Walk Score and Multifamily Default: The Significance of 8 and 80

Gary Pivo, University of Arizona [1]

October 1, 2013

The views expressed in this research paper reflects the personal views of the author, and do not necessarily reflect the views or policies of any other person, including Fannie Mae or its Conservator. Any figures or estimates included in a paper are solely the responsibility of the author.

Introduction

This paper examines the relationship between Walk Score, a widely available indicator of walkability, and mortgage default risk in multifamily rental housing. It shows that very high and very low Walk Scores significantly affect default risk. Where Walk Score is 80 or more (out of 100) the relative risk of default is 60% lower than where Walk Score is less than 80. Where Walk Score is 8 or less, default risk is 121% higher. This was found while controlling for building age and condition, market setting, loan terms, and other factors that impact risk. A Walk Score above 80 indicates a neighborhood is highly walkable, while a score below 8 indicates it is highly auto dependent.

This is the first paper showing that Walk Score affects default risk in multifamily rental housing. It builds on prior work showing that higher Walk Scores are related to lower default risk in single family housing (Rauterkus and Miller 2011) and higher values in office, retail, and apartments buildings (Pivo and Fisher 2011, Kok et al. 2012, Kok and Jennen 2012). For lenders and developers, the findings reported here suggest that Walk Score could be used to help evaluate and underwrite properties and investment risk. For researchers in real estate and urban economics, the findings deepen our knowledge of investment risk correlates and the role of local accessibility in urban economic geography. And for practicing urban planners, developers, policy-makers and others interested in fostering healthier, more sustainable cities, it strengthens the case for walkable urban development.

Background

Walkability is the degree to which an area within walking distance of a property encourages walking trips for functional and recreational purposes (Pivo and Fisher 2011). Several physical and social attributes of an area can affect walkability including street connectivity, traffic volumes, sidewalk width and continuity, topography, block size, safety and aesthetics (Frank and Pivo 1994, Hoehner et al. 2005,

---

[1] Dr. Pivo is Professor of Urban Planning and Professor of Renewable Natural Resources at the University of Arizona, Tucson, Arizona. This paper is incidental to, and leverages the findings of, a consulting engagement performed by Hoyt Advisory Services (HAS) for Fannie Mae. The author wishes to thank Fannie Mae and HAS for their support and comments on earlier drafts. He also is grateful to Dr. Jeffrey Fisher and Dr. Grant Thrall for their help with data. As always, the author is solely responsible for the content, which does not necessarily reflect the opinions of Fannie Mae or HAS.

Cao, Handy and Mokhtarian 2006, Lee and Moudon 2006, Parks and Schofer 2006, Freeman et al. 2012). However, research indicates that the presence of desired destinations, such as stores, parks and transit stops, is the most significant driver of walkability (Hoehner et al. 2005, Lee and Moudon 2006, Sugiyama et al. 2012). Handy (1993) refers to this dimension of urban space as “local accessibility”. More than 30 years ago, Li and Brown (1980) noted that local accessibility was an important aspect of overall accessibility in urban areas even though accessibility was more commonly measured in relation to urban centers.

Local accessibility is the particular dimension of walkability that is measured by Walk Score, although Walk Score is correlated with other walkability correlates, such as intersection, residential, and retail destination density (Duncan et al. 2011). Studies have shown Walk Score to be a reliable and valid estimator of neighborhood features linked to walking (Carr et al. 2010, Carr et al. 2011, Duncan et al. 2011, and Duncan et al. 2013). It is also a better predictor of walking for non-work trips than other related indices (Manaugh and El-Geneidy 2011).

Walk Score rates the walkability of an address by determining the distance from a location to educational (schools), retail (groceries, books, clothes, hardware, drugs, music), food (coffee shops, restaurants, bars), recreational (parks, libraries, fitness centers) and entertainment (movie theaters) destinations. Points are assigned to the location based on distance to the nearest destination of each type. If the closest establishment of a certain type is within a quarter mile, Walk Score assigns the maximum points for that type. No points are given for destinations beyond a mile. Each type of destination is weighted equally. Points for each category are summed and scores are normalized to produce a total from 0 to 100. Pivo and Fisher (2011) discuss some of the limitations and other caveats related to Walk Score. A newer version that addresses certain concerns is currently in development.

Walk Score has certain advantages over other systems for measuring walkability (Moudon and Lee 2003, Parks and Schofer 2006). One advantage is that it measures the best predictor of walking, proximity to desired destinations. Another is that it is available for all addresses nationwide. Weidema and Wesnæs (1996) developed data quality indicators including reliability, completeness, temporal and geographical correlation with the time and place being assessed, and further technical correlation, including whether the data actually represent the process of concern. Walk Score scores well on such metrics.

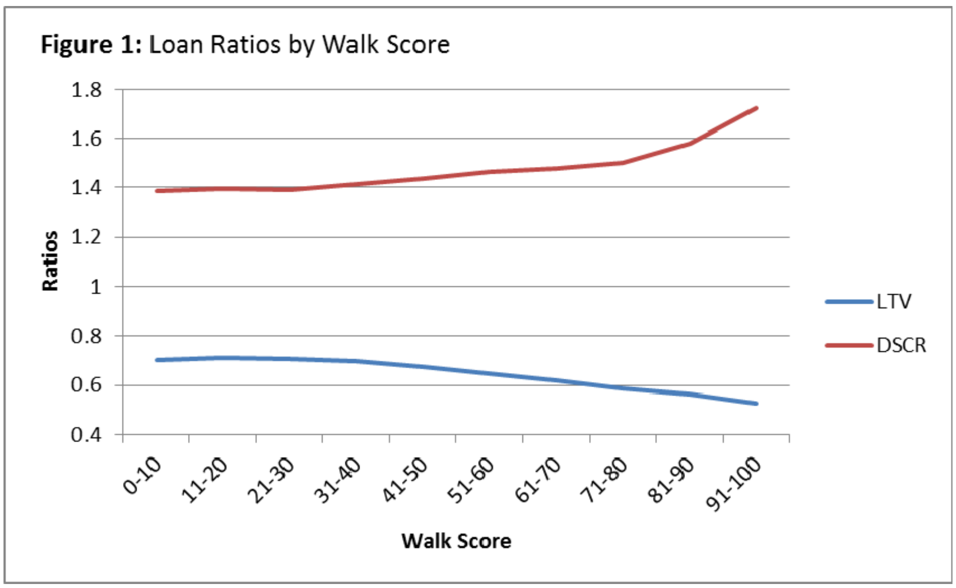
Increasing urban walkability is increasingly viewed as a major goal by urban planners, sustainability scientists and public health experts for social and environmental reasons. The expected benefits remain an ongoing research topic, though a considerable body of evidence is emerging from well-controlled studies. Environmental benefits may include less air pollution, auto use and gasoline consumption (Frank, Stone and Bachman 2000, Ewing and Cervero 2001, Frank and Engelke 2005, Handy, Cao and Mokhtarian 2005, Cao, Handy and Mokhtarian 2006). In fact walking has been recognized as one of the main options for mitigating climate change in the transport sector (Chapman 2007, Bosch and Metz 2011). Social benefits may include better public health as a result of more physical activity (Lee and Buchner 2008, World Cancer Research Fund/American Institute for Cancer Research 2009, Berrigan et al. 2012) and increased social capital including more community cohesion, political participation, trust, and social activity (Leyden 2003, du Toit et al. 2007, Rogers et al. 2009, Wood et al. 2010). Social capital has in turn been linked to the capacity of cities to transition toward greater sustainability (Portney 2005, Geels 2012)

