Using MLS Data to Predict Residential Foreclosure

Ronald C. Rutherford

Elmo J. Burke Jr. Endowed Chair in Building/Development

Professor of Finance and Real Estate

University of Texas, San Antonio

Thomas A. Thomson

Associate Professor of Finance and Real Estate

University of Texas, San Antonio

6900 North Loop 1604 West

San Antonio, TX 78249-0637

tthomson@utsa.edu

210-458-5306 (voice)

210-458-6320 (fax)

Using MLS Data to Predict Residential Foreclosure

Abstract

Mortgage loan default studies typically evaluate the effects of the mortgage contract (fixed versus variable rates, term, down payment), the borrower characteristics (credit score, education, employment), the economic conditions, the legal situation or some combination of these. This paper takes an alternate approach by examining the contribution Multiple Listing Service (MLS) data can make to the understanding of mortgage default. We find: i) Houses that eventually foreclosed sold for about 3% less than predicted by a hedonic model, ii) Houses that eventually foreclosed sold at less of a discount to list price than houses that did not, iii) Houses that eventually foreclosed took about 12% longer to sell than other houses, and iv) A hedonic model based predicted house price is a very strong covariate of future foreclosure.

Using MLS Data to Predict Residential Foreclosure

1. Introduction

Mortgage default continues to be a much-studied part of residential mortgage finance. An overall goal is to understand factors that lead to default so that loan underwriting can be done more accurately and mortgages can be appropriately priced to their underlying risk. Early research on mortgage default evaluated empirical factors that lead to default (see for example von Furstenberg 1969 or Gau 1978). The next line of research focused on theoretical models of mortgage default models based on options analysis that show when the equity in a home hits a certain negative point, the best option for the borrower is to default on that loan (see for example, Kau et al 1994 or Capozza et al 1998). Empirical models have been refined in response to insights from option pricing models, and options pricing models have been refined to introduce more realism. The major purchasers of residential mortgages have developed automated underwriting models that presumably include the insights learned from this research. Until more recently, there has been less emphasis in determining which properties are more prone to default.

A large body of empirical mortgage studies confirms that most defaults occur on loans with low down payments and in areas where house prices have been flat or fallen (see for example, Capozza et al. 1997 or Ambrose and Deng 2001). Where low down payments intersect with a falling house price, a substantial number of borrowers may have negative equity and some of these borrowers will default on their mortgage. Mortgage default studies, while accepting the pivotal role of the importance of equity in the default decision, implicitly accept the starting loan to value ratio as determined by the initial down payment. Some recent studies, however, address

the role of hedonic appraisal models in helping to understand default. Our study continues this line of research, by using MLS data to examine the relationship between a hedonically estimated house value and default, and then proceeds to explore some additional questions.

Given the well known results from mortgage default studies (both empirical and theoretical) that negative equity is a primary determinant of default, it is reasonable to conjecture that when the purchase price is above the true market value of the house, one will see a greater tendency for such houses to have their mortgages default as their true equity is less than is believed at loan origination. In an early study of mortgage default, Gau (1978) found that the ratio of the appraised value to the purchase price to be a mildly useful default covariate. A direct evaluation of actual selling price compared to a hedonic model appraisal is investigated in a recent study by LaCour-Little and Malpezzi (2003, hereafter LLM). Using a sample of 113 defaults from a single credit union in Alaska during a time of volatile house prices (a rapid rise followed by a rapid decline), they find that houses that defaulted tended to be those that a simple hedonic model would have shown to be overvalued at the time of purchase. They incorporate a measure of over appraisal into a hazard model of default and conclude that this variable helps explain default – that is, houses that are appraised at an amount above that determined by a hedonic appraisal model, are more likely to default. They conclude, "appraisal does matter, at least in the high volatility housing market of Alaska during the 1980's, to subsequent mortgage default probability." (p. 229). More particularly they find that "over-appraisal is significantly related to default, whereas under-appraisal has no effect." (p. 229). Due to the somewhat unique situation and small data sample they have, they suggest further work in this area is appropriate to

determine if this phenomena is wide spread. It is reasonable to speculate whether their finding represents an unusual case or a more general problem.

Lending support to the above study, Ong, Neo, and Spieler (2004, hereafter ONS) describe a somewhat similar study, which reports a parallel outcome. Their sample may also be considered somewhat unusual as it deals primarily with condominium units in Singapore. During their study period, the property price index, after rising rapidly through the early nineties, fell by about 1/3 in the late nineties. Their sample consisted of 136 foreclosures on high-rise properties and 149 foreclosures on low-rise properties, which they analyzed separately. They find that the average price per square foot for foreclosed properties is higher than that for non-foreclosed properties. To control for other factors they assess the residuals from a hedonic model and find that foreclosed properties had a statistically significant 9% premium. In logit models for predicting foreclosure, they find statistical support for using the appraisal premium as a predictor of foreclosure. ONS also address whether foreclosed properties sell at a discount in a subsequent sale, and find statistically significant discounts of about 3% and 7% respectively for high-rise and low-rise buildings. They suggest that the greater foreclosure probability from the over priced properties is a market disciplining effect of the poor initial decision. In both the ONS and the LLM study , there was a period of rapid price appreciation, followed by a period of sharply falling prices. As noted by LLM, "While this phenomena is sometimes observed in Real Estate markets, it is not the norm."

Capozza et al (2005) take a different tact in studying appraisal and default. They find that the variables that matter in a hedonic appraisal for normal sale may not be the same as those that are

most appropriate for appraising a unit for a distressed sale. Using manufactured housing data they determine that an atypical unit, compared to other units, sells at a premium in a normal market. At a foreclosure sale, however, they find that an atypical unit sells at a discount. They also find that the relative values placed on the features of a property are not the same for a normal sale versus a distressed sale. They infer that a lender could be better off appraising a property using its distressed sale characteristics rather than its normal market characteristics if mitigating the loss on reposed properties is an important concern of lenders.

There are also some facts about the house selling process that have not been addressed in a foreclosure context but may impact the interplay between house selling and default. Haurin (1988) addresses the marketing time for houses in a search context and suggests that atypical houses will take longer to sell. "Charm prices," or their opposite, which have been referred to as "off-dollar" pricing assesses the effects for houses that are not listed at prices ending in 000, 500, or 900 (see Allen and Dare 2004, Palmon et al 2004 and Salter et al 2005). These studies indicate that choice of list price effects the time it takes houses to sell, while the effect on selling price remains empirically less established.

