variable	Explanatory Variable	Adj. R^2	Coefficients	Standard Error	t Stat
Model I	STDEV	GEOMEAN	51%	0.191	0.026	7.442

	Dependant Variable	Explanatory Variable	Adj. R^2	Coefficients	Standard Error	t Stat
Model II	STDEV	TEMPL	48%	-1.85×10^{-8}	1.18×10^{-7}	-0.157
		POP		1.47×10^{-7}	9.86×10^{-8}	1.488
		INC		-1.23×10^{-7}	4.17×10^{-8}	-2.945
		DIST		-0.001	0.000	-3.571
	GEOMEAN	TEMPL		2.41×10^{-8}	2.88×10^{-8}	0.837
		POP		-2.33×10^{-8}	2.44×10^{-8}	-0.954
		INC		-1.21×10^{-9}	1.04×10^{-8}	-0.116
		DIST		-0.001	7.26×10^{-5}	-9.615

	Dependant Variable	Explanatory Variable	Adj. R^2	Coefficients	Standard Error	t Stat
Model III	lnSTDEV	EPOP	67%	0.004	0.038	0.104
		INC		-2.35×10^{-6}	6.19×10^{-7}	-3.790
		DIST		-0.009	0.006	-1.444
		DENS		2.52×10^{-5}	6.93×10^{-6}	3.629
		AREA		0.010	0.003	3.214
	lnGEOMEAN	EPOP	65%	-0.007	0.020	-0.372
		INC		7.18×10^{-8}	3.27×10^{-7}	0.220
		DIST		-0.015	0.003	-4.601
		DENS		8.30×10^{-6}	3.66×10^{-6}	2.266
		AREA		0.002	0.002	1.103

Notes:
All samples are restricted to zip code areas within 15 miles of downtown Boston.
Model I: This sample uses zip codes restricted to zip code areas with greater than 2000 single family detached units.
Model II: This smaple is not restricted in number of single-family detached units.
Model III: This sample uses zip codes restricted to zip code areas with greater than 2000 single family detached units. Natural log of dependant variables are used in Model III. Coefficients in these regressions are therefore percent change in dependant variable per unit change in explanatory variable.

Appendix B: Data

Note: The following data is used in Models I & III only. Model II includes additional zip codes that have less than 2,000 single-family units (STOCK.)

Zip Code	City	Std. Dev. of Return (STDEV)	Natural Log of STDEV (lnSTDEV)	Geometric Mean Return (GEOMEAN)	Natural Log of GEOMEAN (lnGEOMEAN)
01730	Bedford	4.80%	-303.6%	2.68%	-361.9%
01803	Burlington	4.94%	-300.7%	2.75%	-359.3%
01867	Reading	5.16%	-296.4%	2.72%	-360.3%
01880	Wakefield	5.72%	-286.2%	2.99%	-351.1%
01887	Wilmington	5.17%	-296.2%	2.74%	-359.9%
01890	Winchester	5.37%	-292.5%	2.97%	-351.6%
01904	Lynn	5.86%	-283.7%	2.86%	-355.4%
01905	Lynn	6.81%	-268.7%	3.26%	-342.4%
01906	Saugus	5.65%	-287.3%	2.92%	-353.2%
01907	Swampscott	5.73%	-286.0%	2.83%	-356.5%
01940	Lynnfield	5.19%	-295.9%	2.72%	-360.5%
01945	Marblehead	5.33%	-293.1%	2.71%	-360.8%
01960	Peabody	5.87%	-283.6%	2.87%	-355.2%
01970	Salem	5.84%	-284.0%	2.85%	-355.7%
02021	Canton	6.56%	-272.4%	3.11%	-347.1%
02026	Dedham	5.88%	-283.3%	3.19%	-344.4%
02043	Hingham	6.04%	-280.7%	3.01%	-350.2%
02045	Hull	6.25%	-277.3%	3.26%	-342.5%
02062	Norwood	5.07%	-298.2%	2.65%	-363.2%
02090	Westwood	5.12%	-297.2%	2.72%	-360.4%
02124	Boston	9.14%	-239.3%	3.49%	-335.5%
02126	Mattapan	7.93%	-253.5%	3.63%	-331.7%
02131	Roslindale	7.04%	-265.4%	3.26%	-342.3%
02132	West Roxbury	5.43%	-291.3%	3.01%	-350.3%
02136	Hyde Park	6.19%	-278.2%	3.18%	-344.9%
02148	Malden	6.44%	-274.3%	3.24%	-342.8%
02149	Everett	7.31%	-261.5%	3.43%	-337.4%
02155	Medford	6.26%	-277.0%	3.20%	-344.3%
02169	Quincy - 1	6.27%	-277.0%	3.02%	-349.9%
02170	Quincy - 2	6.77%	-269.3%	3.22%	-343.6%
02171	Quincy - 3	7.62%	-257.5%	3.26%	-342.4%
02176	Melrose	5.83%	-284.2%	3.00%	-350.6%
02180	Stoneham	5.57%	-288.8%	2.90%	-354.0%
02184	Braintree	5.86%	-283.7%	2.95%	-352.4%
02186	Milton	6.23%	-277.5%	3.02%	-349.9%
02188	Weymouth	5.93%	-282.5%	2.93%	-352.9%
02343	Holbrook	5.56%	-289.0%	2.70%	-361.2%