Walkability can be created by developing larger scale mixed-use development projects or by infilling development in currently walkable locations. There is evidence that it is more difficult to finance walkable projects because they are perceived to be riskier, leading to more expensive financing. Financiers could be concerned about disamenities from non-residential uses, uncertainty about the performance of mixed use buildings, entitlement risk for infill projects, or weaker economic conditions in walkable, mixed use neighborhoods. One older study focused on residential developments that were planned to be compact, scaled for pedestrians, and designed to include activities of daily living within walking distance of homes (Gyourko and Rybczynski 2000). It found that developers, financiers and investors perceived such projects to be “inherently riskier and more costly…arising from the multiple-use nature of the developments”. On the other hand, the study also found that urban infill risk premiums could be quite small where communities were willing to accept high densities. More recently, Leinberger and Alfonzo (2012) pointed out that “walkable urban places remain complex developments that still carry high risk and, as such, costly capital (both equity and debt financing)”. Of course, not all projects in walkable locations are mixed use or complex and the Urban Land Institute recently reported that “demand and interest in apartments in ‘American infill’ locations remain hot” (PwC and the Urban Land Institute 2012). Thus, while experts have noted that more walkable projects are more difficult to finance because of their riskier reputation, the degree to which this is true for all walkable projects is unclear because they can vary in location, scale and complexity. It is also unclear exactly what it is about the projects that are cause for concern.

According to Grovenstein et al. (2005), mortgage lenders often respond to perceived risk by limiting how much they will lend. They point out that lenders could also increase interest rates on riskier projects, but that approach is constrained because higher rates can increase default risk. Assuming a given cash flow and value, limiting the amount loaned reduces the loan-to-value ratio (LTV) and increases the debt service coverage ratio (DSCR). For borrowers, a lower loan-to-value ratio means that more walkable projects would produce a lower return on equity compared to what could be earned on more conventional projects with higher loan ratios, all else being equal, as long as positive leverage is possible (i.e. when the cost of debt financing is lower than the overall return generated by the property return on assets). A lower return on equity could cause investors to disfavor walkable investments, decrease capital flows to walkable properties and slow the movement toward more walkable cities.

In the pool of nearly 37,000 multifamily mortgages examined in this study (see Methods, below for details), there is evidence that lenders treated projects in more walkable locations as if they were perceived to be riskier loans. As shown in Figure 1, in the study sample, as Walk Score increased, LTV fell and DSCR increased. These trends in LTV and DSCR relative to Walk Score are consistent with lenders reducing the size of loans relative to property value and income in more walkable locations in response to perceived risk.

As suggested above, less favorable loan terms for more walkable locations may not be caused by lenders’ views about walkability per se but rather by concern about other features of the properties or their location such as disamenities, entitlement risk, or economic conditions. This may seem counterintuitive if one simply assumes that places with higher Walk Scores are correlated with more supply constrained markets. It is true that in the sample there was a very weak correlation between higher Walk Score and higher supply constraint as measured by vacancy rates and price change. However, higher Walk Scores were also correlated with more poverty and lower income households in the neighborhood and with smaller loans and building size, all of which can raise the level of expected risk. It goes beyond the scope of this paper to determine precisely why loan terms appear to have been less favorable in more walkable neighborhoods. The reasons, however, probably result from a number of social and economic conditions that distinguish more and less walkable locations. In the modeling presented below, however, the effect of factors beyond Walk Score that may affect default risk are statistically controlled so as to determine how walkability itself affects default risk, all else being equal.



This paper takes a closer look at this risk issue by comparing default risk in more and less walkable properties (i.e. properties in more and less walkable locations). It shows that default risk for multifamily properties in highly walkable neighborhoods is lower, not higher, than the default risk for projects in less walkable locations.

The hypothesis for this paper is as follows:

> *Hypothesis: Greater walkability, as measured by higher Walk Scores, reduces mortgage default risk in multifamily housing.*

Previous studies have shown that walkability improves property values (Pivo and Fisher 2011, Kok et al. 2012, Kok and Jennen 2012, Pivo 2013). The higher values appear to result from both stronger cash flows and lower capitalization rates, suggesting that walkable properties are favored in both the space (i.e., rental) markets and the capital markets (Pivo and Fisher, 2010). This relationship between walkability and value should be expected, given the long known understanding that accessibility, in this case local accessibility, plays in the formation of property value. Pivo and Fisher (2011) discuss this in the context of a recent summary of the literature on the determinants of urban land and property values.

If more sustainable buildings produce better cash flows and property values, then they should also exhibit lower default risk because default risk is inversely related to cash flow and value (Titman and Torous 1989, Kau et al. 1990, Vandell 1984, Vandell 1992, Vandell et al. 1993, Goldberg and Capone 1998, Goldberg and Capone 2002, Archer et al. 2002, Pivo 2013). However, as Pivo (2013) has noted, adding information on walkability to the loan origination process would only be helpful if its impact on cash flow and value was not already fully accounted for in the loan origination process. The assumption here is that the walkability premium was not fully considered in past lending decisions. That is not to say it was completely ignored, just not recognized as important in property markets as it appears to be today. Indeed, loan proposal documents regularly address locational advantages such as access to public transportation and other amenities.

Methods

Logistic regression models were used to test the effects of Walk Score on default risk. Logistic regression models have been used in several prior studies to estimate the effects of explanatory variables on the probability of mortgage default (Vandell et al. 1993, Goldberg and Capone, 1998, Goldberg and Capone 2002, Archer et al. 2002, Ruaterkus et al., 2010). Logistic regression is a statistical method for predicting the value of a bivariate dependent variable (Menard 1995). A bivariate variable is one with two possible values (e.g., in default/not in default in the present study). The value of the dependent variable predicted by a logistic regression model is the probability that a case will fall into the higher of the two categories of the dependent variable, which normally indicates the event (e.g., default) occurred, given the values for the case on the independent variables. In other words, it is the probability that an event will occur under various conditions characterized by the independent variables. The predicted value of the dependent variable is based on observed relationships between it and the independent variable or variables used in the study.

