# The Effect of Sustainability Features on Mortgage Default Prediction and Risk in Multifamily Rental Housing

**Author** Gary Pivo

**Abstract** This study examines the relationship between transportation-, location-, and affordability-related sustainability features and default risk in multifamily housing. It finds that sustainability features can be used to improve the prediction of mortgage default and reduce default risk. The study uses 37,385 loans in the Fannie Mae multifamily portfolio at the end of 2011:Q3. The results suggest two implications for practice. First, certain aspects of sustainability can be fostered without increasing default risk by adjusting conventional lending standards. Second, lenders could improve their risk management practices by taking stock of sustainability features when loans originate.

This study examines the relationship between sustainability features and mortgage default risk in multifamily rental housing. The borrowers are multifamily rental building owners with loans held by Fannie Mae. The sustainability features pertain to the social and environmental performance of properties including affordability, walkability, auto dependence, exposure to pollution, and proximity to protected open space.

The results show that information on the sustainability of multifamily buildings improves our ability to predict mortgage defaults. The results also show that loans on properties with sustainability features have a much lower risk of default. For example, defaults were 58% less likely for loans on properties in less auto-dependent locations, where 30% or more of the workers living there commute by subway or elevated train. Similarly impressive findings were found for each of the sustainability features examined in the study.

## Mortgage Default and Multifamily Housing

Mortgage loan defaults are a risk for multifamily lenders and investors. In a study of 495 securitized multifamily mortgages originating between 1989 and 1995, Archer, Elmer, Harrison, and Ling (2002) found a default rate of nearly 12%. More recently, in the fourth quarter of 2011, the default rate for multifamily mortgages held by depository institutions in the United States was 3.7% (Chandan, 2011).

Previous studies show that the major risk factors for multifamily loan default are cash flow and property value. Default risk increases if declining cash flow prevents loan repayment or if falling property value produces negative net equity (Vandell, 1984, 1992; Titman and Torous, 1989; Kau, Keenan, Muller, and Epperson, 1990; Vandell, Barnes, Hartzell, Kraft, and Wendt, 1993; Goldberg and Capone, 1998, 2002; Archer, Elmer, Harrison, and Ling, 2002). In these studies, cash flow and equity are commonly measured in terms of debt service coverage ratio (DSCR), or the ratio of income to required loan payments, and loan-to-value ratio (LTV), or the ratio of loan amount to property value. A lower DSCR and a higher LTV, both at origination and over the life of the loan, have been linked to greater default risk.

## The Connection between Mortgage Default and Sustainability

A growing number of econometric studies show that buildings with sustainability features generate more cash flow and value. The value premium appears to come from both the stronger cash flow and lower capitalization rates, suggesting that more sustainable properties are favored in both the space and capital markets by renters and investors. Stronger cash flow has also been tied to lower operating expenses. Specific sustainability features linked to these effects include LEED certification, ENERGY STAR labeling, historic preservation, design excellence, proximity to open space, good transit service, walkability, revitalizing urban location, and reduced exposure to pollution and natural hazards (Vandell and Lane, 1989; Harrison, Smersh, and Schwartz, 2001; McGreal, Webb, Adair, and Berry, 2006; Simons and Saginor, 2006; Miller, Spivey, and Florance, 2008; Eichholtz, Kok, and Quigley, 2009; Pivo and Fisher, 2010, 2011). Perhaps this should be unsurprising. After all, sustainability features promote qualities long associated with higher property value including health, safety, operating efficiency, occupant amenities, and accessibility.

If more sustainable buildings have better cash flow and value, then they should also exhibit lower default risk because, as noted above, default risk is inversely related to cash flow and value. However, adding information on sustainability features to the loan origination process would only be helpful if their impact on cash flow and value was not already fully accounted for in that process. If the financial benefits of sustainability were already fully reflected in the cash flow and value projections made at loan origination, then rational lenders may have already increased the size of the loans or reduced loan interest rates for the more sustainable properties (Grovenstein, Harding, Sirmans, Thebpanya, and Turnbull, 2005). The higher loan amounts or lower interest rates for the more sustainable properties would in turn have increased their LTV or lowered their DSCR, resulting in a probability of default more like that found in more conventional properties. That is to say, the effect of the sustainability variables would have already been fully endogenous to the loan origination process (Archer, Elmer, Harrison, and Ling, 2002). That would have caused the ex post default risk for the more sustainable properties to end up being similar to the risk for the conventional ones. In that case, adding the sustainability features to a default prediction model, which already includes the original LTV and DSCR, would not improve the ability of the model to predict default because the information carried by the sustainability regressors would already be included in the LTV and DSCR.

However, if the financial benefits of sustainability had not been fully accounted for when the loans were originated, but instead future cash flows and values for more sustainable properties were expected to be closer to those of conventional properties than they actually turned out to be, then as any unrecognized financial benefits from the sustainability features materialized post-origination, the average post-origination DSCR would turn out to be higher and the average post-origination LTV would turn out to be lower for the more sustainable properties than for the conventional ones. This would produce a lower default rate for the more sustainable properties compared to the conventional ones, because there would be a lower risk of negative cash flow or a lower risk of negative equity.

This is not to say that lenders would have completely ignored sustainability features at loan origination. Indeed, some of them, such as being in a transit-oriented location, are commonly recognized by lenders as locational advantages. But even if some of the sustainability features were previously recognized as advantages by lenders, as long as their full effect on cash flow, value, and default risk were not fully reflected in the original loan terms, the more sustainable properties would have a lower post-origination default rate than the more conventional ones.

In a modeling context, if the financial benefits of sustainability were not completely reflected in the original DSCR and LTV, then adding information on sustainability features to a default prediction model could significantly improve the accuracy of that model. Put another way, if sustainability features help buffer properties from losing cash flow and value in the face of energy shocks, recessions, competition, or other events that can push cash flow and value into the default "danger zone" (Bradley, Cutts, and Follain, 2001), then knowing that a property is more or less sustainable should help lenders mitigate default risk.

These expectations led to Hypotheses 1 and 2.

Hypothesis 1: If certain transportation-, location-, and housing-related sustainability features are added to a model of default risk, the accuracy of the model will improve.

This is based on the expectation that most, if not all, of the impact that sustainability features have on future cash flow and value was unanticipated when multifamily loans were originated. Even though some advantages of the features may have been considered by underwriters, this hypothesis proposes that their positive effect on cash flow or value was not fully accounted for in the original loan terms.

Hypothesis 2: Sustainability features will be associated with a lower risk of default, ceteris paribus.

This should be the case because prior research shows that sustainability features tend to be associated with increased cash flow and value, which are in turn related to lower default risk.

## Methods

### *Logistic Regression Model*

A logistic regression model was used to test the hypotheses. This model has been used in several prior studies to estimate the effects of explanatory variables on the probability of mortgage default (Vandell, Barnes, Hartzell, Kraft, and Wendt, 1993; Goldberg and Capone, 1998, 2002; Archer, Elmer, Harrison, and Ling, 2002, Rauterkus, Thrall, and Hangen, 2010). A primer on logistic regression and its alternatives for studying default risk is given in the Appendix.