Appendix B: Data (cont)

Zip Code	City	Std. Dev. of Return (STDEV)	Natural Log of STDEV (lnSTDEV)	Geometric Mean Return (GEOMEAN)	Natural Log of GEOMEAN (lnGEOMEAN)
02368	Randolph	5.70%	-286.6%	2.70%	-361.1%
02420	Lexington	4.93%	-301.0%	3.00%	-350.7%
02421	Lexington	5.02%	-299.2%	3.06%	-348.7%
02445	Brookline	6.59%	-271.9%	3.62%	-331.9%
02453	Waltham	5.06%	-298.5%	3.10%	-347.5%
02459	Newton Center	5.17%	-296.1%	3.18%	-344.8%
02465	West Newton	4.99%	-299.7%	3.20%	-344.2%
02472	Watertown	5.50%	-290.1%	3.12%	-346.6%
02474	Arlington	5.30%	-293.8%	3.03%	-349.8%
02476	Arlington	5.19%	-295.8%	2.95%	-352.3%
02478	Belmont	5.34%	-293.1%	3.08%	-347.9%
02481	Wellesley Hills	4.91%	-301.5%	2.98%	-351.4%
02482	Wellesley	4.87%	-302.2%	2.92%	-353.2%
02492	Needham	4.82%	-303.3%	2.89%	-354.5%
02493	Weston	4.81%	-303.5%	2.82%	-356.8%
02494	Needham Hts	4.93%	-300.9%	2.92%	-353.4%

Appendix B: Data (cont)