To build logistic regression models for the present study, data were provided by Fannie Mae on all the loans in its multifamily portfolio at the end of Q32011. The sample included mortgages with fixed and adjustable rates and with a wide variety of seasoning, originating anywhere from September, 1971 through September, 2011. In the study, each loan was treated as a separate case or observation. For each case, data were available on the loan age, type, terms, and lender, on various financial, physical, and locational attributes of the property, and on the number of days the loan was delinquent, if any. In addition to these data on the loans, Walk Score data and other data on neighborhood and regional attributes were collected from other sources for use in the model. More details on these variables and those from other sources are discussed below.

Following Archer et al. (2002), cases in the Fannie Mae database with extreme values on certain variables were excluded from the study in order to filter out possible measurement error. The extreme value filters ensured that all loans used in the study had an original note interest rate greater than the 10-year constant maturity risk-free rate at their origination date, an original LTV ratio of 100% or less, an original debt service coverage ratio greater than 0.9 and less than 5.0, and an original note interest rate greater than 3% and less than 15%. After these filters were applied, there were 36,922 loans in the sample out of 42,474 loans originally provided for the study by Fannie Mae (including affordable, student and senior housing).

Variables

Dependent and Explanatory Variables

DEFAULT was the dependent variables. It was binary, indicating whether (1) or not (0) a loan was in default as of Q32011. A loan was classified as being in default if it was delinquent on its payments by 90 days or more as of Q32011. This is an industry standard definition and it matches that used by Archer et al. (2002) who pointed out that such a broad definition is useful because other resolutions in addition to foreclosure can be used to resolve defaults and they all involve delinquency-related costs to the lender.

WALKSCORE was the explanatory variable of interest in the study. It captures the walkability of the area where each apartment building was located. As noted above, it has been found to be a reliable and valid estimator of neighborhood features linked to walking and a better predictor of walking for non-work trips than other similar indices.

Control Variables

The expectation was that WALKSCORE was related to default risk because it affects cash flow and value to a degree that was unaccounted for in the DSCR or LTV ratios at loan origination. However, it could also be the case that WALKSCORE is correlated with other factors that affect financial outcomes, such as other loan, property, neighborhood or macroeconomic variables. In that case WALKSCORE could simply be a proxy for other drivers of cash flow and value, such as neighborhood vacancy rate. Therefore, in order to separate the effects of WALKSCORE on DEFAULT from other possible drivers, several control variables suggested by prior research were used in the models. The controls fall into four groups including loan, property, neighborhood and economic characteristics.

Loan Characteristics

OLTV and ODSCR measured the loan-to-value and debt service coverage ratios at loan origination. These are commonly used to predict default risk. Higher OLTV and lower ODSCR were expected to be associated with greater default risk. LOAN_AGE_MONTHS was the number of months from the loan origination date to the observation date (Q32011). Previous researchers have shown that default risk declines with age, though the pattern is nonlinear, increasing rapidly in the first few years and then declining (Snyderman 1991, Esaki et al. 1999, Archer et al. 2002). The same pattern was observed in this study sample. Consequently, some degree of non-linearity in the logit (i.e., a nonlinear relationship with the logit form of DEFAULT) was detected for LOAN_AGE_MONTHS using the Box-Tidwell transformation (Menard 1995). Transformations of LOAN_AGE_MONTHS were tried in the models but they did not improve the results and were discarded to simplify interpretation of the results. ARM_FLAG was a dummy indicating whether the loan was adjustable (1) or fixed (0).

Property Characteristics

NO_CONCERNS was a dummy indicating whether or not there were no substantial concerns about the property condition at the time of loan origination. This should reduce default risk by decreasing the need to divert cash flow to deferred maintenance. BUILT_YR was the year the property was built. Archer et al. (2002) found that default rates increased with building age, so BUILT_YR was expected to be inversely related to default risk (i.e., older buildings would default more often). This was the expectation for the nation as a whole, although it could be true that in some areas the historic or design qualities associated with older buildings may be desired and that could influence how age is related to default risk by increasing demand, cash flow and value for older buildings. TOT_UNTS_CNT was the total number of units in the property. Smaller properties have been reported to experience more financial distress (Bradley et al. 2000). Perhaps this is because of the characteristics of borrowers on smaller properties who may have less experience, less access to capital and less of a tendency to use professional property managers. Archer et al. (2002), however, looked at unit count in a multivariate analysis and found that size (and value) was unrelated to default, even though their univariate analysis showed that smaller properties had less default risk, contrary to Bradley et al. (2000). So the expected effect in this study was ambiguous.

Neighborhood and City Scale Geographic Characteristics

Researchers have found that stress on properties is related to geographical effects. In fact, Archer et al. (2000) found geographical effects to be one of the most important dimensions for predicting default. Five control variables were created to control for these sorts of effects at the city and neighborhood level. MEDHHINC000 was the median household income in the census tract from the 2000 census. Higher income was expected to be linked with lower default rates. PROP_CRIME_MIL was the annual number of property crimes per million persons at the city scale, reported by local police departments to the US Department of Justice. Higher crime in the city was expected to increase default risk. VACANCY_RATE was the vacancy rate for housing in the census block group as determined by the 2007-11 American Community Survey conducted by the US Census. It was used to control for the effect of housing supply constraint on default rates in order to rule out the possibility that WALKSCORE is a proxy for constrained supply. PRINCIPAL_CITY was a dummy indicating whether the property is located in a Principal City, defined by the US Census as the largest incorporated or census designated place in a Core Based Statistical Area. The goal was to control for whether or not the property was centrally located within a larger metropolitan or micropolitan area because many such areas have outperformed less central, suburban locations in the past decade and Walk Score tends to be higher in central cities. Properties in Principal Cities were expected to have lower default risk. URB_RUR was also used to measure regional centrality. It was based on the 11 Urbanization Summary Groups available from the ESRI Tapestry Segmentation System, which groups locations into an urban-rural continuum from Principal Urban Centers to Small Towns and Rural places. The system also divides each urbanization group into places with higher and lower affluence; however, that element was ignored for URB_RUR.

Finally, TOP25CITY was a dummy variable indicating whether the property was in one of the 25 largest US cities.

Regional and National Economy

Certain regional and national variables were included to control for difference in the national and regional economic conditions faced by properties since loan origination. Nine dummies were created to indicate whether a property was located in each of the nine census divisions. Vandell et al. (1993) used a similar variable. Additional variables designed to capture regional effects were dummies for whether the property was located in New York City (NYC) or Washington, DC (DC), and changes in vacancy rates and prices in the metropolitan area in the most recent six-year period. AVG_PRICE_6 and AVG_OCC_6 were computed using the NCREIF Apartment Index for metropolitan statistical areas. They describe the average increase in apartment prices and the average occupancy rate in the metro area for each property over the last 6 years prior to the study observation date. Prior research updates LTV and DSCR over time on the theory that negative equity or cash flow will trigger default. Both are affected by NOI, which are in turn affected by vacancy rates and rental price indices. Therefore, changes in vacancy rates and rental price indices at the metropolitan scale can be used to capture changes in market conditions that strengthen or weaken mortgages over time (Goldberg and Capone 1998, Goldberg and Capone 2002).