The data on individual mortgages used to build the model was provided by Fannie Mae and then combined by the author with data from other sources in order to measure sustainability and control variables. The Fannie Mae data included information on every loan in its multifamily portfolio at the end of 2011:Q3, making the study cross-sectional rather than longitudinal. The cross-sectional design raises some concern about the external validity of the findings (i.e., how far the findings can be generalized beyond the study sample) because the relationships between the regressors and default risk could change over time. For example, proximity to transit could reduce default rates by a greater amount when gas prices are peaking and demand is higher for apartments near public transit. Unfortunately, longitudinal data were unavailable for this study. It would be useful to confirm the results reported here in a follow-up study using longitudinal data. Another external validity issue comes from the fact that the Fannie Mae mortgage pool had an average default rate that was about one-fourth the rate found for mortgages held by depository institutions. It would be important to know whether the effects found in this study apply to those mortgages as well.

In the present work, 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 collateral property, and on the number of days the loan was delinquent, if any. More details on these variables and those from other sources are given below.

Following Archer, Elmer, Harrison, and Ling (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 the cases used 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 DSCR 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, 37,385 loans remained in the sample out of the 42,474 cases. The sample included mortgages with fixed and adjustable rates and with a wide variety of seasoning, originating anywhere from September, 1971 to September, 2011.

The variables evaluated in the models are described in the next section. Exhibit 1 gives their definitions and summary statistics.

## Variables

*DEFAULT* was a binary variable indicating whether or not a loan was in default as of 2011:Q3. Loans were classified as in default if they were delinquent on their payments by 90 days or more as of 2011:Q3.

Seven sustainability variables were analyzed. Sustainability is a multi-dimensional construct with multiple distinct but related dimensions treated as a single theoretical concept. As such, the variables used in this study could not capture every dimension normally used to assess building sustainability (Pivo, 2008). Key omissions due to data limitations included operational energy and water efficiency. It would be useful to test the effect of these and other sustainability metrics on mortgage outcomes in future studies. If they affect cash flow and value, they are likely to affect default risk as well.

For some of the sustainability variables, a dummy that indicated whether their value fell above or below a cut-point produced a better result in the models than a continuous variable. This is common when a change in the dependent variable associated with a one-unit change in the independent variable is nonlinear and there is a suspected or assumed threshold effect (Williams, Mandrekar, Mandrekar, Cha, and Furth, 2006). Where nonlinearity was suspected, possible cut-points were examined by categorizing the relevant continuous sustainability variable into quantiles (e.g., quartiles or deciles) and comparing the default rates for each group. Default rates significantly higher for all mortgages above or below a certain quantile indicated a discontinuity and suggested a likely cut-point where a threshold effect may occur. The cut-point was then used to create a new dummy variable for testing in the model. This procedure was used iteratively to find the ''optimal'' cut-point which produced a dichotomous variable that was most useful in predicting default. In a practical setting, cut-points can be more useful than continuous indicators because they allow a simple risk classification into ''high'' and ''low'' and communicate clearly the threshold above (or below) which default risk will be consistently above (or below) average.

Three variables were used to capture the nature of the journey to work by residents living in the census tract where each property was located. These variables address the air, water, and wildlife issues linked to commuting and related social issues including traffic accidents, physical activity, and social interaction. *COMMUTE TIME* was the average commute time in minutes for people 16 years of age and older living in the tract who worked outside the home. *SUBWAY30* indicated whether a property was in a tract where at least 30% of the workers take a subway or elevated train to work. The 30% cut-point was selected using the ''optimal cut-point'' method described above. *PCTWALK* indicated the percentage of people in the tract who walk to work. All commuting data were from the 2000 U.S. Census. Newer data were unavailable at the time of the study. It is unlikely, however, that changes in commuting patterns over the past decade were sufficient to significantly alter the results.

**Exhibit 1** | Definitions and Summary Statistics for Variables in Final Models ($n = 37{,}385$)

| Variable | Definition | Min. | Max. | Mean | Std. Dev. |
|---|---|---|---|---|---|
| Dependent Variable | | | | | |
| *DEFAULT* | Binary variable indicating whether loan was (1) or was not (0) in default. Default was defined as >90 days delinquent as of 2011:Q3. | 0 | 1 | 0.01 | 0.09 |
| Sustainability Variables | | | | | |
| *COMMUTE TIME* | Mean commute time in 2000 for residents in the census tract (minutes). | 2.10 | 75.50 | 26.72 | 6.72 |
| *SUBWAY30* | In census tract where at least 30% of workers use a subway or elevated train for work (1 = yes, 0 = no). | 0 | 1 | 0.08 | 0.27 |
| *PCTWALK* | Percent of people in census tract that walked to work in 2000. | 0 | 100 | 4.92 | 7.40 |
| *RETAIL16* | Sixteen or more retail establishments in census block group (1 = yes, 0 = no). | 0 | 1 | 0.40 | 0.49 |
| *AFFORDABLE* | Meets FNMA affordable housing standards (1 = yes, 0 = no). | 0 | 1 | 0.10 | 0.30 |
| *FREEWAY1000FT* | Within 1,000 feet of an interstate freeway (1 = yes, 0 = no). | 0 | 1 | 0.40 | 0.49 |
| *PROTECTED1MILE* | Within 1 mile of a protected area (1 = yes, 0 = no). | 0 | 1 | 0.71 | 0.45 |
| Control Variables | | | | | |
| *OLTV* | Loan-to-value ratio at origination. | 0.59 | 100 | 61.30 | 16.29 |
| *ODSCR* | Debt service coverage ratio at origination. | 0.90 | 5.0 | 1.52 | 0.549 |
| *LOAN_AGE_MONTHS* | Number of months since loan origination. | 0.00 | 468.00 | 73.15 | 52.91 |
| *LOAN_SIZE_GP* | Original loan size; ordinal variable (1 ≤ \$3MM, 2 = \$3–5MM, 3 = \$5–25MM, 4 ≥ \$25MM). | 1 | 4 | 1.70 | 0.97 |
| *ARM_FLAG* | Dummy for adjustable-rate mortgage (1 = yes, 0 = no). | 0 | 1 | 0.31 | 0.46 |
| *NOCONCERNS* | Dummy indicating absence (1) or presence (0) of concerns about physical condition at origination. | 0 | 1 | 0.28 | 0.40 |
| *BUILT_YR* | The year the property was built. | 1800 | 2011 | 1967.83 | 26.25 |
| *TOT_UNTS_CNT* | Total housing units in the property. | 2 | 3284 | 94.65 | 125.05 |
| *PRINCIPAL_CITY* | Dummy for whether or not located in U.S. Census Principal City (1 = yes, 0 = no). | 0 | 1 | 0.60 | 0.49 |
| *URB_RUR* | Ordinal variable indicating whether property is in Principal Urban Center, Metro City, Urban Outskirts, Suburban Periphery, Small Town or Rural location. | 1 | 7 | 1.92 | 1.16 |
| *MEDHHINC000* | Median household income in census tract in 2000 (000 dollars). | 0.0 | 200.00 | 42.70 | 16.94 |
| *PROP_CRIME_MIL* | Annual number of property crimes per million persons in the city. | 0.0 | 2849.85 | 407.48 | 165.31 |