City	Emp. per Resident (EPOP)	Total Emp. (TEMPL)	Total Population (POP)	Median Household income in 1999 (INC)	Dist to 02108 (DIST)	Density (DENS) (POP/AREA)	Land Area (AREA)	Units in Structure, One, Detached (STOCK)
Bedford	1.65	20,715	12,566	87,962	14.7	998	12.6	3,461
Burlington	1.46	33,285	22,876	75,240	12.3	1,935	11.8	6,488
Reading	0.32	7,564	23,708	77,059	12.5	2,373	10.0	6,553
Wakefield	0.51	12,532	24,804	66,117	10.1	3,148	7.9	6,227
Wilmington	1.03	21,876	21,335	70,741	14.9	1,274	16.7	6,353
Winchester	0.31	6,393	20,810	94,049	7.8	3,308	6.3	5,569
Lynn	0.20	3,690	18,052	56,284	10.6	5,373	3.4	4,849
Lynn	0.37	8,601	23,410	37,099	9.2	4,226	5.5	2,987
Saugus	0.42	10,985	26,041	55,364	8.1	2,247	11.6	7,002
Swampscott	0.19	2,754	14,412	71,089	11.5	4,664	3.1	3,493
Lynnfield	0.39	4,470	11,542	80,626	12.4	1,100	10.5	3,705
Marblehead	0.21	4,202	20,377	73,968	14.2	4,600	4.4	6,082
Peabody	0.48	23,287	48,129	54,829	13.1	2,863	16.8	10,959
Salem	0.41	16,434	40,407	44,033	13.3	4,928	8.2	4,915
Canton	0.96	19,859	20,614	69,260	12.9	1,053	19.6	5,210
Dedham	0.56	13,272	23,527	61,631	9.6	2,193	10.7	6,747
Hingham	0.51	10,042	19,882	83,018	13.4	870	22.9	6,116
Hull	0.07	815	11,050	52,377	10.8	3,708	3.0	3,879
Norwood	0.99	28,329	28,587	58,421	13.5	2,710	10.6	6,138
Westwood	0.81	11,371	14,100	87,638	12.0	1,279	11.0	4,342
Boston	0.12	6,258	51,344	36,025	4.8	14,382	3.6	2,971
Mattapan	0.07	1,947	28,147	38,581	5.9	12,400	2.3	2,643
Roslindale	0.15	4,818	32,404	46,722	5.9	10,222	3.2	3,948
West Roxbury	0.26	6,089	23,826	57,179	7.3	4,833	4.9	5,502
Hyde Park	0.13	3,807	28,337	41,144	7.8	6,828	4.2	4,293
Malden	0.33	18,463	56,340	45,654	5.1	10,961	5.1	6,032
Everett	0.27	10,386	38,037	40,661	3.5	10,962	3.5	2,923
Medford	0.30	17,108	56,185	52,512	5.1	6,571	8.6	8,085
Quincy - 1	0.51	26,322	52,117	45,056	8.4	4,275	12.2	7,745
Quincy - 2	0.11	2,219	19,327	51,607	6.7	9,160	2.1	3,772
Quincy - 3	1.09	18,021	16,581	50,999	5.2	5,545	3.0	2,761
Melrose	0.23	6,277	27,134	62,811	6.8	5,724	4.7	6,423
Stoneham	0.31	6,790	22,219	56,605	8.3	3,311	6.7	5,013
Braintree	0.88	29,783	33,828	61,790	11.1	2,351	14.4	9,153
Milton	0.22	5,779	26,062	78,985	8.0	1,952	13.4	7,209
Weymouth	0.29	4,258	14,789	48,212	11.7	3,871	3.8	3,286
Holbrook	0.28	2,970	10,682	54,252	14.8	1,500	7.1	2,981

Appendix B: Data (cont)

City	Emp. per Resident (EPOP)	Total Emp. (TEMPL)	Total Population (POP)	Median Household income in 1999 (INC)	Dist to 02108 (DIST)	Density (DENS) (POP/AREA)	Land Area (AREA)	Units in Structure, One, Detached (STOCK)
Randolph	0.29	9,173	31,227	55,358	12.5	2,913	10.7	7,091
Lexington	0.42	5,864	14,126	92,911	10.3	2,153	6.6	4,202
Lexington	0.51	8,328	16,239	100,358	10.6	1,584	10.3	4,833
Brookline	0.19	5,012	26,259	73,197	4.1	10,100	2.6	3,409
Waltham	0.39	11,405	29,146	50,189	9.1	7,650	3.8	3,337
Newton Center	0.29	5,182	17,893	99,076	7.2	3,495	5.1	4,733
West Newton	0.14	1,538	10,995	82,301	8.3	5,114	2.2	2,734
Watertown	0.26	8,413	32,986	59,764	5.8	7,910	4.2	3,293
Arlington	0.08	1,945	25,682	62,846	6.4	7,902	3.3	4,314
Arlington	0.14	2,269	16,697	66,631	6.9	7,521	2.2	3,542
Belmont	0.12	3,001	24,173	80,300	6.4	5,100	4.7	4,651
Wellesley Hills	0.56	7,937	14,128	123,622	11.1	2,564	5.5	4,408
Wellesley	0.24	3,033	12,485	105,738	12.7	2,527	4.9	2,979
Needham	0.16	3,092	19,697	93,342	10.7	2,031	9.7	5,955
Weston	0.18	2,080	11,469	153,918	12.0	659	17.4	3,407
Needham Hts	0.99	9,117	9,214	79,169	9.4	3,123	3.0	2,378