We address a related set of concerns in this paper, while taking a distinctly different approach than the studies cited. Our primary objective is to determine what Multiple Listing Service (MLS) sales data can inform us about future foreclosure. We first address the use of hedonic models for identifying foreclosures. The main difference vis a vis LLM and ONS is that we use MLS sales data from a large U. S. metropolitan area encompassing many lenders. We address whether houses that eventually foreclosed were more likely to have been over-priced at their

initial sale date, relative to the value estimated using a hedonic model. We then address the difference between list and sale price to detect any difference for homes that eventually foreclose. In these inquiries, in addition to using the standard hedonic pricing covariates, we use atypicality and off-dollar pricing to control for any effect these variables may have. We then examine the relationship between marketing time and foreclosure. One reason a house may end up in foreclosure is that during the delinquency period that precedes foreclosure, the house could not be sold to satisfy the mortgage. We addresses whether houses that eventually foreclosured were slower to sell initially. Finally we employ logistic regression models to directly assess factors that lead to foreclosure, including, atypicality, off dollar pricing, hedonic pricing model residuals, actual and predicted selling price, and discounts from list price.

In the following sections of this paper we describe our data collection, our statistical techniques and then present the results. A primary finding is that homes that went through foreclosure are more likely to be under-priced relative to what would be expected from a hedonic model using MLS sales data. This result is the opposite of the prior two studies that detected a price premium on units that foreclosed. We confirm some results regarding off dollar pricing and have mixed results on atypicality. We find two predictors of foreclosure that have not hitherto been reported. The first result is that the longer a property takes to sell, the more likely it will end up in foreclosure. Second, the better the "deal", that is, the lower the ratio of sales price to list price, the less likely the property is to foreclose.

2. The Data

We focus our analysis on houses that are known to enter foreclosure sometime after an initial sale. We evaluate MLS data from Tarrant County, a part of the Dallas–Fort Worth metropolitan area, over the period 1998-2000. We were able to identify 911 houses that were identified as foreclosures for which we could find a corresponding prior sales record during the 1998-2000 period. A primary question is whether these houses had a statistically different value at the original sale compared to other houses during that period. Given previous research findings, we wish to determine if they were systematically over-priced at their first sale date, as measured by a hedonic model. To complete the hedonic model we drew an additional large sample of houses that did not have a future foreclosure. We use a sample of about 130,000 other houses to develop the hedonic price model and serve as control data.

Our analysis variables, along with their descriptive statistics, are provided are Table 1. The variables for the most part are commonly used in hedonic house price models and are self-explanatory. In addition to the variables presented in Table 1, we use a control set of set of 11 quarterly time dummies, and 87 location dummies that we do not report.

As noted earlier, Capozza et al. (2005) find interesting results regarding the interplay between atypicality and estimated value. To determine how this might play out in a very different data set we follow Haurin (1988) and Capozza et al. (2005) to create a variable to account for any disparities that arise in valuing homes that are unusual. We use implicit marginal prices from a hedonic regression of home sales prices on various characteristics to penalize absolute deviations

from the average. We then aggregate these values. Our measure of Atypical for the ith home is as follows:

$$
\begin{aligned}
Atypical_i = & \left| P_{SqFt} \cdot \left(SqFt_i - \overline{SqFt} \right) \right| + \left| P_{Age} \cdot \left(Age_i - \overline{Age} \right) \right| \\
& + \left| P_{Bedrooms} \cdot \left(Bedrooms_i - \overline{Bedrooms} \right) \right| + ...
\end{aligned}
\tag{1}
$$

where P_{SqFt}, P_{Age} , and $P_{Bedrooms}$, etc. are estimated regression coefficients, analogous to those reported in Table 2, from a first pass hedonic regression of the log of sales price on the physical characteristics of the houses, and where $\overline{SqFt}$, $\overline{Age}$, and $\overline{Bedrooms}$ etc, are the sample means from Table 1 for area in 100 Square feet, home age in decades, the number of bedrooms, and so on.

Also as noted above, we want to address the effect that "off-dollar" pricing may have on selling price and foreclosure. In particular, the dummy variables "TripleZero", "FiveHundred", and "NineHundred" represent that the listing price ended in 000, 500, or 900 respectively. The sum of the means of these variables shows that 12 percent of the houses were listed at "off dollar" prices, that is, prices not covered by these three indicator variables.

3. Results

We begin by presenting a cross tab plot of the foreclosure rate versus log of house selling price decile (Figure 1). This plot shows that defaults are concentrated in the lower cost houses indicating we must be sensitive to any distortion in results that could be due to the simple fact that low cost houses are more likely to default.

Table 2, Model 1 presents a standard hedonic house price model. The physical characteristics

which affect value, are as expected. Larger houses, newer houses with more bathrooms, a pool,

a fireplace, and a large lot are more valuable. More bedrooms, in the same amount of space, or a

multistory house detract from value. A vacant house, or a house with a tenant sells for less.

The effect of listing price indicates that houses listed at off-dollar amounts sell for statistically

less than similar houses. Houses listed at the five hundreds sell for just a little more, and houses

listed at the even thousand (triple zero) sell for almost 3% more. This finding lends support to

Palmon et al (2004) who find a small premium for triple zero and a non-significant, but small

premium for nine hundreds. The results here show similar signs, but stronger magnitudes. Salter

et al (2005) find no effect of off dollar listing on sales price. The result here is the opposite of

Allen and Dare (2004) who find that houses listed at a triple zero sell for 7-10% less while

houses at the five hundred or nine hundred sell for about 5% more.

It is well known that the foreclosure rate for FHA VA loans is much higher than for conventional

loans. When a lender commits to grant a mortgage, that lender knows whether the chosen

instrument is an FHA VA loan. The hedonic model shows such houses are worth almost 5% less

than houses financed with a conventional loan. This result suggests that buyers who choose

FHA VA loans are simultaneously choosing houses that are less valuable than other houses that

share a similar set of physical attributes.

We also evaluate the effects of atypicality. The estimated coefficient for Atypical is statistically

negative, providing the opposite result of Capozza et al. (2005) in their study of manufactured

houses. This result acknowledges that atypicality plays a role in house valuation, but in contrast

to Capozza et al. (2005) this result indicates that atypical houses sell at a discount at the outset.

Within the hedonic regression model, we include an indicator variable, "Foreclose", that takes

the value 1 if the house is identified as a future foreclosure and the value zero otherwise. The

sign and statistical significance of the estimated coefficient for this variable indicates over or

under-pricing. The results of LLM and ONS imply that such a coefficient will be positive and

statistically significant. A similar result would indicate that houses that are initially over-priced,

vis a vis a hedonic model, are the ones more often observed in a future foreclosure. Our results

show a statistically significant coefficient of about -2.7% for the Foreclose covariate. In other

words, houses that end up in foreclosure appear to be under-priced by about 2.7% at their initial

sale – a result that contradicts the earlier studies of LLM and ONS. Figure 2 plots the quarterly

dummies from this model showing that sales prices were steadily, albeit moderately, increasing

over the study period.