**Exhibit 1** | (continued)

Definitions and Summary Statistics for Variables in Final Models ($n$ = 37,385)

| Variable | Definition | Min. | Max. | Mean | Std. Dev. |
|---|---|---|---|---|---|
| *US_PRICE_CHANGE* | Percent change in NCREIF U.S. Apartment Index from quarter loan originated to 2011:Q3. | −13.54 | 137.89 | 16.26 | 27.22 |
| *NEWENGLAND* | Dummy for New England region. | 0 | 1 | 0.03 | 0.17 |
| *MIDATLANTIC* | Dummy for Mid-Atlantic region. | 0 | 1 | 0.14 | 0.35 |
| *ENCENT* | Dummy for East North Central region. | 0 | 1 | 0.08 | 0.26 |
| *WNCENT* | Dummy for West North Central region. | 0 | 1 | 0.04 | 0.19 |
| *SOATLANTIC* | Dummy for South Atlantic region. | 0 | 1 | 0.09 | 0.29 |
| *ESOCENTRAL* | Dummy for East South Central region. | 0 | 1 | 0.02 | 0.14 |
| *WSOCENT* | Dummy for West South Central region. | 0 | 1 | 0.08 | 0.27 |
| *MTN* | Dummy for Mountain region. | 0 | 1 | 0.05 | 0.22 |
| *PACIFIC* | Dummy for Pacific region. | 0 | 1 | 0.47 | 0.50 |
| *AVG_PRICE_6* | Average apartment price change in percent from 2005:Q3 to 2011:Q3 in percent in the MSA. | −50.30 | 20.79 | −1.33 | 3.51 |
| *AVG_OCC_6* | Average apartment occupancy rate in percent from 2005:Q3 to 2011:Q3 in percent in the MSA. | 33.43 | 100 | 91.04 | 3.66 |
| *TOP25CITY* | Dummy for in one of 25 most populous U.S. cities (1 = yes, 0 = no). | 0 | 1 | 0.23 | 0.42 |
| *NYC* | Dummy for in New York City (1 = yes, 0 = no). | 0 | 1 | 0.03 | 0.16 |
| *DC* | Dummy for in Washington, DC (1 = yes, 0 = no). | 0 | 1 | 0.06 | 0.08 |

*RETAIL16* captured walkability where the apartment buildings were located or the degree to which the area within walking distance of a property encourages walking from the property to other destinations. Walkability has been linked to various social and environmental benefits and increases with the number of desired destinations within walking distance of a property (Pivo and Fisher, 2011; Federal Highway Administration, 2012). *RETAIL16* was a dummy indicating whether the property was in a census block group with at least 16 retail establishments in 2011 and also was selected using the "optimal cut-point" method. The data were collected from Nielsen Claritas, which estimated the number of establishments in 2011 based on the 2007 U.S. Economic Census. The author is currently completing work on a follow-up study that examines the relationship between Walk Score—a widely available walkability metric—and multifamily mortgage risk.

*AFFORDABLE* indicated whether or not the loan was part of the Fannie Mae Targeted Affordable Segment, which focuses on financing properties with rent subsidies or income restrictions. Family well-being can be in jeopardy if too much of a household's budget is required for housing, leaving too little money for food, health care, childcare or other essentials (Bratt, 2002).

*FREEWAY1000FT* indicated whether a property was located within 1,000 feet of a freeway. There is growing evidence that living close to a freeway increases risk for autism, cancers, and respiratory disease (Gauderman et al., 2007; Volk, Hertz-Picciotto, Delwiche, Lurmann, and McConnell, 2011; Office of Health Hazard Assessment, 2012; Cakmak, Mahmud, Grgicak-Mannion, and Dales, 2012). Data on highway locations was obtained from the 2011 National Transportation Atlas Database.

*PROTECTED1MILE* indicated whether a property was located within a mile of a Protected Area according to the U.S. Protected Area Database. It includes public lands at all government levels held for conservation and voluntarily provided privately protected areas. Protected open space helps sustain resource-based industry, recreation, wildlife, watersheds, and other ecosystem services such as greenhouse gas absorption and heat island mitigation. Access to parks and recreation has also been linked to lower childhood obesity and other social benefits (Wolch et al., 2011).

In this study, the expectation was that if certain sustainability features were related to default risk, it is because they affect cash flow and/or value to a degree unaccounted for in the DSCR or LTV ratio at loan origination. However, it could also be true that the sustainability features are correlated with other factors that affect financial outcomes and default risk, such as other loan, property, neighborhood, or macroeconomic variables, raising questions about whether sustainability is a proxy for these other drivers of cash flow and value. Therefore, to separate the effects of sustainability features on default risk from these other factors, 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.

*OLTV* and *ODSCR* measure the LTV and debt service coverage ratios at loan origination. Higher *OLTV* and lower *ODSCR* were expected to be associated with greater default risk. *LOAN_SIZE_GP* was an ordinal variable for the loan amount at origination. Esaki, L'Heureux, and Snyderman (1999) found that smaller commercial loans had lower default rates. *ARM_FLAG* was a dummy indicating whether the loan is adjustable or fixed. *LOAN_AGE_MONTHS* was the number of months from the loan origination date to the observation date. 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, L'Heureux, and Snyderman, 1999; Archer, Elmer, Harrison, and Ling, 2002). The same pattern was observed in this study sample. Linearity is not a requirement of the logistic regression model and it was unnecessary to transform *LOAN_AGE_MONTHS* to obtain significant results. However, some non-linearity in the logit was detected for *LOAN_AGE_MONTHS* using the Box-Tidwell transformation (Menard, 1995), so transformations of *LOAN_AGE_MONTHS* were tried but they did not improve the results.

*NO_CONCERNS* was a dummy indicating whether 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, Elmer, Harrison, and Ling (2002) found that default rates increased with building age, so *BUILT_YR* was expected to be inversely related to default risk, although in some areas historic buildings may be prevalent, which could influence how age relates to risk.

*TOT_UNTS_CNT* was the total number of units in the property. Smaller properties have been reported to experience more financial distress (Bradley, Cutts, and Follain, 2000). Archer, Elmer, Harrison, and Ling (2002), however, found that size was unrelated to default, although their univariate analysis showed that smaller properties had less risk, contrary to Bradley, Cutts, and Follain (2000). So the expected effect in this study was ambiguous.

Four control variables were created to control for geographical effects at the city and neighborhood level. Archer, Elmer, Harrison, and Ling (2002) found geographical effects to be one of the most important dimensions for predicting multifamily mortgage default. *PRINCIPAL_CITY* indicated whether the property was located in a Principal City, defined by the U.S. Census as the largest incorporated or census designated place in a core-based statistical area. Its purpose was to control for whether or not a property was centrally located in a metro- or micropolitan area because many central areas have outperformed suburban locations over the past decade and several sustainability features, such as walkability, are more common 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 defined in the ESRI Tapestry Segmentation system, which groups locations along an urban-rural continuum from "Principal Urban Centers" to "Small Towns and Rural" places. *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 the U.S. Department of Justice. Higher crime was expected to increase default risk.