Borrower Characteristics

Lenders consider borrower characteristics to be crucial to reducing default rates. Relevant variables include borrower character, experience, financial strength and credit history. Unfortunately, data on these issues were not available for this study. It is likely, however, that lenders adjusted the original loan terms based in part on their assessment of borrower characteristics. Therefore, OLTV, ODSCR and ARM_FLAG may be proxies for borrower characteristics. TOT_UNTS_CNT may also be correlated with borrower characteristics, as mentioned above. In linear regression, omitted orthogonal variables (i.e., variables that are not correlated with the other independent variables) that are determinants of the dependent variable do not bias the parameter estimates. However, in logistic regression, Cramer (2007) showed that omitted orthogonal variables depress the estimated parameters of the remaining regressors toward zero. Therefore, it is possible that the estimated effects of the sustainability variables on default risk reported below would be even larger if borrower characteristics were included in the analysis.

Collinearity

Correlation among the independent variables is indicative of collinearity. Collinearity can create modeling problems including insignificant variables, unreasonably high coefficients and incorrect coefficient signs (e.g., negatives that should be positive). Collinearity will not affect the accuracy of a model as a whole, but it can produce incorrect results for individual variables. Tolerance statistics, which check for a relationship between each independent variable and all other independent variables, were used as an initial check for collinearity and they raised no concerns (Menard 1995). A pairwise correlation matrix among the independent variables also uncovered no issues.

Results

Univariable analysis

The process of building the logistic regressions began with a univariable analysis of each variable as recommended by Hosemer and Lemeshow (2000). For the dummy and ordinal variables, this was done by using a contingency table to compare outcomes for properties that did and did not default. The significance of the differences was determined with the Likelihood Ratio and Pearson Chi-Square tests. For the continuous variables, means for the default and not-default groups were compared using the two-sample *t*-test.

The results are shown in Table 1 along with descriptive statistics for the total sample. Other than TOP25CITY and a few of the regional dummies, all of the variables, including WALKSCORE, were significantly related to DEFAULT.

Logistic regressions

Following the univariable analysis, several different models were produced. Each one had a specific purpose which is described below. The statistics for each model are given in Table 2. Particular attention was paid to changes in the WALKSCORE coefficients across the various models.

Model 1 included all of the scientifically relevant variables. This allowed the effect of removing insignificant variables on the variables that remained in subsequent models to be observed.

The size and direction of the relationships are indicated by the unstandardized coefficients (B). B gives the change in the risk of default associated with a 1-unit change in the variable while other variables are held constant. If B is positive, then default risk increases with a 1-unit increase in the variable. If B is negative, the relationship is inversed. For example, in Model 1, the B coefficient for WALKSCORE (-0.018) indicates that as WALKSCORE rises, the risk of DEFAULT falls, holding the other variables constant. All of the variables in Model 1 were related to DEFAULT in the expected direction even though some of the relationships were statistically insignificant.

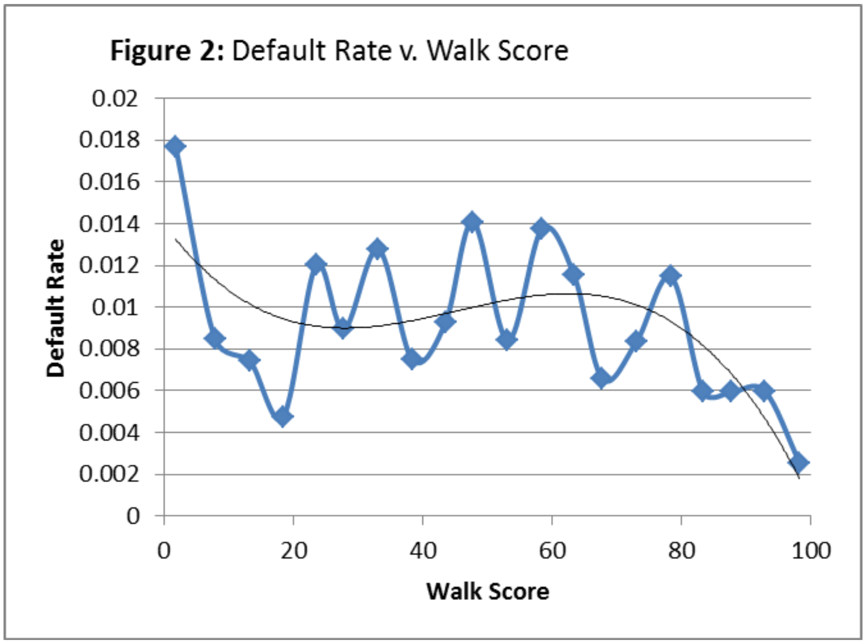
The Exp(B) statistic is the odds ratio or the number by which one would multiply the odds of default for each 1-unit increase in the independent variable. An Exp(B) greater than 1 indicates the odds increase when the independent variable increases and an Exp(B) less than 1 indicates the odds decrease when the independent variable increases. For WALKSCORE in Model 1, Exp(B) indicate that a 1-unit increase resulted in a 1.8% decrease in the odds of default (i.e., the odds of DEFAULT are multiplied by .018, which is .982 less than 1). Odd ratios can also be interpreted as relative risk when the outcome occurs less than 10% of the time, which is the case for DEFAULT in the study sample (Hosmer and Lemeshow 2000). So, we can say that for every 1-unit increase in Walk Score, the relative risk of default declines by 1.8%. If, for example the default rate for properties with a particular Walk Score was 0.9% (the mean for the sample), then according to Model 1, a 1-point increase in Walk Score would decrease the risk of default from 0.90% to 0.88% (i.e., 0.90 x (1 – 0.018)).

Model 2 is the reduced version of Model 1. Insignificant variables are dropped to produce a more parsimonious model to achieve the best fit with the fewest parameters. Using irrelevant variables increases the standard error of the parameter estimates and reduces significance (Menard 1995). Removing controls did not alter the coefficient or significance of WALKSCORE, indicating that its relationship with DEFAULT was unaffected by any relationships between DEFAULT and the variables that were eliminated for Model 2.