In response to the earlier noted concern that house price may play an important role, we plotted

the residuals of the Table2, Model 1 regression against house price decile and found a trend for

low and high price houses. In particular, Model 1 tends to over predict the house price of the

lowest decile (PriceDecile 1), and under predict the price of the two highest deciles (PriceDecile

9, PriceDecile 10). This result is entirely reasonable as houses in the lowest decile are selling at

prices below that expected given their size, age, etc. due to reasons that are not captured in the

model – perhaps due to their state of repair or micro location. On the other end of the spectrum,

house often sell above that expected for their size, age, etc., perhaps due to extra high quality finish, or prime location, which are not captured in the available covariates.

Table 2, Model 2 addresses this concern by including indicator variables for the lowest and two highest deciles, and by interacting these price decile indicators with the foreclosure indicator. The results for the price deciles are as expected, that is, the PriceDecile 1 indicator variable is significantly negative, and the upper price decile indicators are significantly positive. These variables, interacted with the Foreclosure variable demonstrate interesting results. For the lowest price decile, we find that houses that eventually foreclosed initially sold at a statistically significant premium of 2.5% to those that did not, a result that is in agreement with LLM and ONS. This holds, however, for only the lowest 10% of the houses. For the 70% of houses in the middle price zone, houses that defaulted sold at 4.8% discount to those that did not, indicating that the market had discounted the value of these houses at their original purchase date. For the top twenty percent of houses, we observe no statistically significant foreclosure effect suggesting foreclosure is a random event relative to these higher house prices.

With the indicator variables for the high or low priced houses, the model is freer to choose coefficients that better reflect the 70% of the houses in the middle. Some model coefficients show changes in response to this formulation. The FiveHundred price dummy is now reduced to near zero and is no longer statistically significant, indicating that such houses sell for no different amount than the off-dollar price homes. The NineHundred and TripleZero variables have reduced coefficients but remain significant. One coefficient with a rather large change is that for atypicality, whose magnitude doubled in the new formulation.

Table 2, Model 3 investigates whether the atypicality effect is linear. While Model 2 results indicated that highly atypical homes sell at a discount, its also possible that rather ordinary houses might also sell at a somewhat of a discount with this effect masked in a linear formulation of the variables. "Atyp-High" indicates the home is in the 10% that are furthest above the norm. "Atyp-MedHigh" indicates the home is in the next 15%. "Atyp-Low" indicates the home is in the 10% that deviate least from the average house, while "Atyp-MedLow" are the next 15%. The left out category represents the 50% of homes that remain in the middle. The regression coefficients of Model 3 indicate the atypicality effect is basically linear. Homes that are most atypical sell for 6.7% less while those that are unusually typical sell for 5.9% more. A similar, symmetry with smaller coefficients is observed for the medium high and medium low categories.

The results thus far demonstrate a robust finding that homes, in the middle 70% of the value range, that end up in foreclosure, sell at a discount to similar homes in the epoch prior to foreclosure. Homes in the lowest decile sell at a premium and those in the top 20% show no effect. These results suggest that a hedonic check appraisal may have limited value in determining which houses are more likely to default.

A related issue is not what the houses actually sold for, but what was the estimated value of the house as measured by the listing price In other words, does the "under-pricing" result for houses that will foreclose also apply to the price at which the house was listed? Under-pricing detected at the listing phase would indicate that houses that end up in foreclosure were viewed prior to being placed on the market as somehow flawed, as measured by the discount relative to a

hedonic model. Table2, Model 4 presents a regression where the dependent variable is the log of

list price. The FHA VA indicator is not included in this regression as the type of financing is

unknown when the house is listed. The results show that houses that will foreclose, and are in

the lowest house price decile, are not differently priced from those that will not foreclose. Our

previous results showed that at the time of sale, they sold at a premium relative to other houses.

For the seventy percent of house in the middle, the list price for foreclosed houses was 6.3%

below others demonstrating that houses that will end up in foreclosure were viewed as somewhat

undesirable at the listing phase. This 6.3% list price discount is about 1.5% greater than the

discount realized at the time of sale suggesting that the buyers (who are more likely to be less

experienced in real estate transactions) may not recognize the discount as easily as those setting

the list price (the real estate professionals). For the two highest deciles, there is no statistically

significant effect from interacting the price decile with the foreclosure variable relative to the

prices the properties were listed at. For comparison to earlier models, Table 2, Model 5 adds the

FHA VA indicator. Adding the FHA VA indicator to the list price model has the same effect as

adding it to the sale price model, that is, it is significantly negative while other variables retain

their same impact.

Comparing the future foreclosure discount between the list price and sales price one observes

that the discount is about 1½ % smaller at the time of sale than at the time of listing. This result

suggests that on average, individuals buying houses that will eventually go through foreclosure

did not bargain quite as well as the average individual buying other houses because the discount

at purchase is less than the discount at the time of listing. To further study the relationship

between list price, selling price, and the foreclosure discount, Table 3 uses as the dependent

variable the difference between the list price and the sales price. The descriptive statistics of

Table 1 show the average list price is higher than the average sales price. A probe of the data

shows that 21% of the houses sold for the list price, and 15% sold at a premium to the list price[1].

Table 3, Model 1 shows that houses that later went into foreclosure received a $1270 lower

discount than similar homes that did not go into foreclosure, indicating that the initial bargaining

by buyers who went into foreclosure were not as successful as the buyers of similar houses that

did not foreclose. As with the previous analysis, Table 3, Model 2 separates out the lowest and

two highest decile classes to determine whether this effect is uniform across selling price points.

Compared to the 70% of homes in the middle, homes that were in the lowest decile, but did not

foreclose achieved an average additional discount of $430. The additional discount for homes at

the top of the sales range was near $1000, which is smaller, relative to the sales price. For the

homes that became foreclosures, however, all received less of a discount than their non-

foreclosing counterparts. The homes in the lowest decile received about a $700 lower discount,

which is about 1.2% of the sales price. The $940 discount for homes in the middle represents

paying about 0.8% more than those that did not foreclosure. For the top two deciles, the homes

that foreclosed saw about an $11,000 smaller discount compared to those that did not foreclose.

This represents about a 5% discount for the next to highest decile and about a 3% smaller

discount than similar homes for the highest decile. These results indicate that houses that went

into foreclosure, received less of a discount from list price compared to other houses, and this

result holds across house prices.

---

[1] Because the log of non-positive values is not defined, it is not possible to use the log of the price difference to evaluate this discount.