Regional and national variables were used to control for differences in the economic context experienced by properties since loan origination. Dummies were created for the nine census divisions. Vandell, Barnes, Hartzell, Kraft, and Wendt (1993) used a similar variable. Other variables were tested but found to be insignificant. They included state, metropolitan area, and city location. Additional variables designed to capture regional economic effects were whether the property was in one of the 25 largest cities (*TOP25CITY*), 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 metro area in the most recent six-year period. *USPRICE_CHANGE* was a national indicator that captured the percentage change in the National Council of Real Estate Investment Fiduciaries (NCREIF) U.S. Apartment Index that occurred from the time the loan was originated to the observation date (2011:Q3). *AVG_PRICE_6* and *AVG_OCC_6* were computed using the NCREIF Apartment Index for metro areas. They described the average increase in apartment prices and the average occupancy rate in the metro area for each property over the last six years prior to the study observation date. Prior researchers have used updates of LTV and DSCR over time to predict default on the theory that negative equity or cash flow will trigger default. Both are affected by the property's net operating income, which is in turn affected by vacancy rates and rental price indices. Therefore, changes in vacancy rates and rental price indices at the metro scale can be used to capture changes in market conditions that strengthen or weaken mortgages over time (Goldberg and Capone, 1998, 2002).

Lenders consider borrower characteristics to be crucial in predicting default rates. Relevant variables include borrower character, experience, financial strength, and credit history. In their "simple model of default probability," Archer, Elmer, Harrison, and Ling (2002) theorized that losses from loans depend on the risk characteristics of the borrower, among other things, though such variables were not included in their models. Vandell, Barnes, Hartzell, Kraft, and Wendt (1993) used borrower type (individual, partnership, corporation, other) in their analysis of commercial mortgage defaults, as did Ciochetti, Deng, Gao, and Yao (2003), who expected individuals to represent a lower risk to lenders, although neither study found these variables to be significant. Unfortunately, due to privacy concerns, data on borrowers were unavailable for this study. It is likely, however, that lenders adjusted the original loan terms based in part on their assessment of borrower characteristics. Therefore, the *LTV*, *DSCR*, and *ARM_FLAG* variables may serve as proxies for borrower characteristics. It is inappropriate to make assumptions, however, about the effects of omitting variables in logistic regression. It is known that omitting relevant variables introduces bias into linear regression, but less is known about how it may bias logistic regression (Dietrich, 2003). One study showed that omitted orthogonal variables (i.e., variables that are uncorrelated with other independent variables) can depress the estimated parameters of the remaining regressors toward zero (Cramer, 2007). That would make the findings about sustainability in this study appear to be weaker than they actually are. It would be helpful to include borrower characteristics in future work building on the present study.

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, which makes it more of a concern for Hypothesis 2 than Hypothesis 1. Tolerance statistics, which check for a relationship between each independent variable and all other independent variables, were used as an initial check and raised no concerns (Menard, 1995). A pairwise correlation matrix among the independent variables, however, did indicate possible issues. *LOAN_AGE_MONTHS* and *USPRICE_CHANGE* were moderately correlated (0.737), as were *TOT_UNTS_CNT* and *LOAN_SIZE_GP* (0.662), both of which make logical sense. *SUBWAY30* was also correlated with *MIDATLANTIC* (0.684) and *NYC* (0.601). Correlations at this level do not automatically mean there will be collinearity issues, but they do raise the need for further tests, which were done and are reported below.

## Results

Exhibit 2 gives the statistics for the three final models produced for the study. The first model predicts *DEFAULT* only using conventional explanatory variables

**Exhibit 2** | Logistic Regression Results for *DEFAULT*

| | Model 1: Without Sustainability | | Model 2: With Sustainability | | Model 3: Significant Variables Only | |
|---|---|---|---|---|---|---|
| | $\beta$ (sig.) | Exp($\beta$) | $\beta$ (sig.) | Exp($\beta$) | $\beta$ (sig.) | Exp($\beta$) |
| Loan | | | | | | |
| *OLTV* | 0.041 (.000) | 1.042 | 0.044 (.000) | 1.045 | 0.043 (.000) | 1.044 |
| *ODSCR* | −0.868 (.003) | 0.420 | −1.037 (.001) | 0.355 | −1.043(.001) | 0.352 |
| *LOAN_AGE_MONTHS* | −0.005 (.002) | 0.995 | −0.004 (.019) | 0.996 | −0.004 (.020) | 0.996 |
| *ARM_FLAG* | 0.578 (.000) | 1.782 | 0.468 (.001) | 1.596 | 0.477 (.000) | 1.611 |
| Property | | | | | | |
| *NOCONCERNS* | −0.902 (.000) | 0.406 | −0.820 (.000) | 0.440 | −0.827 (.000) | 0.437 |
| *BUILT_YR* | −0.015 (.000) | 0.985 | −0.016 (.000) | 0.985 | −0.016 (.000) | 0.984 |
| *TOT_UNTS_CNT* | −0.004 (.000) | 0.996 | −0.004 (.000) | 0.996 | −0.004 (.000) | 0.996 |
| Neighborhood | | | | | | |
| *PRINCIPAL_CITY* | 0.145 (.297) | 1.156 | 0.285 (.057) | 1.330 | | |
| *URB_RUR* | 0.041 (.481) | 0.960 | 0.020 (.745) | 0.980 | | |
| *MEDHHINC000* | −0.028 (.000) | 0.973 | −0.033 (.000) | 0.968 | −0.035 (.000) | 0.966 |
| *PROP_CRIME_MIL* | 0.001 (.001) | 1.001 | 0.001 (.000) | 1.001 | 0.001(.000) | 1.001 |
| Economy | | | | | | |
| *TOP25CITY* | −0.358 (.032) | 0.699 | −0.564 (.002) | 0.569 | −0.419 (.011) | 0.658 |
| *DC* | −1.158 (.115) | 0.314 | −1.420 (.056) | 0.242 | −1.247 (.091) | 0.287 |
| *REGION* | unreported | unreported | unreported | unreported | unreported | unreported |
| *AVG_PRICE_6* | 0.004 (.790) | 1.004 | 0.005 (.742) | 1.005 | | |
| Sustainability | | | | | | |
| *COMMUTE TIME* | | | 0.041 (.000) | 1.042 | 0.037 (.000) | 1.037 |
| *SUBWAY30* | | | −0.821 (.014) | 0.440 | −0.878 (.008) | 0.416 |
| *PCTWALK* | | | −0.031 (.008) | 0.969 | −0.031 (.009) | 0.969 |
| *RETAIL16* | | | −0.417 (.002) | 0.659 | −0.421 (.002) | 0.656 |
| *AFFORDABLE* | | | −0.959 (.000) | 0.383 | −0.964 (.000) | 0.381 |
| *FREEWAY1000FT* | | | 0.455 (.044) | 1.576 | 0.464 (.040) | 1.590 |
| *PROTECTED1MILE* | | | −0.401 (.009) | 0.669 | −0.393 (.010) | 0.675 |
| Constant | 24.620 (.000) | 4.926E10 | 25.841 (.000) | 1.670E11 | 25.492 (.000) | 1.178E11 |

*Notes:* In all models, $n$ = 37,385. In Model 1, chi-square = 549.54 (.000), −2 log-likelihood = 3,097.149, Nagelkerke $R^2$ = .157, and under ROC curve = 0.829. In Model 2, chi-square = 625.55 (.000), −2 log-likelihood = 3,019.951, Nagelkerke $R^2$ = .179, and under ROC curve = 0.841. In Model 3, chi-square = 621.54 (.000), −2 log-likelihood = 3,023.967, Nagelkerke $R^2$ = .178, and under ROC curve = 0.841

unrelated to sustainability. The second model repeats the first but adds the sustainability variables. The third model is a reduced version of the second. It drops insignificant variables to produce a more parsimonious model in order 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). Insignificant variables were kept in the second model so their effect on the sustainability variables could be considered.