The Goodness of Fit statistics are reported in the last four rows of Table 2. They measure how well all the explanatory variables in each model, taken together, predict DEFAULT. The higher the chi-square and the lower the -2 log likelihood, the better the model predicts DEFAULT. Comparing these statistics for Models 1 and 2 indicates that goodness of fit declines slightly as variables are removed, which normally occurs when variables are eliminated. Goodness of fit was also tested using the Area Under the Receiver Operating Characteristic (ROC) Curve. It measures a model's ability to discriminate between loans that do and do not default. It represents the likelihood that a loan that defaults will have a higher predicted probability than a loan that does not. If the result is equal to 0.5, the model is no better than flipping a coin. For all the models, ROCs were .83 to .85 indicating excellent discrimination (Hosemer and Lemeshow 2000). In other words, all the models did an excellent job distinguishing between loans that did and did not default.

A degree of non-linearity in the logit was detected for WALKSCORE using the Box-Tidwell transformation. Following that approach, a multiplicative term in the form of WALKSCORE times the log-normal form of WALKSCORE was added to Model 2. Statistically significant interaction terms indicated that linearity may not be a reasonable assumption for WALKSCORE.

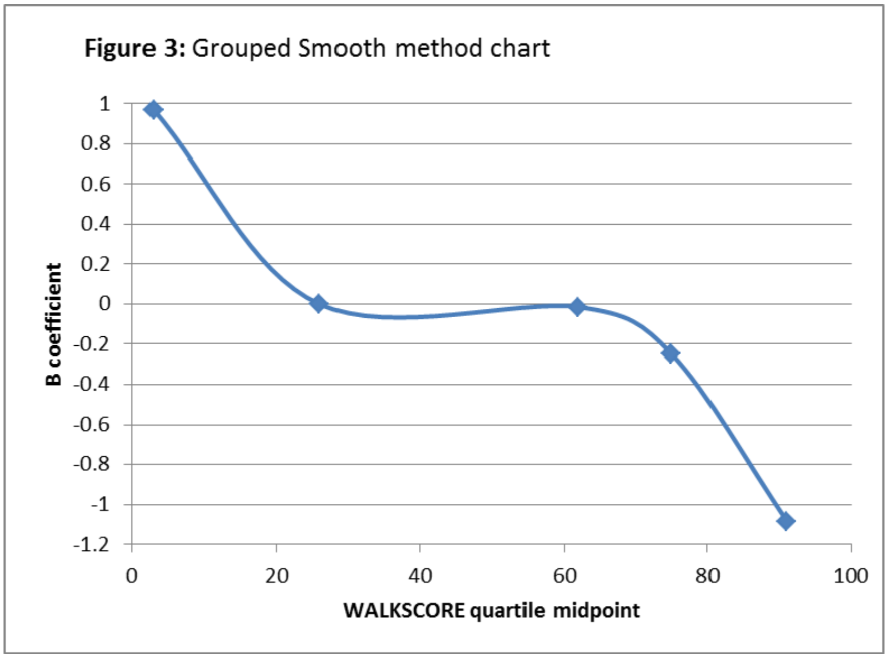
Two graphical methods were used to further investigate the shape of the nonlinear relationship between WALKSCORE and DEFAULT. In the first approach 20 groups of cases were created using 5 point increments of WALKSCORE. The average WALKSCORE for each group was then plotted against the average DEFAULT for each group. The result is shown in Figure 2 along with a 3 order polynomial fitted line. The patterns suggested there were two thresholds; one at a Walk Score of about 8 below which there was a marked increase in default risk and one at a Walk Score of about 80 above which there was a marked decrease in default relative to the normal default rate of about 0.9%.



This first graphical method for investigating nonlinearity does not use control variables. In order to take the controls into consideration, the Grouped Smooth Method suggested by Hosmer and Lemeshow (2000) was employed. First, the quartiles of the distribution of WALKSCORE were determined. Next, a categorical variable with 4 levels was created using the three cut-points based on the quartiles. An additional categorical variable was also created using 8 on WALKSCORE as the cut-point, in order to investigate the threshold of 8 found in the prior graphic analysis. Then, the multivariable model (Model 2) was refitted, replacing the continuous WALKSCORE variable with the 4-level categorical variable and the dummy for 8 or less, using the lowest quartile as the reference group. The coefficients for each of the 3 categorical variables were then plotted against the midpoints for WALKSCORE in each of the groups. A coefficient equal to zero was also plotted at the midpoint of the first quartile. The resulting data and plot were as follows:

Table 3: Estimated logistic regression coefficients vs. quartile midpoints

Range	Midpoints	B (sig.)
0-8	3	0.966(.019)
52-69	62	.020(.888)
69-83	75	-.222(.173)
83-100	91	-1.063*(.000)



The Grouped Smooth Method confirmed that the relationship between WALKSCORE and DEFAULT was nonlinear while holding control variables constant. It also showed the existence of the previously discovered thresholds. As shown in Table 3 and as suggested by the shape of the line in Figure 3, in the middle range of WALKSCORE, the coefficients were small and insignificant. That suggests that the middle range of WALKSCORE is unhelpful for predicting DEFAULT. However, at the lowest and highest levels the coefficients were larger and significant.

In an applied setting, cut-points can be more useful than continuous indicators because they allow a simple risk classification of cases into “high” and “low” and they communicate clearly the threshold above (or below) which risk will consistently be above (or below) average (Williams et al. 2006). In this case thresholds could identify the cut-points for WALKSCORE above which default risk is consistently below average and below which it is consistently above average.

Using a method for finding optimal cut-points recommended by Williams et al. (2006), candidate cut-points were evaluated by comparing the default rates above and below each candidate WALKSCORE value and computing a p-value for the difference using the chi-square test. This method indicated that 80 was the most significant WALKSCORE cut-point at the upper level and 8 was the most significant at the lower level.

Based on this analysis, Model 3 was produced using dummy variables indicating whether or not a property had a Walk Score of 80 or more (WALKSCORE80+) or 8 or less (WALKSCORE80-). Model 3 had better goodness of fit statistics than Model 2, meaning that it did a better job predicting DEFAULT than the prior model that treated WALKSCORE as a continuous variable. In Model 3, the Exp(B) for WALKSCORE80+ was 0.397, indicating that when a property had a WALKSCORE of 80 or more, it had a

60.3% decrease in the odds of default. In terms of relative risk, we can say that the relative risk of default was 60.3% lower for the properties with a Walk Score above 80 than those below 80. Similarly, Exp(B) for WALKSCORE8- was 2.208, indicating that properties with Walk Scores of 8 or less had a 121% increase in the odds of default (the odds of default for properties with Walk Scores greater than 8 are multiplied by 2.208).

Model 4 was the final model produced in order to show that using WALKSCORE in the default model improved its goodness of fit. It includes the same variables as Model 3, except for WALKSCORE80+ and WALKSCORE8-. Comparison of the goodness of fit statistics for Models 3 and 4 shows that goodness of fit was better for Model 3, when the Walk Score variables were in the model. That indicates that Walk Score can be used to improve our ability to predict default and discriminate between loans that do and do not default.