The remaining results of Table 3 illustrate how other characteristics affect the list to sale price discount. Houses priced at the 500, 900, or 000, received lower discounts than other houses suggesting that off-dollar pricing leads to a bigger cut in selling price. Houses that were purchased with a FHA VA loan also received lower discounts. The more atypical houses (Atyp-High, Atyp-MedHigh) realized less of a discount though the results are weak. Homes that are the most typical received more of discount than the average home. New houses received less of a discount and houses with a pool, fireplace, on a large lot, with a tenant, or vacant received a higher discount from list.

Before a house enters the foreclosure process there is a period of mortgage delinquency. In fact, many houses that experience even a severe delinquency, that is a delinquency period in excess of 90 days, often do not result in foreclosure. For conventional loans, Phillips and VanderHoff (2004) found that only about 30 percent of loans that were 90 days delinquent resulted in foreclosure. For FHA loans, Ambrose and Capone (1996) show an overall transition to foreclosure of 38%, and Ambrose and Capone (1998) show an overall transition to foreclosure of 32%. One reason that a severely delinquent loan may not end up in foreclosure is that the home may be sold in advance of the scheduled foreclosure. Not all houses, however, are equally easy to sell in a short period. It is reasonable to hypothesize that houses that are hard to sell quickly are more likely to result in foreclosure, as they cannot be sold quickly to satisfy a delinquent mortgage. Table 4 presents survival models where the dependent variable is the log of the days until the house sells. We employ a Weibull distribution for analysis. The foreclose indicator of Table 4, Model 1 shows that it takes about 12% longer to sell a property that will eventually end up in foreclosure. This result supports the hypothesis that some houses wind up in foreclosure

because they sell slower to and thus are less likely to complete a sale during the delinquency period that precedes the foreclosure. Other factors that add to selling time include large lot, tenant, or vacant. Houses listed at round numbers sell about 2-4% faster than those listed at off dollar prices. This result suggests that a homeowner with a delinquent mortgage who is trying to sell his house before the scheduled foreclosure date should list at a 500 or 000 figure.

Table 4, Model 1, like Haurin (1988), indicates that atypical houses take longer to sell. He suggests the reason for this result is that it takes longer to find a match for an atypical house (longer search period), and this results in a longer duration until sale. This argument equally suggests, however, that houses that are the most typical should also take longer to sell, as extremely ordinary houses are as uncommon as highly atypical houses, and thus would require a commensurate amount of search to find. To test for the non-linearity that such an argument produces, Table 4, Model 2 uses the set of indicator variables noted earlier to note the degree of atypicality. Model 2 shows the non-linear effect of atypicality by demonstrating that the plainest houses are somewhat slower to sell than the average house, while the more atypical houses sell at significantly slower pace than the average house.

Table 4, Model 3 investigates the price decile effects in marketing time. The only price decile shown to be different from the middle is the lowest price houses that take almost 20% longer to sell. Considering the foreclosed subset, Model 3 shows that those in the lowest decile took an additional 13% longer to sell, over and above the 19% delay that other houses in this decile took. For the 70% of houses in the middle, foreclosed house took about 12% longer to sell than the

ones that did not foreclose. For the two highest deciles, we detect no change in selling period for foreclosed versus other houses.

An interesting note is that with the decile information included, the atypicality variable loses its explanatory power. To determine whether there may be some non-linear effects of atypicality, the indicator variables for atypicality employed in earlier models are used in Table 4, Model 4. In this formulation there is some significance only for the plainest houses (least atypical), in that they exhibit a slightly longer marketing time. It appears that atypicality is in some way accounting for the price decile effect as lower valued houses are no doubt smaller than average, and higher value houses larger, which is part of the atypicality measure.

While results presented thus far indicate that homes that will eventually foreclose generally sell at a discount to what a hedonic model predicts, and that such homes also take longer to sell, the most compelling analysis may focus directly on the likelihood of foreclosure. Most default studies focus on the borrower and loan characteristics along with economic and house price trends, and legal remedies in default. Because our data is from a single metropolitan area, all houses are subject to similar area house price trends, economic conditions, and legal environment. We can, however, examine the role that the physical collateral contributes to default, as well as other information that can be derived from the MLS data and hedonic models. In Table 5 we present the results of logistic regressions whose dependent variable is the foreclosure indicator. We use the hedonic pricing model covariates from the previous models, plus we add some others.

Considering the house attribute covariates, Table 5, Model 1, shows that larger houses and newer

houses are less likely to proceed to foreclosure. Houses with more bedrooms are more likely to

end up in default as are houses with a pool. Multistory houses are more prone to default, as are

properties that were vacant when purchased. Listing the house at charm prices does not appear

to effect future default. The FHA VA indicator is strongly positive which confirms the well

known fact that the foreclosure rate on FHA VA loans is much higher than for conventional.

Atypical houses are neither more likely nor less likely to end up in foreclosure. The Days on

Market variable is very significant. Homes that take longer to sell are more likely to end up in

foreclosure.

ONS and LLM found that the residuals of a hedonic model to be helpful in predicting

foreclosure. In Model 1 of Table 5 we also include the residuals from a hedonic sales price

model (LSP Residual). The hedonic model used to compute the residuals is like Table 2, Model

3, except the foreclosure dummies were excluded. The regressions results of Table 5, Model 1

show a statistically important effect from including the hedonic model residuals.

LLM note the non-linear effect of the residuals in that those that "over-pay" have higher default

rates while those that "under-pay" show no effect. To test whether there is a region where the

residuals matter, we created a set of indicator variables that are analogous to those for

Atypicality. That is, we grouped the residuals into the highest 10%, the next 15%, the next 50%,

the next 15% and the remaining 10%. The 50% in the middle is the hold out category for the

results presented in Table 5, Model 2. This formulation confirms the non-linear nature of the

results. In particular, the 25% of the houses with low residuals are about 40% more likely to

default. Houses in the medium high residuals group are about 30% less likely to default. Curiously those with the highest residuals, are not statistically different the 50% holdout of the middle group.

Figure 1 showed there is a strong price trend effect on default; thus, Table5, Model 3, rather than using residuals, uses the log of sales price as a default covariate. This variable is very significant and the overall model fit improves greatly by adding this variable. A natural progression is to simultaneously consider both the residuals, and the sales price. The Predicted Log of Sale Price combines these two variables, and is used as a covariate in Table 5, Model 4. This covariate is very significant, and its magnitude is much greater than for actual sale price. Overall model fit is much better than the previous model as demonstrated by the reduction in AIC and increase in CorrSq[2].

With the addition of the predicted price as a covariate, the inference from several other variables changes, as they may have been proxies for the predicted house value. The interpretation for their coefficients now, is their effect given the predicted house price. The signs on the regression coefficients for size, age and number of bathrooms reversed. Given the predicted house price, larger size, a newer building, and with more bathrooms, the more likely a foreclosure is in its future. Homes with a pool and fireplace retain their positive sign, but the magnitude of the coefficient grows. The effect of a large lot flips sign so that given the predicted selling price, a large lot leads to higher foreclosure. The magnitude of the coefficient for FHA VA declines.