Some variables were excluded from the three final models due to collinearity issues. *LOAN_AGE* was insignificant and had the wrong sign when it was included in the models with *USPRICE_CHANGE*. This was corrected when *LOAN_AGE* was used without *USPRICE_CHANGE*. *LOAN_AGE* was kept instead of *US_PRICE CHANGE* because it captured the information included in *USPRICE_CHANGE* (since they were correlated), and because it captured the effect of other forces that may have affected *DEFAULT* due to the seasoning of the loan and the market conditions when and since the loan entered the market.

*TOT_UNTS_CNT* was insignificant when included with *LOAN_SIZE_GP*; however both were significant on their own. *TOT_UNTS_CNT* was used in the final models because it had more informational content than *LOAN_SIZE_GP*, which was only ordinal rather than continuous.

The coefficient for *SUBWAY30* was inflated when *MIDATLANTIC* and *NYC* were in the model, so both were excluded, which produced more conservative results (smaller effects) for *SUBWAY30* in both cases. *NYC* was reintroduced in a robustness check, which is discussed below.

Since collinearity tends to produce unreasonably high regression coefficients, a rough indication of remaining collinearity in the models would be if any of the unstandardized coefficients ($\beta$) were greater than 2 (Menard, 1995). This was not the case, indicating the steps taken to reduce collinearity were sufficiently effective.

In the final models, all the variables had the expected signs except for *URB_RUR*; however its results were insignificant. Also, larger properties had smaller default rates, supporting the findings by Bradley, Cutts, and Follain (2000).

## Goodness-of-Fit with and without Sustainability

The first hypothesis was that if certain sustainability features are added to a model of default risk, the accuracy of the model will improve. This hypothesis was tested by comparing the goodness-of-fit of models with and without sustainability features included. Goodness-of-fit refers to how well all the explanatory variables in a logistic regression model, taken together, predict the dependent variable.

Goodness-of-fit statistics are reported in the notes for Exhibit 2. For all of the statistics except $-2$ log-likelihood, a higher value indicates a better fitting model. All four statistics show there was less discrepancy between the observed values for *DEFAULT* and the values produced by the model when sustainability features were included. That supports the acceptance of Hypothesis 1.

The model chi-square measures the total reduction for all the cases in default prediction errors that occurs when the independent variables are in the model, compared to when they are not. Comparing the model chi-squares across the models in Exhibit 2 indicates that the models with the sustainability features predicted default more accurately than the model without them. This is indicated by a chi-square of 550 for Model 1 versus 626 and 622 for Models 2 and 3, respectively.

The −2 log-likelihood statistic is similar to the model chi-square since it is based on the total error made by the model in predicting default for all the cases combined. But −2 log-likelihood measures the total error made by the model with the independent variables included rather than the difference between the error with and without the independent variables. That is, it measures how poorly a model fits the data with all the independent variables in the equation. That is why a better model has a smaller −2 log-likelihood. Here again, comparing across the models in Exhibit 2, the results show that the models with sustainability variables more accurately predicted default than the model without them, as indicated by −2 log-likelihood values of 3,020 for Model 1 versus 3,024 and 3,097 for Models 2 and 3.

The Nagelkerke R-Square is a "pseudo R-square" that can be computed in logistic regression analysis. Pseudo R-squares are not analogous to the R-square in linear regression and are easily misinterpreted by readers familiar with linear regression models, according to Hosmer and Lemeshow (2000). For that reason, they recommend against reporting them. The Nagelkerke R-square is a measure of improvement from the null model to the fitted model (i.e., the improvement in each model produced by adding the independent variables). It is most useful for comparing multiple models predicting the same outcome with the same dataset, as in the present case. When used that way, the models with the higher R-squares are the ones that better predict the outcome. As Exhibit 2 shows, the models with sustainability features included had higher Nagelkerke R-squares (.178 and .179 vs. .157).

Goodness-of-fit was also tested using the area under the receiver operating characteristic (ROC) curve. This test measures the model's ability to discriminate between loans that do and do not default and is 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. In the present study, ROCs exceeded 0.8. Values at this level indicate excellent discrimination according to Hosmer and Lemeshow (2000). In other words, the models did an excellent job distinguishing between loans that will and will not default. Here again, the models with sustainability features did a better job than the model without (0.841 for both models with sustainability vs. 0.829 for the model without).

The Hosmer–Lemeshow test is another commonly recommended statistical test for goodness-of-fit in logistic regression models. However, its assumption that the expected frequencies are large was violated because default was a relatively rare event in this study, so its results would be invalid in the present case (Hosmer and Lemeshow, 2000).

According to all four goodness-of-fit measures, the results support the first hypothesis that if certain sustainability features are added to a model of default risk, the accuracy of the model improves.

## *Interpretation of Sustainability Coefficients*

The second hypothesis was that sustainability features will be associated with a lower risk of default. This hypothesis was tested by examining and interpreting the regression coefficients in the reduced model (Model 3). According to the significance tests ("sig." in Exhibit 2), all of the sustainability features were significantly related to default risk. This means it is highly improbable that the sustainability features would be so strongly related to default in a model with this sample size if there actually were no relationships.

The size and direction of the relationships are indicated by the unstandardized coefficients ($\beta$). In Models 2 and 3, $\beta$ gives the change in the risk of default associated with a one-unit change in the sustainability variables, while the control variables are held constant. If $\beta$ is positive, then default risk increases with each one-unit increase in the sustainability variable. For example, in Model 3, a $\beta$ of 0.041 for *COMMUTE TIME* indicates that the risk of default increases as the average commute time increases in the area where the property is located. If $\beta$ is negative, then default risk decreases with each one-unit increase in the sustainability variable. For example, in Model 3, a $\beta$ of $-0.878$ for *SUBWAY30* indicates that the risk of default decreases when a property is located where 30% or more of the people commute to work by subway or elevated train. And since these findings were estimated when the control variables were also in the model, we can say they are true regardless of neighborhood income (*MEDHHINC000*), whether or not the loan is adjustable (*ARM_FLAG*), the debt service coverage ratio at origination (*ODSCR*), etc.

Together, the significance tests and unstandardized coefficients in Models 2 and 3 indicate that when the sustainability features studied here are present, there is less risk of default, all else being equal. Note, however, that for some of the sustainability variables, a larger raw score indicates the property is less sustainable. That is the case for *COMMUTE TIME* and *FREEWAY1000FT*. For these variables, a positive $\beta$ means that a lower value (which indicates less sustainability), such as a longer commute time, is associated with less default risk. For the other variables, a negative $\beta$ means that a higher value (which indicates more sustainability), such as more walking to work, is associated with less default risk.