Discussion and Conclusion

The hypothesis was that greater walkability, as measured by higher Walk Scores, reduces mortgage default risk. The results supported the hypothesis; however the relationship was not linear. Instead, there were thresholds at Walk Scores of 8 and 80, above which significant declines in mortgage risk occurred.

A key implication of this study is that walkability could be fostered for highly walkable properties by relaxing lending terms without increasing default risk. For example, in terms of the impact on default rate, Model 3 predicts that the risk of default would be 0.9% for a property with a Walk Score between 9 and 79 and average values on the other model variables. This includes an OLTV of .61 and an ODSCR of 1.52, which are the sample means. However, if WALKSCORE were 80 or more, the OLTV for the same average property could be increased to .83, the ODSCR could be reduced to 1.23 and the property would still have a predicted default risk of 0.9%, according to Model 3. Inversely, with a Walk Score of 8 or less, the loan terms would need to be tightened to an OLTV of .51 and an ODSCR of 2.01 in order to produce a default risk of 0.9%. Figures for these scenarios are given in Table 3.

If higher LTV ratios at origination could be obtained by borrowers on more walkable properties, they could achieve a higher return on their equity as long as positive leverage is possible (i.e., when the cost of debt financing as indicated by the mortgage constant is lower than the overall return generated by the property as indicated by the return on assets). They could also use the unused portion of their equity funds for other projects that could diversify their investment portfolios. All else being equal, more attractive loan terms could make walkable property investments more attractive to investors, increase capital flow to more sustainable buildings, and foster a transition toward more sustainable cities.

Walkability has several potential social and environmental benefits, not the least of which include improved public health and mitigation of global climate change and other environmental impacts linked to motorized transportation. Fortunately, as this paper shows, properties in highly walkable locations, as indicated by a Walk Score of 80 or more, can also reduce mortgage default risk by more than 60%. This means that lenders could be willing partners in the promotion of more walkable cities by offering better terms for walkable property investments without increasing the exposure by lenders to default risk.

Socially responsible property investing has been described as maximizing the positive effects and minimizing the negative effects of property investment on society and the natural environment in a way that is consistent with investor goals and fiduciary responsibilities (Pivo and McNamara 2005). If it is possible to promote, as this study suggests, social and environmental goals through greater walkability without increasing default risk, then it seems that ethical, responsible lenders should offer better terms for more walkable properties. It may even be possible to promote walkability while improving business outcomes. In that case, investment in walkable places is simply a smarter way of doing business.

References

1. Archer WR, Elmer PJ, Harrison DM and Ling DC (2002), Determinants of Multifamily Mortgage Default, Real Estate Economics 30(3), 445-473.
2. Berrigan D, Carroll DD, Fulton JE, Galuska DA, Brown DR, Dorn JM, Armour B, Paul P (2012), Vital Signs: Walking Among Adults — United States, 2005 and 2010, Morbidity and Mortality Weekly Report 61(31): 595-601.
3. Bosch P and Metz B (2011), Options for Mitigating Climate Change: Results of Working Group III of the Fourth Assessment Report of the IPCC, in H.G. Brauch et al. (eds.), Coping with Global Environmental Change, Disasters and Security, Springer-Verlag, Berlin 2011.
4. Bradley DS, Cutts AC and Follain JR (2001), An Examination of Mortgage Debt Characteristics and Financial Risk among Multifamily Properties, Journal of Housing Economics 10, 482-506.
5. Cao, X., S.L. Handy and P.L. Mokhtarian (2006), The Influences of the Built Environment and Residential Self-Selection on Pedestrian Behavior: Evidence from Austin, TX. Transportation 33: 1–20.
6. Carr LJ, Dunsiger SI and Marcus BH (2010), Walk Score™ As a Global Estimate of Neighborhood Walkability, American Journal of Preventative Medicine 39(5):460–463.
7. Carr, LJ, Dunsiger, SI, and Marcus, BH. (2011), Validation of Walk Score™ for Estimating Access to Walkable Amenities. British Journal of Sports Medicine, 45:1144-1148.
8. Chapman L (2007), Transport and climate change: a review, Journal of Transport Geography 15: 354–367
9. Cramer, JS (2007), Robustness of Logit Analysis: Unobserved Heterogeneity and Mis-specified Disturbances, Oxford Bulletin of Economics and Statistics 69(4), 545-555.
10. du Toit, L., E. Cerin, E. Leslie and N. Owen. 2007. Does Walking in the Neighborhood Enhance Local Sociability? Urban Studies 44(9): 1677–1695.
11. Duncan DT, Aldstadt J, Whalen J, Melly SJ and Gortmaker SL (2011), Validation of Walk Score® for Estimating Neighborhood Walkability: An Analysis of Four US Metropolitan Areas, International Journal of Environmental Research and Public Health 8: 4160-4179.
12. Duncan DT, Aldstadt J, Whalen J, Melly SJ (2013) Validation of Walk Scores and Transit Scores for estimating neighborhood walkability and transit availability: a small-area analysis, GeoJournal 78:407–416.

13. Esaki H, L'Heureux S and Snyderman M (1999), Commercial Mortgage Defaults: An Update, Real Estate Finance 16(1), 80-86.
14. Ewing R. and R. Cervero, (2001), Travel and the Built Environment: A Synthesis, Transportation Research Record 1780: 87–114.
15. Frank, L.D. and P. Engelke (2005), Multiple Impacts of the Built Environment on Public Health: Walkable Places and the Exposure to Air Pollution, International Regional Science Review 28(2): 193–216.
16. Frank, L. and G. Pivo (1994), The Impacts of Mixed Use and Density on the Utilization of Three Modes of Travel: The Single Occupant Vehicle, Transit, and Walking, Transportation Research Record 1466: 44–52.
17. Frank, L.D., B. Stone and W. Bachman. 2000. Linking Land Use with Household Vehicle Emissions in the Central Puget Sound: Methodological Framework and Findings. Transportation Research Part D 5: 173–196.Freeman et al. 2012
18. Geels, FW (2012), A socio-technical analysis of low-carbon transitions: introducing the multi-level perspective into transport studies, Journal of Transport Geography 24: 471–482.
19. Goldberg L and Capone CA (1998), Multifamily Mortgage Credit Risk: Lessons from Recent History, Cityscape 4(1), 93-113.
20. Goldberg L and Capone CA (2002), A Dynamic Double-Trigger Model of Multifamily Mortgage Default, Real Estate Economics 30(1), 85-113.
21. Grovenstein RA, Harding JP, Sirmans CF, Thebpanya S and Trunbull GK (2005), Commercial mortgage underwriting: How well do lenders manage the risks? Journal of Housing Economics 14, 355-383.
22. Gyourko JE and Rybczynski W (2000), Financing New Urbanism projects: Obstacles and solutions, Housing Policy Debate 11 (3): 733-750.
23. Handy, S. (1993), Regional Versus Local Accessibility: Implications for Nonwork Travel, Transportation Research Record 1400: 58–66.
24. Handy, S., X. Cao and P. Mokhtarian (2005), Correlation or Causality between the Built Environment and Travel Behavior? Evidence from Northern California, Transportation Research Part D 10: 427–444.
25. Hoehner, CM, Ramirez LKB, Elliott MB, Handy SL and Brownson SC (2005), Perceived and Objective Environmental Measures and Physical Activity among Urban Adults, American Journal of Preventative Medicine 28(2S2): 105–116.
26. Hosemer DW and Lemeshow S (2000), Applied Logistic Regression, 2nd Edition, John Wiley & Sons, Inc, New York.Grovenstein RA,
27. Harding JP, Sirmans CF, Thebpanya S and Trunbull GK (2005), Commercial mortgage underwriting: How well do lenders manage the risks? Journal of Housing Economics 14, 355-383.
28. Kok N and Jennen M (2012), The impact of energy labels and accessibility on office rents, Energy Policy 46: 489-497.
29. Lee, C. and A.V. Moudon, (2006), The 3Ds + R: Quantifying Land Use and Urban Form Correlates of Walking, Transportation Research Part D 11: 204–215.
30. Leinberger CB (2007), Back to the Future: The Need for Patient Equity in Real Estate Development Finance, Research Brief, The Brookings Institution, Washington, DC.