---

[2] AIC is the Akaike Information Criterion. The CorrSq is the square of the correlation between the actual and predicted values of the dependent variable and in this way is analogous to the R-square of linear regression (Maddala 1978).

Atypical homes are less likely to default. Days on market remains an important default covariate.

In addition to measuring the importance of predicted price, one could also measure over or under payment relative to the list price. Table 3, Model 1 showed that houses that eventually foreclosed had received a $1270 smaller discount from list price compared to other houses with the same characteristics. An alternate approach to evaluate a discount effect is to determine whether the size of the discount (or premium) to list price is a useful covariate for predicting Foreclose. Table 5, Model 5 adds the covariate of the percent discount of the Sale Price to the List Price. The estimated coefficient for this variable is negative and highly significant demonstrating that the larger the discount from list, the less likely that a foreclosure will occur in the future. It seems prudent to also investigate the linearity of this discount effect. We followed a similar approach to that for the residuals, with a small modification in grouping. It was noted earlier that 15% of the sales were at a premium to list price, or alternately stated, at a negative discount. The PD Low variable indicates the discount was in the lowest 5%, which means they are the 5% that paid the highest percent markup over list price. PD MedLow represents the remaining 10% of houses on which a premium over list price was paid. PD High represents the 10% of sales that were at the highest discount, and PD MedHigh represents the next 15%. The remaining 60% of the data includes the 21% that were sold at list price, and the remaining 39% that were sold at a small discount from list price. Table 5, Model 6 presents this formulation of the sale to list price discount and it demonstrates a non-linear effect. Those who paid the highest premium over list price are much more likely to default, while those with the biggest discount show a lower likelihood of foreclosure though with a smaller magnitude than those who paid the

highest premium. The medium high and medium low groupings show about the same magnitude of the effect with opposite signs.

4. Conclusions

This paper investigates what MLS data can reveal about foreclosure. We began with a re-examination of studies finding that homes which are over-appraised, relative to a hedonic model, are more likely to end up in foreclosure. The earlier studies of LLM and ONS used data from situations that may be considered unusual (Alaska and Singapore); thus, it seemed prudent to evaluate whether their findings hold in a more typical American housing market. For the Dallas-Fort Worth area, which experienced moderately increasing house prices, the houses that will later end up as foreclosures sell for about 2.7% less than predicted by a hedonic appraisal model. Further analysis showed this discount is not evenly distributed by house price decile. For the lowest price decile, houses that eventually defaulted sold at 2.5% premium to those that did not default. For houses selling in the middle 70%, however, those that foreclosed sold at 4.8% discount. There was no statistical trend for houses in the top 20% of sale prices.

When evaluating the list price, rather than the sale price, the Foreclose indicator for the bottom decile was non significant, and for the 70% in the middle it was somewhat more negative than for sales price. This result suggests that factors beyond those captured in the hedonic model explain the lower price for properties that will foreclose and these factors are realized at the listing phase, and not just at the selling phase. To further uncover the relationship between the list and selling price, we found that the foreclosed houses had about a $1300 smaller discount from list, ceteris paribus at the time of their initial sale. The discount is not symmetrically

distributed around the sales price. For the most expensive 20% of the homes, there was much

less of a discount between list and sales price for those that foreclosed which suggests the buyers

may have over paid for these homes.

Our results indicate that houses that foreclosed were somewhat less desirable houses as indicated

by the fact that in general they both list and sell for less than a hedonic model would suggest. If

these properties were also hard to sell, we would expect a higher default rate from such

properties. Using a survival model that measure the days until a listing sells, we find that homes

that end up as foreclosures took about 12% longer to sell compared to houses that shared similar

characteristics. This sales delay was especially high for the lower priced homes, but was not

evident in the higher priced homes.

A logistic model of foreclosure shows that after controlling for predicted selling price, larger and

newer houses with a pool or fireplace that are situated on a large lot are more likely to end up in

foreclosure. List pricing strategies in terms of "charm pricing" show no effect on future

foreclosures. As expected, homes with FHA VA loans are more likely to end up in foreclosure.

The magnitude of the FHA VA effect, however, declined across models suggesting that one

reason for high default rates on FHA VA loans is the type of house that is financed, rather than

simply the choice of financing vehicle. After controlling for predicted price and sales discount,

the more atypical the house, the less likely it will foreclose. The longer the house was on the

market for its initial sale, the more likely it was to eventually result in foreclosure.

Like previous studies we find a relationship between hedonic model residuals and the likelihood of foreclosure. The effect of the residuals, however, are most usefully incorporated with the sale price – in other words via using the predicted sales price of the house to help determine its likelihood of future foreclosure. In addition, we find, that homes with the smallest percentage discount from list price (that is, the home that paid the highest premium over list), showed a markedly higher foreclosure rate. Homes that sold with a significant discount from list had a lower foreclosure rate. The interaction between home selling price and foreclosures thus revolves around the predicted house price and the discount from list price and not so much hedonic model residuals.

Mortgage default continues as an important vein of research as more advanced theoretical and empirical analysis allows more accurate pricing of mortgage risk. The results of this study suggest further analysis of the role that MLS data can add to understanding mortgage risk. The results here show the type of houses and effective sales price bargaining that foreshadow default, just as other studies show that measures of borrower credit worthiness effect default. Borrowers with low credit capacity will buy low priced houses, and this study confirms that low price houses are more prone to default. While LLM were able to combine some borrower data with house data, expanded research that uses both types of data in a broad analysis would help determine the degree of additional discrimination power that underwriters would gain by incorporating information like that presented here, in addition to the pricing covariates currently in use.

6. References

Allen, M. T and W. H. Dare. 2004. The effects of charm listings prices on house transaction prices. *Real Estate Economics* 32(4):695-713.

Ambrose, B. W. and C. A. Capone. 1996. Do lenders discriminate in processing defaults? *CityScape* 2(1):89-98

Ambrose, B. W. and C. A. Capone. 1998. Modeling the conditional probability of foreclosure in the context of single-family mortgage default resolutions. *Real Estate Economics* 26(3):391-429.

Ambrose, B. W. and Y. Deng. 2001. Optimal put exercise: An empirical examination of conditions for mortgage foreclosure. *Journal of Real Estate Finance and Economics* 23(2):213-234.

Capozza, D. R. , R. Israelsen and T. A. Thomson. 2005. Agency, appraisal and atypicality: Evidence from foreclosures. *Real Estate Economics*. Forthcoming.