Appendix C: Tests of Volatility in Index

Distribution of Data

This section reviews tests performed on standard deviation and geometric mean return for zip code areas that exhibit high volatility.

High volatility in certain zip codes could be caused by random error in zip codes with a small amount repeat transactions. Since data on the number of transactions included in each zip code area index was not available, the amount of single family detached dwellings (STOCK) in each zip code area is used to identify indexes that might be formulated from a small number of repeat transactions.

Zip code areas with a small amount of single-family detached dwellings (STOCK) however, are not highly correlated with zip codes with high volatility (STDEV) [correlation (STOCK, STDEV): -12%.] Zip code areas with a small amount of single-family detached dwellings (STOCK) are only slightly correlated with zip codes located close to the city (DIST) where volatilities are the greatest [correlation (STOCK, DIST): 30%] [correlation (DIST, STDEV): -57%]

Moreover, zip codes with high volatility also exhibit high geometric mean return [correlation (STDEV, GEOMEAN): 72%.] It is not likely that high geometric mean return would result from small sample error. For this to situation to take place, houses would necessarily transact only during boom times. Therefore, this study concludes that high volatility in certain zip code area indexes is not the result of random error caused by a small number of transacting properties in these indexes.

Additional tests restricting sample size

This section compares results of models using a full sample of zip code areas (unrestricted with respect to number of single-family units) with modified samples of zip code areas comprising no less than 2,000 and 3,000 units of single-family detached dwellings (STOCK.)

Model III, Modified Sample I

This model restricts the sample of zip code areas to those with more than two thousand single-family detached dwellings (STOCK > 2,000 units.) Using the natural log of standard deviation (lnSTDEV) as the dependant variable, the explanatory power of this model is consistent with the model using the full sample of zip code areas [Adj. R^2 of 63% and 62%, respectively.] Signs and magnitudes of coefficients, standard error, and t-stats remain largely unchanged between these models (see regression results below.)

Using the natural log of geometric mean return (lnGEOMEAN) as the dependant variable, the explanatory power of model using Modified Sample I is less than the model using the full sample [Adj. R2 of 65% and 71%, respectively.] Signs and magnitudes of coefficients, standard error, and t-stats remain largely unchanged between these models (exception: Median Household Income (INC) is no longer significant.) The result of this comparison for geometric mean return (GEOMEAN) however does not cause concern since the geometric mean return is not affected by volatility.

In the base case for Model III, the Modified Sample I is used. This sample is restricted to zip codes with no less than 2,000 single-family units. (See Appendix A, Model III for results.)

Model III, Modified Sample II

This model restricts the sample of zip code areas to those with no less than three thousand single-family detached dwellings (STOCK > 3,000 units.) Although the explanatory power of this model was reduced for dependant variables standard deviation and geometric mean return (over one-third of sample was removed,) the results using Modified Sample II were largely consistent with the results using the full sample and the Modified Sample I.

Summary results: Full Sample & Modified Sample II

See Appendix A – Summary results for Modified Sample II (Base Case.)