31. Leinberger CB and Alfonso M (2012), Walk this Way: The Economic Promise of Walkable Places in Metropolitan Washington, D.C., The Brookings Institution, Washington, DC.
32. Leyden, K.M. (2003), Social Capital and the Built Environment: The Importance of Walkable Neighborhoods, American Journal of Public Health 93(9): 1546–1551.
33. Li, M.M. and H.J. Brown (1980), Micro-Neighborhood Externalities and Hedonic Housing Prices, Land Economics 56(2): 125–141.
34. Manaugh K and El-Geneidy A (2011), Validating walkability indices: How do different households respond to the walkability of their neighborhood? Transportation Research Part D 16: 309–315
35. Menard, S (1995), Applied Logistic Regression, Sage University Paper series on Quantitative Applications in the Social Sciences, 07-106, Sage, Thousand Oaks, CA.
36. Parks JR and Schofer JL (2006), Characterizing Neighborhood Pedestrian Environments with Secondary Data, Transportation Research Part D 11: 250–263.
37. Pivo G (2013), The Effect Of Transportation, Location, And Affordability Related Sustainability Features On Mortgage Default Prediction And Risk In Multifamily Rental Housing, Fannie Mae, Washington, DC.
38. Pivo G and Fisher J (2011), The Walkability Premium in Commercial Real Estate Investments, Real Estate Economics 39(2), 185-219.
39. Pivo G and McNamara P (2005) Responsible Property Investing, International Real Estate Review 8(1), 128-143.
40. Portney K (2005), Civic Engagement and Sustainable Cities in the United States, Public Administration Review 65(5): 579-591.
41. PwC and the Urban Land Institute (2012), Emerging Trends in Real Estate 2013, Washington, D.C.: PwC and the Urban Land Institute.
42. Rauterkus S and Miller N (2011), Residential Land Values and Walkability, Journal of Sustainable Real Estate 3(1): 23-43.
43. Rogers SH, Halstead JM, Gardner KH, Carlson CH (2011), Examining Walkability and Social Capital
44. as Indicators of Quality of Life at the Municipal and Neighborhood Scales, Applied Research in Quality Life 6: 201–213.
45. Snyderman, M (1991), Commercial mortgages: default occurrence and estimated yield impact, Journal of Portfolio Management 18(1), 6-11.
46. Sugiyama T, Neuhaus M, Cole R, Giles-Corti B, Owen N (2012), Destination and Route Attributes Associated with Adults' Walking: A Review, Medicine & Science In Sports & Exercise 44(7):1275-86.
47. Titman S and Torous W (1989), Valuing Commercial Mortgages: An Empirical Investigation of the Contingent-Claims Approach to Pricing Risky Debt, The Journal of Finance 59, 165-206.
48. Vandell KD (1984), On the Assessment of Default Risk in Commercial Mortgage Lending, AREUEA Journal 12(3), 270-296.
49. Vandell KD (1992), Predicting Commercial Mortgage Foreclosure Experience, Journal of the American Real Estate and Urban Economics Association, 20(1), 55-88.
50. Vandell K, Barnes W, Hartzell D, Kraft D and Wendt W (1993), Commercial Mortgage Defaults: Proportional Hazards Estimation Using Individual Loan Histories, Journal of the American Real Estate and Urban Economics Association 4(21), 451-480.

51. Weidema, P and Wesnæs, MS (1996), Data quality management for life cycle inventories – an example of using data quality indicators, Journal of Cleaner Production 4(3/4): 167-74.
52. Williams BA, Mandrekar JN, Mandrekar SJ, Cha SS, Furth AF (2006), Technical Report Series No. 79, Finding Optimal Cutpoints for Continuous Covariates with Binary and Time-to-Event Outcomes, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota.
53. Wood L, Frank LD, Giles-Corti B (2010), Sense of community and its relationship with walking and neighborhood design, Social Science & Medicine 70: 1381–1390.
54. World Cancer Research Fund/American Institute for Cancer Research. 2009. Food, Nutrition, and Physical Activity: A Global Perspective. American Institute for Cancer Research: Washington, DC.