Capozza, D. R., D. Kazarian and T. A. Thomson. 1998. The conditional probability of mortgage default. *Real Estate Economics* **26**(3):359-389.

Capozza, D. R., D. Kazarian and T. A. Thomson. 1997. Mortgage default in local markets. *Real Estate Economics* **25**(4):631-655.

Gau, G. W. 1978. A taxonomic model for the risk-rating of residential mortgages. *Journal of Business* 51(4):687-707.

Haurin, D., 1988, The Duration of Marketing Time of Residential Housing, *AREUEA* 16(4): 396-410.

Kau, J. B, D. Keenan, and T. Kim. 1994. Default probabilities for mortgages. *Journal of Urban Economics* 35:278-296. .

LaCour-Little, M. and R. K. Green. 1998. Are minorities or minority neighborhoods more likely to get low appraisals. *Journal of Real Estate Finance and Economics* 16(3):301-315.

LaCour-Little, M. and S. Malpezzi. 2003. Appraisal quality and residential mortgage default: Evidence from Alaska. *Journal of Real Estate Finance and Economics* 27(2):211-233.

Maddala, G. S. 1988. Introduction to econometrics. MacMillan Publishing Company, New York. 472 p.

Ong, S. E., P. H. Neo, and A. Spieler. 2004. Price premium and foreclosure risk. Working Paper

Palmon, O, B. A. Smith and B. J. Sopranzetti. 2004. Clustering in Real Estate prices: Determinants and consequences. *Journal of Real Estate Research* 26(2): 115-136

Phillips, R. A. and J. H. VanderHoff. 2004. The conditional probability of foreclosure: An empirical analysis of conventional mortgage loan defaults. *Real Estate Economics*, 32(4):571-587.

Salter, S. P., K. H. Johnson, and W. P. Spurlin. 2005. Off-dollar pricing, residential property prices and marketing time. Paper presented at the 2005 American Real Estate Society Annual Meeting, Santa Fe, NM.

von Furstenberg, G. M. 1969. Default risk on FHA-insured home mortgages as a function of the terms of financing: a quantitative analysis. Journal of Finance 23:459-477.

Table 1. Descriptive statistics. N = 130,693.

	Mean	Stdev	Min	Max
Sale Price	147877	99618	25000	1000000
List Price	151223	102786	30000	1000000
Day on Market	86.15	64.73	0	723
Foreclosure Indicator	0.0070	0.0832	0	1
Square Feet	20.80	8.03	8	70
Age	1.90	1.66	0	10
Bedrooms	3.37	0.67	1	6
Bathrooms	2.30	0.72	1	6
Pool	0.16	0.37	0	1
Fireplace	0.93	0.51	0	4
Large Lot	0.10	0.30	0	1
Stories	1.29	0.46	1	3
Tenant	0.03	0.18	0	1
Vacant	0.23	0.42	0	1
Atypical	0.25	0.24	-0.24	1.65
FHA VA	0.34	0.47	0	1
List Price – Sale Price	3345	7167	-100000	100000

Table 2. Hedonic house price regression results based on Sales Price and List Price. N = 130,693. All estimated regression coefficients have p-values <0.0001. Dependent variables are as noted.

| | Log of Sales Price | | | | | | Log of List Price | | | |
| | Model 1 | | Model 2 | | Model 3 | | Model 4 | | Model 5 | |
	Est Coef	p-value	Est Coef	p-value	Est Coef	p-value	Est Coef	p-value	Est Coef	p-value
Intercept	10.708	<.0001	11.068	<.0001	10.961	<.0001	11.021	<.0001	11.074	<.0001
SqFt	0.533	<.0001	0.346	<.0001	0.417	<.0001	0.396	<.0001	0.364	<.0001
SqFtSq	-0.023	<.0001	-0.004	<.0001	-0.019	<.0001	-0.009	<.0001	-0.006	<.0001
Age	-0.106	<.0001	-0.084	<.0001	-0.085	<.0001	-0.086	<.0001	-0.081	<.0001
Agesq	0.012	<.0001	0.009	<.0001	0.009	<.0001	0.010	<.0001	0.009	<.0001
New	0.035	<.0001	0.013	<.0001	0.016	<.0001	0.009	<.0001	0.009	<.0001
Bedrooms	-0.032	<.0001	-0.017	<.0001	-0.017	<.0001	-0.022	<.0001	-0.019	<.0001
Bathrooms	0.075	<.0001	0.027	<.0001	0.023	<.0001	0.028	<.0001	0.028	<.0001
Pool	0.080	<.0001	0.068	<.0001	0.063	<.0001	0.069	<.0001	0.067	<.0001
Fireplace	0.062	<.0001	0.041	<.0001	0.039	<.0001	0.042	<.0001	0.042	<.0001
Large Lot	0.065	<.0001	0.056	<.0001	0.052	<.0001	0.065	<.0001	0.062	<.0001
Stories	-0.063	<.0001	-0.044	<.0001	-0.043	<.0001	-0.047	<.0001	-0.044	<.0001
Tenant	-0.071	<.0001	-0.054	<.0001	-0.055	<.0001	-0.049	<.0001	-0.049	<.0001
Vacant	-0.038	<.0001	-0.028	<.0001	-0.028	<.0001	-0.024	<.0001	-0.024	<.0001
FiveHundred	0.008	<.0001	0.001	0.42	0.001	0.40	-0.001	0.66	-0.001	0.39
NineHundred	0.014	<.0001	0.011	<.0001	0.011	<.0001	0.009	<.0001	0.008	<.0001
TripleZero	0.027	<.0001	0.017	<.0001	0.017	<.0001	0.016	<.0001	0.015	<.0001
FHA VA	-0.047	<.0001	-0.050	<.0001	-0.050	<.0001			-0.060	<.0001
Atypical	-0.108	<.0001	-0.210	<.0001			-0.210	<.0001	-0.206	<.0001
Atyp High					-0.067	<.0001				
Atyp MedHigh					-0.032	<.0001				
Atyp MedLow					0.037	<.0001				
Atyp Low					0.059	<.0001				
Foreclose	-0.027	<.0001								
PriceDecile 1			-0.243	<.0001	-0.250	<.0001	-0.229	<.0001	-0.231	<.0001
PriceDecile 9			0.219	<.0001	0.212	<.0001	0.220	<.0001	0.217	<.0001
PriceDecile 10			0.434	<.0001	0.426	<.0001	0.423	<.0001	0.425	<.0001
PriceDecile 1 x Foreclose			0.025	0.006	0.024	0.007	0.005	0.61	0.010	0.26
PriceDecile 2-8 x Foreclose			-0.048	<.0001	-0.049	<.0001	-0.063	<.0001	-0.056	<.0001
PriceDecile 9 x Foreclose			0.011	0.73	0.005	0.88	-0.037	0.24	-0.037	0.22
PriceDecile 10 x Foreclose			-0.002	0.95	-0.004	0.92	-0.031	0.36	-0.036	0.28
Adjusted R-square	0.9043		0.9319		0.9316		0.9315		0.9335	

Table 3. Hedonic regression estimates for dependent variable of List Price minus Sales Price.   N = 130,693.