Model	Dependant Variable	Explanatory Variable	Adj. R^2	Coefficients	Standard Error	t Stat
Model III - Full Sample	STDEV	EPOP	65%	-0.035	0.035	-0.985
		INC		-1.80×10^{-6}	5.10×10^{-7}	-3.529
		DIST		-0.008	0.006	-1.260
		DENS		2.32×10^{-5}	5.33×10^{-6}	4.354
		AREA		0.012	0.003	3.937
	GEOMEAN	EPOP	74%	-0.013	0.018	-0.716
		INC		5.19×10^{-7}	2.59×10^{-7}	2.002
		DIST		-0.014	0.003	-4.408
		DENS		1.15×10^{-5}	2.7×10^{-6}	4.234
		AREA		0.003	0.002	1.777

Model	Dependant Variable	Explanatory Variable	Adj. R^2	Coefficients	Standard Error	t Stat
Model III - Modified Sample II	STDEV	EPOP	51%	-0.045	0.037	-1.193
		INC		-1.90×10^{-6}	5.72×10^{-7}	-3.336
		DIST		-0.006	0.006	-1.004
		DENS		1.90×10^{-5}	7.67×10^{-6}	2.481
		AREA		0.010	0.003	3.326
	GEOMEAN	EPOP	55%	-0.009	0.023	-0.385
		INC		1.10×10^{-7}	3.56×10^{-7}	0.308
		DIST		-0.015	0.004	-4.132
		DENS		6.45×10^{-6}	4.78×10^{-6}	1.349
		AREA		0.002	0.002	1.235

Note:
All samples are restricted to zip code areas within 15 miles of downtown Boston.

Bibliography

Abraham, Jesse M., Goetzmann, William M., and Wachter, Susan M., 1994. Homogeneous Groupings of Metropolitan Housing Market. Journal of Housing Economics **3**, pp 186-206, 1994.

Abraham, Jesse M., Hendershott, Patric H., 1994. Bubbles in Metropolitan Housing Markets. Journal of Housing Research, vol. 7, no. 2, 1996, pp.191-207.

Adair, Alastair, Berry, Jim, McGreal, Stanley, Deddis, Bill, Hirst, Suzanne, 1999. Evaluation of Investor Behaviour in Urban Regeneration. Carfax Publishing Company.

Bourassa, Steven C., Hoeslib, Martin, and Peng, Vincent S., 2002. Do housing sub-markets really matter?

Case, Bradford, Pollakowski, Henry O., and Wachter, Susan M., 1991. On Choosing Among House Price Index Methodologies. Journal of the American Real Estate and Urban Economics Association 19, no. 3.

Case, Karl E. and Mayer, Christopher J., 1996. Housing Price Dynamics Within a Metropolitan Area. *Regional Science and Urban Economics* 26 (under permission, *Elsevier Science.*)

Case, Karl E., 1986. The Market for Single-Family Homes in the Boston Area. New England Economic Review, May/June.

Census 2000, Summary File 1 (SF1), Census of Population and Housing, Massachusetts.

DiPasquale, Denise and Somerville, C. Tsuriel, 1995. Do House Price Indices Based on Transacting Units Represent the Entire Stock? Evidence from the American Housing Survey. *Journal of Housing Economics* 4, 1995.

Geltner, David and Miller, Norman G., 2001. Commercial Real Estate Analysis and Investments. South-Western Publishing, MA.

Hendershott, Patric H., Hendershott, Robert J. 2002. On measuring Real Estate Risk. Real Estate Finance. Boston: Winter 2002, Vol. 18, Iss. 4; pp. 35-41.

Mayer, C.J., 1993. Taxes, Income Distribution, and the Real Estate Cycle: Why All Houses do not Appreciate at the same rate. New England Economic Review, May-June, pp. 39-50.

Office of Federal Housing Enterprise Oversight, First Quarter 2003 Report.

Poterba, J.M., 1991. House Price Dynamics: the Role of Tax Policy and Demography,. Brookings Papers on Economic Activity 2, pp. 143-183.

Smith, B.A. and W.P. Tesarek, 1991. House Prices and Regional Real Estate Cycles: Market Adjustments in Houston. Journal of the American Real Estate and Urban Economics Association 19, no. 3, pp. 396-416.

U.S. Census Bureau, ECON^{97} Report Series, Economic Census. U.S. Department of Commerce, Economics and Statistics Administration, 1997.