The unstandardized coefficients can be used to obtain estimated odds ratios by exponentiating the coefficients (i.e., computing its base e anti-log). An odds ratio is the odds of an outcome in one group (e.g., the default rate for properties near a protected area) divided by the odds of an outcome in another group (e.g., the default rate for properties not near a protected area). It is analogous to relative risk (Grimes and Schulz, 2008). The odds ratio associated with each independent variable is given as Exp($\beta$) in Exhibit 2. As explained by Menard (1995), the odds ratio is the number by which we would multiply the odds of default for each one-unit increase in the independent variable. An Exp($\beta$) greater than 1 indicates

**Exhibit 3** | Summary of Sustainability Effects on Default Risk in Multifamily Mortgages

| Variable | Definition | Effect on Relative Risk of Default |
|---|---|---|
| *COMMUTE TIME* | Per 1 minute increase in the commute time to work. | 3.7% more |
| *SUBWAY30* | Where ≥ 30% commute by subway or elevated. | 58.4% less |
| *PCTWALK* | Per 1-unit increase in the percent who walk to work. | 3.1% less |
| *RETAIL16* | Where there are 16 or more retail establishments in the block group. | 34.4% less |
| *AFFORDABLE* | When property meets FNMA definition for affordable housing. | 61.9% less |
| *FREEWAY1000FT* | Where property is located within 1,000 feet of a freeway corridor. | 59.0% more |
| *PROTECTED1MILE* | Where property is located with 1 mile of protected open space. | 32.5% less |

that the odds of default increase when the independent variable increases and an Exp($\beta$) less than 1 indicates that the odds of default decrease when the independent variable increases. For example, a one-unit increase in *COMMUTE TIME* (i.e., a one-minute increase, according to the definition column in Exhibit 1) results in a 3.7% increase in the odds of default (the odds of default are multiplied by 1.037). Similarly, a one-unit increase in *SUBWAY30* results in a 58.4% decrease in the odds of default (the odds of *DEFAULT* are multiplied by 0.584, which is 0.416 less than 1). For a dummy variable that can only be scored as 0 or 1, a one-unit increase is equivalent to referring to all cases where the dummy variable has a score of 1 or Yes. So, for example, for *SUBWAY30*, if we say that a one-unit increase decreases the odds of default by 58.4%, we are saying that when *SUBWAY30* is scored 1 (i.e., when a property is located in a location where 30% or more of the people commute by subway or elevated to work), the odds of default are 58.4% lower than in cases where the property is not located in such a location and *SUBWAY30* is scored 0.

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). Relative risk is the ratio of the probability of an event occurring in a group with and without a certain characteristic. We can say, for example, that in locations where commute time is 28 minutes (or about one minute above average), multifamily mortgages are 3.7% more likely to default than in locations with a 27-minute commute time. Similarly, in locations where 30% or more residents take a subway or elevated train to work, owners are 58% less likely to default in comparison to locations where fewer than 30% commute by subway or elevated. And, as noted above, the reductions in risk associated with the each of the sustainability features are unrelated to differences in regional location, neighborhood income, and the other control variables included in the models.

Exhibit 3 summarizes the effects of a one-unit change in the sustainability variables on default risk in relative risk terms. In every case, the effects are large, indicating that these sustainability features have a very significant effect on default risk, independent of other factors commonly used to predict default. In addition, the direction of each relationship indicates that in all cases more sustainability is related to less default risk. These findings confirm Hypothesis 2 that sustainability features are associated with a lower risk of default.

When dealing with continuous variables, such as *COMMUTE TIME* and *PCTWALK*, a one-unit change is not always interesting. For example, a one-minute increase in commuting or a 1% increase in the percentage of people who walk to work may be too small to be considered important. Hosmer and Lemeshow (2000, p. 63) show that as in the case of a one-unit increase, the effect from a multi-unit change in an independent variable on the odds of default can be determined from an estimated odds ratio. In the case of a multi-unit change, the odds ratio is estimated by exponentiating the product of the unstandardized coefficient ($\beta$) for a given variable times the number of units of change. If, for example, one were interested in how a 10-minute increase in *COMMUTE TIME* affected the odds of default, the odds ratio would be computed by exponentiating the product of $10 \times \beta$ for *COMMUTE TIME*. Applying this method using the coefficients from Model 3 (Exhibit 2), we find that for every 10-minute increase in the mean commute time for residents in a census tract, the risk of default increases by 45%. Similarly, for every five-unit increase in the percentage of people in a tract who walk to work, the risk of default decreases by 15%.

As discussed above, *NYC* was excluded from the models. This was because *SUBWAY30* was inflated with *NYC* in the model due to collinearity issues. However, it was important to know if the effects of *SUBWAY30* were due to its association with *NYC*. To answer that question, Model 3 was rerun using the 33,733 cases that were not located in *NYC*. The unstandardized coefficient for *SUBWAY30* ($\beta$) was $-1.062$ (0.003) and the Exp($\beta$) was 0.356. Compared to the results in Exhibit 2, this indicates that the effect of *SUBWAY30* outside of *NYC* was even stronger. When cases in New York City were excluded from the sample, the relative risk of default was 64.4% less for properties located where at least 30% of the tract residents commuted by subway or elevated, compared to 58.4% less when cases in New York City were included in the sample. This indicates that the effects of *SUBWAY30* on default were not because a disproportionate share of the cases in subway-oriented locations were in New York City.

## Discussion

The findings support the hypotheses. When certain transportation-, location-, and affordability-related sustainability features are included in a default probability model for multifamily housing, the model predicts default more accurately. Also, properties with the sustainable features studied are much less likely to default. These findings suggest two important implications for practice: one pertaining to sustainability and the other to risk management.

First, certain aspects of sustainability can be fostered without increasing default risk by adjusting conventional lending standards for properties with the sustainability features studied here to achieve the same risk of default associated with less sustainable properties. For example, in the study sample, the mean *OLTV* and *ODSCR* were 0.61 and 1.52, respectively. However, according to Model 3, if a property is located in a subway-oriented census tract (where at least 30% of the people commute by subway or elevated train), then the *OLTV* could have been raised to nearly 0.75 and the *ODSCR* could have been lowered to 1.28 (which maintains the normal ratio between the two found in the sample) and the probability of default would have remained virtually equal to that found for properties not located near subway stations, all else being equal. According to the model, the extra risk produced by a higher *OLTV* and lower *ODSCR* would be offset by the lower risk produced by locating in a more sustainable location, leaving the total risk unchanged. Similar results can be produced with various other combinations of the studied sustainability features and conventional lending standards. Of course, lenders have programmatic constraints that set maximums for the *OLTV* and minimums for the *ODSCR*. Those constraints establish practical limits on how far lenders could go in adjusting for sustainability features unless they are willing to make exceptions to the programmatic constraints for that purpose.

If higher LTV ratios at origination could be obtained by borrowers for more sustainable properties, the borrowers would achieve a higher return on equity 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 asset). Ceteris paribus, this should cause investors to prefer more sustainable investments, increase capital flow to more sustainable buildings, and foster a transition to more sustainable cities.