	Model 1		Model 2	
	Est Coef	p-value	Est Coef	p-value
Intercept	2137.30	<.0001	2479.36	<.0001
SqFt	-740.03	<.0001	-1017.61	<.0001
SqFtSq	785.91	<.0001	799.69	<.0001
Age	215.88	<.0001	255.72	<.0001
Agesq	-3.08	0.61	-10.75	0.076
New	-1215.20	<.0001	-1291.43	<.0001
Bedrooms	-518.94	<.0001	-491.00	<.0001
Bathrooms	413.05	<.0001	358.07	<.0001
Pool	404.33	<.0001	389.12	<.0001
Fireplace	573.97	<.0001	613.99	<.0001
Large Lot	1410.28	<.0001	1404.79	<.0001
Stories	-184.23	0.0003	-182.79	0.000
Tenant	487.24	<.0001	505.97	<.0001
Vacant	610.62	<.0001	613.21	<.0001
FiveHundred	-756.35	<.0001	-757.24	<.0001
NineHundred	-753.41	<.0001	-748.83	<.0001
TripleZero	-366.52	<.0001	-398.46	<.0001
FHA VA	-1013.18	<.0001	-1006.16	<.0001
Atypical				
Atyp-High	274.62	0.002	-147.72	0.112
Atyp-MedHigh	172.17	0.002	-102.83	0.081
Atyp-MedLow	62.76	0.23	169.56	0.001
Atyp-Low	99.00	0.12	227.51	0.0004
Foreclose	-1270.50	<.0001		
PriceDecile 1			429.45	<.0001
PriceDecile 9			864.61	<.0001
PriceDecile 10			1244.46	<.0001
PriceDecile 1 x Foreclose			-693.37	0.09
PriceDecile 2-8 x Foreclose			-940.95	0.0002
PriceDecile 9 x Foreclose			-10753.00	<.0001
PriceDecile 10 x Foreclose			-11027.00	<.0001
Adjusted R-square	0.2245		0.2263	

Table 4. Survival regression models for days on market. N = 130,693. Dependent variable is log of days on market.

	Model 1		Model 2		Model 3		Model 4	
	Est Coef	p-value	Est Coef	p-value	Est Coef	p-value	Est Coef	p-value
Intercept	4.345	<.0001	4.353	<.0001	4.254	<.0001	4.237	<.0001
SqFt	0.273	<.0001	0.264	<.0001	0.302	<.0001	0.313	<.0001
SqFtSq	-0.016	<.0001	-0.015	<.0001	-0.019	<.0001	-0.022	<.0001
Age	-0.033	<.0001	-0.031	<.0001	-0.036	<.0001	-0.035	<.0001
Agesq	0.005	<.0001	0.005	<.0001	0.005	<.0001	0.005	<.0001
New	0.297	<.0001	0.295	<.0001	0.293	<.0001	0.293	<.0001
Bedrooms	-0.030	<.0001	-0.030	<.0001	-0.032	<.0001	-0.033	<.0001
Bathrooms	-0.004	0.47	0.001	0.79	0.016	0.001	0.017	0.0007
Pool	-0.060	<.0001	-0.059	<.0001	-0.053	<.0001	-0.054	<.0001
Fireplace	-0.021	<.0001	-0.019	<.0001	-0.003	0.53	-0.003	0.49
Large Lot	0.251	<.0001	0.252	<.0001	0.256	<.0001	0.256	<.0001
Stories	0.046	<.0001	0.045	<.0001	0.037	<.0001	0.036	<.0001
Tenant	0.192	<.0001	0.192	<.0001	0.188	<.0001	0.188	<.0001
Vacant	0.117	<.0001	0.116	<.0001	0.112	<.0001	0.112	<.0001
FiveHundred	-0.039	<.0001	-0.038	<.0001	-0.037	<.0001	-0.036	<.0001
NineHundred	-0.033	<.0001	-0.033	<.0001	-0.032	<.0001	-0.032	<.0001
TripleZero	-0.023	0.0003	-0.024	0.0003	-0.024	0.0002	-0.024	0.0002
FHA VA	0.020	<.0001	0.021	<.0001	0.022	<.0001	0.022	<.0001
Atypical	0.074	<.0001			-0.019	0.16		
Atyp-High	0.122	<.0001	0.064	<.0001			0.010	0.26
Atyp-MedHigh			0.036	<.0001			0.003	0.65
Atyp-MedLow			-0.002	0.65			0.001	0.80
Atyp-Low			0.018	0.003			0.022	0.0005
Foreclose			0.121	<.0001				
PriceDecile 1					0.191	<.0001	0.187	<.0001
PriceDecile 9					-0.005	0.51	-0.005	0.47
PriceDecile 10					0.001	0.89	-0.003	0.78
PriceDecile 1 x Foreclose					0.131	0.001	0.130	0.001
PriceDecile 2-8 x Foreclose					0.120	<.0001	0.120	<.0001
PriceDecile 9 x Foreclose					0.099	0.48	0.094	0.50
PriceDecile 10 x Foreclose					-0.011	0.94	-0.014	0.92
Scale	0.610		0.609		0.608		0.608	
Weibull Shape	1.641		1.641		1.645		1.645	
Log Likelihood	-130965		-130938		-130643		-130637	

Appraisal and Default   Page 30

Table 5. Logistic models of foreclosure. Dependent variable is foreclosure indicator. N = 130,693.