Data Tables

Table 1: Descriptives Statistics

	All Loans		Defaulted Loans		Non-defaulted Loans		Difference Tests		
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	*t*-test	Likelihood Ratio	Pearson Chi-Square
Dependent Variable									
Fraction of loans defaulting	0.86%		100%		0%				
Walkability Variable									
Walk Score	66.0	21.8	61.6	21.0	66.1	21.8	0.000		
Loan characteristics									
Loan-to-value ratio at origination	61.20%	16.30%	70.40%	11.50%	61.20%	16.30%	0.000		
Debt coverage ratio at origination	1.5	0.6	1.3	0.3	1.5	0.6	0.000		
Loan age in months	73.2	52.9	67.9	33.1	73.2	53.0	0.005		
ARM flag	0.31	0.462	0.39	0.49	0.31	0.46			
Property characteristics									
No concerns	0.29	0.45	0.12	0.32	0.29	0.45		0.000	0.000
Year built	1968.0	26.3	1955.0	32.1	1968.0	26.2	0.000		
Total units	94.6	125.0	64.2	99.5	94.9	125.2	0.000		
Neighborhood and city characteristics									
Median household income in 2000 census tract	42,694	16,957	34,085	13,483	42,768	16,965	0.000		
Property crime per million capita in city	407.5	165.3	474.5	161.6	406.9	165.2	0.000		
Housing vacancy rate 2011 block group (%)	6.58	5.87	9.85	7.45	6.56	5.85	0.000		
Urban/Rural Continuum	1.92	1.16	2.00	1.08	1.92	1.16		0.001	0.000
Principal City	0.60	0.49	0.68	0.47	0.60	0.49		0.002	0.002
Top 25 City	0.23	0.42	0.19	0.39	0.23	0.42		0.069	0.076
Geographic Variables									
New England	0.03	0.17	0.13	0.34	0.03	0.47		0.000	0.000
Mid Atlantic	0.14	0.35	0.15	0.36	0.14	0.35		0.590	0.586
East North Central	0.08	0.26	0.15	0.36	0.08	0.26		0.000	0.000
East South Central	0.02	0.14	0.02	0.14	0.02	0.15		0.906	0.906
West North Central	0.04	0.19	0.02	0.15	0.04	0.19		0.102	0.131
South Atlantic	0.09	0.29	0.22	0.42	0.09	0.29		0.000	0.000
West South Central	0.08	0.27	0.06	0.24	0.08	0.27		0.287	0.303
Mountain	0.05	0.22	0.06	0.24	0.05	0.22		0.397	0.382
Pacific	0.47	0.50	0.17	0.38	0.47	0.50		0.000	0.000
New York City	0.03	0.16	0.01	0.10	0.03	0.16		0.021	0.045
Washington, D.C.	0.01	0.08	0.01	0.08	0.01	0.08		0.895	0.893
Avg. pct. price change in MSA, past 6 yrs.	-1.3	3.5	-1.6	2.7	-1.3	3.7	0.266		
Avg. pct. leased in MSA, past 6 yrs.	91.0	3.9	90.9	3.7	91.0	3.9	0.127		

TABLE 2: *Logistic Regression Results for DEFAULT*

	Model 1: All variables		*Model 2: Insignificant variables removed*		Model 3: Walk Score 80 plus or 8 minus		Model 4: Without Walk Score	
	B (sig.)	*Exp(B)*	*B (sig.)*	*Exp(B)*	B (sig.)	Exp(B)	B (sig.)	Exp(B)
WALKSCORE	-.018 (.000)	0.982	-.018 (.000)	0.982				
WALKSCORE * ln*(WALKSCORE)*								
WALKSCORE80+					-.924 (.000)	0.397		
WALKSCORE8-					.792 (.046)	2.208		
Loan								
OLTV	.029 (.000)	1.029	.028 (.000)	1.028	.027 (.000)	1.028	.032 (.000)	1.033
ODSCR	-1.120 (.000)	0.326	-1.133 (.000)	0.322	-1.100 (.000)	0.333	-1.072 (.000)	
ARM_FLAG	.719 (.000)	2.053	.758 (.000)	2.135	.657 (.000)	1.929	.775 (.000)	2.17
LOAN_AGE_MONTHS	-.001 (.301)	0.999						
Property								
NOCONCERNS	-.892 (.000)	0.410	-.907 (.000)	0.404	-.879 (.000)	0.415	-.952 (.000)	0.386
BUILT_YR	-.016 (.000)	0.984	-.015 (.000)	0.985	-.018 (.000)	0.982	-.013 (.000)	0.987
TOT_UNTS_CNT	-.005 (.000)	0.995	-.005 (.000)	0.995	-.005 (.000)	0.995	-.005 (.000)	0.995
Neighborhood and City								
MEDHHINC000	-.027 (.000)	0.974	-.029 (.000)	0.972	-.030 (.000)	0.971	-.027 (.000)	0.974
PROP_CRIME_MIL	.001 (.011)	1.001	.001 (.001)	1.001	.001 (.000)	1.001	.001 (.002)	1.001
VACANCY_RATE	.023 (.008)	1.023	.0223 (.006)	1.023	.022 (.008)	1.022	.024 (.004)	1.025
PRINCIPAL_CITY	.313 (.033)	1.368						
URBAN_RURAL	-.154 (.015)	0.858	-0.139 (.024)	0.870				
Regional Economy								
TOP25CITY	-.203 (.239)	0.816						
DC	-1.057 (.151)	0.347						
NYC	-.731 (.212)	0.457						
REGION	unreported		unreported		unreported		unreported	
AVG_PRICE_6	.003 (.857)	1.003						
AVG_PCT_LEASED_6	.021 (.185)	1.021						
Constant	25.926 (.000)	1.82E+11	26.909 (.000)	4.86E+11	32.318 (.000)	1.09E+14	20.288 (.000)	6.47E+08
n	36,922		36,922		36,922		36,922	
Goodness of Fit								
Model Chi-square	621.714		606.523		617.482		582.323	
-2 Log likelihood	3063.855		3079.046		3068.087		3111.265	
Nagelkerke R- Square	0.176		0.172		0.175		0.164	
Under ROC Curve	0.845		0.841		0.844		0.837	

Table 3: Trade-off experiments

Variables	Model 3	Mean case		Walk Score 80+ case		Walk Score 8- case	
	B	value	B x value	value	B x value	value	B x value
WALKSCORE80+	-0.924	0.000	0.000	1.000	-0.924	0.000	0.000
WALKSCORE8-	0.792	0.000	0.000	0.000	0.000	1.000	0.792
OLTV	0.027	61.296	1.679	83.000	2.274	51.000	1.397
ODSCR	-1.100	1.518	-1.669	1.230	-1.353	2.010	-2.210
ARM_FLAG	0.657	0.309	0.203	0.309	0.203	0.309	0.203
NOCONCERNS	-0.879	0.286	-0.252	0.286	-0.252	0.286	-0.252
BUILT_YR	-0.018	1967.834	-35.421	1967.834	-35.421	1967.834	-35.421
TOT_UNTS_CNT	-0.005	94.643	-0.469	94.643	-0.469	94.643	-0.469
MEDHHINC000	-0.030	42.694	-1.276	42.694	-1.276	42.694	-1.276
PROP_CRIME_MIL	0.001	407.479	0.411	407.479	0.411	407.479	0.411
VACANCY_RATE	0.022	6.573	0.142	6.573	0.142	6.573	0.142
New England	0.836	0.031	0.026	0.031	0.026	0.031	0.026
ENCENT	0.612	0.076	0.046	0.076	0.046	0.076	0.046
SoAtlantic	0.924	0.093	0.086	0.093	0.086	0.093	0.086
Pacific	-1.045	0.469	-0.490	0.469	-0.490	0.469	-0.490
Constant	32.318		32.318		32.318		32.318
Sum of B x value			-4.665		-4.677		-4.696
Exp(sum)			0.009		0.009		0.009
1+ Exp(sum)			1.009		1.009		1.009
Predicted Probability Exp(sum)/1+Exp(sum))			0.93%		0.92%		0.90%