The second implication of the findings is that lenders could improve their risk management practices by taking stock of whether a property has certain sustainability features when loans are originated. All the features studied here are based on existing federal datasets provided by the Census Bureau and other agencies. So it would be relatively easy to build an online address-based lookup tool that any lender can use to obtain certain sustainability information on a given property. In addition, recommended adjustments to *OLTV* and *ODSCR* ratios could be made available through a second online tool based on coefficients found in this study and confirmed by additional research. The results clearly show that traditional lending ratios have not fully recognized the presence or absence of certain sustainability features (by recognizing their effect on capitalization rates, values, or cash flows). The most common result is that these ratios have been based on an underestimation of the actual risk of default for properties without those sustainability features. This is because the "normal" rate of default expected for all properties is derived from the experience with more and less sustainable properties combined, without the effect of sustainability being recognized. In this situation, less sustainable properties, according to the findings in this study, should have a rate of risk higher than the norm. There is more risk associated with less sustainable properties (and less risk with more sustainable properties) than is being accounted for in traditional lending ratios. If better information could be made available on the size of these effects and on how lending ratios could be adjusted to offset the effects produced by the presence or absence of sustainability features, the risk of portfolios could be more effectively managed.

Unfortunately, this study is limited to those sustainability features for which data could be obtained. It is likely, however, that other sustainability features that were not studied here also have beneficial effects on default risk. The most likely examples pertain to energy efficiency and green building certification because prior studies show they affect cash flow and value. Other sustainability features that could reduce default risk include rental unit flexibility, urban centrality, noise mitigation, water efficiency, childcare services, and school quality. This is expected because of their likely ''materiality'' to financial performance, rather than their positive effect on the public good (Pivo, 2008). Further work on whether these features do or do not affect default risk would be most useful, but it will have to await the development of better databases to account for them.

Other fruitful avenues for further research include repeating this study using other modeling methods (e.g., longitudinal hazard models) and data sets on higher risk mortgages. It would also be useful to conduct similar studies on other property types, including single-family homes and other commercial property types. And work should begin on a practical tool that helps lenders adjust conventional lending ratios based on sustainability information without increasing normal risk levels.

## Conclusion

Perhaps the most important point of this study is that sustainability is just as much an issue with material consequence for investors as for those interested in social and environmental well-being. Pivo and McNamara (2005) wrote about the common ground emerging between real estate investing and sustainability stating: ''It is probably apparent to anyone who thoughtfully considers real estate that it can both contribute to and be affected by many of the social and environmental issues that face the world's societies. Until recently, however, most real estate investors would likely have said that while they are sympathetic, such issues are...not of direct concern to their investment practices. But today, a new view is emerging...that various social and environmental issues can have significant material consequences for their investment portfolios.''

Sustainability is financially consequential for property investors. Multifamily lenders can mitigate risk by gearing their portfolios toward more sustainable properties. Moreover, and for society this may be even more important, lenders can offer more favorable terms to more sustainable properties without increasing default risk, or as Benjamin Franklin once said, they can ''do well by doing good.''

## Appendix

### Logistic Regression and Alternatives

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, such as pass/fail, or in the present study, default/no default. 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 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.

Logistic regression models are typically used to determine whether the classification of cases into one of the two categories of the dependent variable can be predicted better from information on the independent variables than if the cases were randomly classified into the categories. ''Goodness-of-fit'' refers to how well all the explanatory variables in a logistic regression model, taken together, predict the dependent variable. So, in terms of logistic regression, the first hypothesis in this paper is that a logistic regression model for default will have a better goodness-of-fit if sustainability features are included along with other more conventional explanatory variables in the regression equation.

The estimated parameters of a logistic regression equation can be interpreted as the change in the dependent variable that can be expected from a one-unit change in the dependent variable, holding other dependent variables constant. So again, looking at the second hypothesis in this paper in terms of logistic regression analysis, the expectation is that sustainability features will have estimated parameters that show they reduce the probability of default by a significant amount.

The most common alternative to the logistic regression model in mortgage default research is the proportional hazard model. Hazard models explain the time that passes before some event occurs in terms of covariates associated with that quantity of time. They have been used to estimate the probability that a mortgage with certain characteristics will default in a given period if there has been no default up until that period (Vandell, Barnes, Hartzell, Kraft, and Wendt, 1993; Ciochetti, Deng, Gao, and Yao, 2002; Teo, 2004). Other methods for predicting default also have been explored including neural networks (Episcopos, Pericli, and Hu, 1998) and a maximum entropy approach (Stokes and Gloy, 2007).

A common view of the hazard model is that it is less sensitive to bias from database censuring than logistic regression. Censoring occurs when cases are removed from the database prior to observation (e.g., when a loan is paid off or foreclosed and sold prior to observation) or when the event of interest happens after observation occurs (e.g., when a loan defaults after the study observation date). However, as pointed out by Archer, Elmer, Harrison, and Ling (2002), bias is only an issue in logistic regression when the explanatory variables have a different effect on the censored and uncensored cases. In the present study, there is no reason to expect that sustainability features affected the odds of default differently in censored and uncensored cases. Hazard models also require a time series dataset that reports the occurrence of defaults over time and such a dataset was unavailable at the start of the present study. One effort to predict mortgage pre-payment using both approaches found that the logistic regression model made better predictions (Pericli, Hu, and Masri, 1996), while in another study on insolvency among insurers, the two models produced equally accurate predictions (Lee and Urrutia, 1996). So, while it would be interesting to repeat this study using a hazard model, there is no a priori reason to assume that the logistic regression method used here produced results that are inferior to those that would have come from another method.

## References


Archer, W.R., P.J. Elmer, D.M. Harrison, and D.C. Ling. Determinants of Multifamily Mortgage Default. *Real Estate Economics*, 2002, 30:3, 445–73.

Bradley, D.S., A.C. Cutts, and J.R. Follain. An Examination of Mortgage Debt Characteristics and Financial Risk among Multifamily Properties. *Journal of Housing Economics*, 2001, 10, 482–506.

Bratt, R.G. Housing and Family Well-Being. *Housing Studies*, 2002, 17:1, 13–26.

Cakmak, S., M. Mahmud, A. Grgicak-Mannion, and R.E. Dales. The Influence of Neighborhood Traffic Density on the Respiratory Health of Elementary Schoolchildren. *Environment International*, 2012, 39:1, 128–32.

Chandan, S. Banks' Commercial Mortgage Default Rates Fell—Now What. *New York Observer*, March 3, 2011.

Cramer, J.S. Robustness of Logit Analysis: Unobserved Heterogeneity and Mis-specified Disturbances. *Oxford Bulletin of Economics and Statistics*, 2007, 69:4, 545–55.

Ciochetti, B.A., Y. Deng, B. Gao, and R. Yao. The Termination of Commercial Mortgage Contracts through Prepayment and Default: A Proportional Hazard Approach with Competing Risks. *Real Estate Economics*, 2002, 30:4, 595–633.

Dietrich, J. Under-specified Models and Detection of Discrimination in Mortgage Lending. Office of the Comptroller of the Currency. Economic and Policy Analysis Working Paper 2003-02, 2003.

Eichholtz, P., N. Kok, and J.M. Quigley. *Doing Well by Doing Good? An Analysis of the Financial Performance of Green Office Buildings in the USA*. Royal Institute of Chartered Surveyors, London, 2009.

Episcopos, A., A. Pericli, and J. Hu. Commercial Mortgage Default: A Comparison of Logit with Radial Basis Function Networks. *Journal of Real Estate Finance and Economics*, 1998, 17:2, 163–78.

Esaki, H., S. L'Heureux, and M. Snyderman. Commercial Mortgage Defaults: An Update. *Real Estate Finance*, 1999, 16:1, 80–6.

Federal Highway Administration. *Health and Environmental Benefits of Walking and Bicycling*. fhwa.dot.gov, 2012.

Gauderman, W.J., H. Vora, R. McConnell, K. Berhane, F. Filliland, D. Thomas, F. Lurmann, E. Avol, N. Kunzli, M. Jerrett, and J. Peters. Effect of Exposure to Traffic on Lung Development from 10 to 18 Years of Age: A Cohort Study. *The Lancet*, 2007, 369:9561, 571–77.

Goldberg, L. and C.A. Capone. Multifamily Mortgage Credit Risk: Lessons From Recent History. *Cityscape*, 1998, 4:1, 93–113.

Goldberg, L. and C.A. Capone. A Dynamic Double-Trigger Model of Multifamily Mortgage Default. *Real Estate Economics*, 2002, 30:1, 85–113.

Grimes, D.A. and K.F. Schulz. Making Sense of Odds and Odds Ratios. *Obstetrics and Gynecology*, 2008, 111:2, 423–26.

Grovenstein, R.A., J.P. Harding, C.F. Sirmans, S. Thebpanya, and G.K. Turnbull. Commercial Mortgage Underwriting: How Well Do Lenders Manage the Risks? *Journal of Housing Economics*, 2005, 14, 355–83.

Haas, P.M., C. Makarewicz, A. Benedict, and S. Bernstein. Estimating Transportation Costs by Characteristics of Neighborhood and Household. *Transportation Research Record*, 2008, 2077, 62–70.

Harrison, D., G. Smersh, and A. Schwartz. Environmental Determinants of Housing Prices: The Impact of Flood Zone Status. *Journal of Real Estate Research*, 2001, 21, 3–20.

Hosmer, D.W. and S. Lemeshow. *Applied Logistic Regression*. Second edition. New York: John Wiley & Sons, Inc., 2000.

Kau, J.B., D.C. Keenan, W.J. Muller, and J.F. Epperson. Pricing Commercial Mortgages and their Mortgage-Backed Securities. *Journal of Real Estate Finance and Economics*, 1990, 3, 333–56.

Lee, S.H. and J.L. Urrutia. Analysis and Prediction of Insolvency in the Property-Liability Insurance Industry: A Comparison of Logit and Hazard Models. *Journal of Risk and Insurance*, 1996, 63:1, 121–30.

McGreal, S., J.R. Webb, A. Adair, and J. Berry. Risk and Diversification for Regeneration/Urban Renewal Properties: Evidence from the U.K. *Journal of Real Estate Portfolio Management*, 2006, 12:1, 1–12.

Menard, S. Applied Logistic Regression. Sage University paper series on Quantitative Applications in the Social Sciences, 07-106. Thousand Oaks, CA: Sage, 1995.

Miller, N., J. Spivey, and A. Florance. Does Green Pay Off? J*ournal of Real Estate Portfolio Management*, 2008, 14:4, 385–400.

Office of Health Hazard Assessment. Health Effects of Diesel Exhaust. http://oehha.ca.gov, 2012.

Pericli, A., J. Hu, and J. Masri. The Prepayment of Fixed-Rate Mortgages: A Comparison of Logit with Proportional Hazard Models (June 25, 1996). Available at SSRN: http://ssrn.com.ezproxy2.library.arizona.edu/abstract=7982.

Pivo, G. Responsible Property Investment Criteria Developed Using the Delphi Method. *Building Research and Information*, 2008, 36:1, 30–6.

Pivo, G. and J. Fisher. Income, Value, and Returns in Socially Responsible Office Properties. *Journal of Real Estate Research*, 2010, 32:3, 243–70.

——. The Walkability Premium in Commercial Real Estate Investments. *Real Estate Economics*, 2011, 39:2, 185–219.

Pivo, G. and P. McNamara. Responsible Property Investing. *International Real Estate Review*, 2005, 8:1, 128–43.

Rauterkus, S.Y., G.I. Thrall, and E. Hangen. Location Efficiency and Mortgage Default. *Journal of Sustainable Real Estate*, 2010, 2:1, 117–41.

Shilling, J., J. Benjamin, and C.F. Sirmans. Adjusting Comparable Sales for Floodplain Location. *The Appraisal Journal*, 1985, 53, 429–36.

Simons, R.A. and J.D. Saginor. A Meta-Analysis of the Effect of Environmental Contamination and Positive Amenities on Residential Real Estate Values. *Journal of Real Estate Research*, 2006, 28:1, 71–104.

Snyderman, M. Commercial Mortgages: Default Occurrence and Estimated Yield Impact. *Journal of Portfolio Management*, 1991, 18:1, 6–11.

Stokes, J.R. and B.A. Gloy. Mortgage Delinquency Migration: An Application of Maximum Entropy Econometrics. *Journal of Real Estate Portfolio Management*, 2007, 13:2, 153–60.

Teo, A.H.L. Delinquency Risk in Residential ARMs: A Hazard Function Approach. *Journal of Real Estate Portfolio Management*, 2004, 10:3, 243–58.

Titman, S. and W. Torous. Valuing Commercial Mortgages: An Empirical Investigation of the Contingent-Claims Approach to Pricing Risky Debt. *The Journal of Finance*, 1989, 59, 165–206.

Vandell, K.D. On the Assessment of Default Risk in Commercial Mortgage Lending. *Journal of the American Real Estate and Urban Economic Association*, 1984, 12:3, 270–96.

——. Predicting Commercial Mortgage Foreclosure Experience. *Real Estate Economics*, 1992, 20:1, 55–88.

Vandell, K.D. and J.S. Lane. The Economics of Architecture and Urban Design: Some Preliminary Findings. *Real Estate Economics*, 1989, 17:2, 235–60.

Vandell, K., W. Barnes, D. Hartzell, D. Kraft, and W. Wendt. Commercial Mortgage Defaults: Proportional Hazards Estimation Using Individual Loan Histories. *Real Estate Economics*, 1993, 4:21, 451–80.

Volk, H.E., I. Hertz-Picciotto, L. Delwiche, F. Lurmann, and R. McConnell. Residential Proximity to Freeways and Autism in the CHARGE Study. *Environmental Health Perspectives*, 2011, 119:6, 873–77.

Williams, B.A., J.N. Mandrekar, S.J. Mandrekar, S.S. Cha, and A.F. Furth. Finding Optimal Cutpoints for Continuous Covariates with Binary and Time-to-Event Outcomes. Technical Report Series #79. Department of Health Sciences Research. Mayo Clinic, Rochester, MN, June 2006.

Wolch, J., M. Jerrett, K. Reynolds, R. McConnell, R. Chang, N. Dahmann, F. Filliland, J.G. Su, and K. Berhane. Childhood Obesity and Proximity to Urban Parks and Recreational Resources: A Longitudinal Cohort Study. *Health and Place*, 2011, 17:1, 207–14.

Gary Pivo, University of Arizona, Tucson, AZ 85721 or gpivo@u.arizona.edu.