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	Est Coef	p-value	Est Coef	p-value	Est Coef	p-value	Est Coef	p-value	Est Coef	p-value	Est Coef	p-value
Intercept	-4.484	<.0001	-4.557	<.0001	16.206	<.0001	57.418	<.0001	56.381	<.0001	56.432	<.0001
SqFt	-1.405	0.003	-1.337	0.006	-0.467	0.28	1.921	<.0001	1.976	<.0001	1.994	<.0001
SqFtSq	0.118	0.23	0.104	0.31	0.107	0.22	-0.008	0.93	-0.017	0.85	-0.018	0.84
Age	0.225	0.003	0.195	0.01	0.049	0.52	-0.411	<.0001	-0.391	<.0001	-0.400	<.0001
Agesq	-0.033	0.003	-0.028	0.01	-0.012	0.28	0.049	<.0001	0.047	<.0001	0.048	<.0001
New	-0.460	0.07	-0.474	0.06	-0.427	0.09	-0.401	0.11	-0.485	0.05	-0.507	0.04
Bedrooms	0.296	0.0001	0.292	0.0002	0.220	0.004	-0.015	0.85	-0.026	0.75	-0.024	0.76
Bathrooms	-0.196	0.06	-0.211	0.04	-0.013	0.90	0.327	0.002	0.312	0.004	0.319	0.003
Pool	0.263	0.02	0.269	0.02	0.429	0.0001	0.777	<.0001	0.762	<.0001	0.762	<.0001
Fireplace	0.070	0.42	0.053	0.54	0.221	0.01	0.520	<.0001	0.519	<.0001	0.520	<.0001
Large Lot	-0.041	0.75	-0.046	0.72	0.037	0.77	0.332	0.01	0.364	0.004	0.355	0.01
Stories	0.343	0.001	0.351	0.001	0.248	0.02	0.023	0.83	0.019	0.86	0.017	0.88
Tenant	0.083	0.62	0.092	0.58	-0.013	0.94	-0.257	0.13	-0.241	0.16	-0.255	0.13
Vacant	0.239	0.002	0.240	0.002	0.187	0.02	0.061	0.45	0.064	0.42	0.060	0.45
FiveHundred	-0.114	0.40	-0.115	0.40	-0.107	0.43	-0.074	0.59	-0.067	0.62	-0.059	0.67
NineHundred	-0.025	0.84	-0.028	0.82	0.009	0.94	0.100	0.41	0.102	0.40	0.105	0.39
TripleZero	0.046	0.72	0.049	0.70	0.074	0.57	0.227	0.08	0.225	0.08	0.235	0.07
FHA VA	0.767	<.0001	0.773	<.0001	0.607	<.0001	0.241	0.003	0.174	0.03	0.167	0.04
Atypical	0.154	0.65	0.233	0.51	-0.252	0.43	-0.704	0.031	-0.662	0.05	-0.681	0.04
Days on Market	0.003	<.0001	0.003	<.0001	0.002	<.0001	0.002	<.0001	0.002	<.0001	0.002	<.0001
LSP Residual	-1.514	<.0001										
LSP Resid_High			-0.071	0.57								
LSP Resid_MedHigh			-0.291	0.01								
LSP Resid_MedLow			0.419	<.0001								
LSP Resid_Low			0.382	0.0001								
Log of Sale Price					-1.929	<.0001						
Predicted Log of Sale Price							-5.801	<.0001	-5.698	<.0001	-5.716	<.0001
Percent Discount (PD)									-0.083	<.0001		
PD High											-0.298	0.03
PD MedHigh											-0.432	0.0005
PD MedLow											0.399	<.0001
PD Low											0.805	<.0001
AIC	10271		10256		10178		9940		9858		9852	
CorrSq	0.57%		0.57%		0.65%		1.26%		1.31%		1.39%	

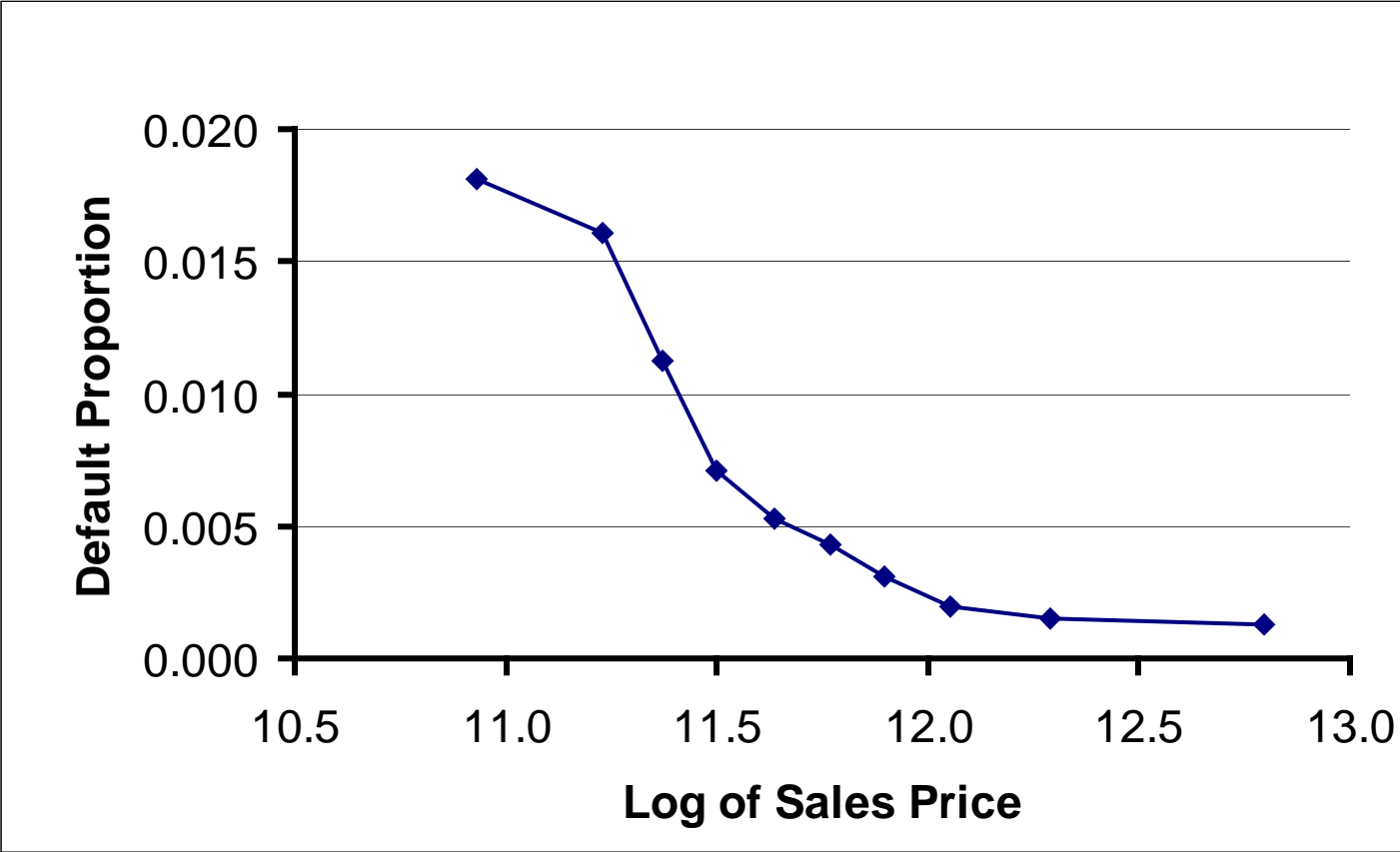Figure 1. Foreclosure proportion versus log of sales price (cross tab by decile).

Figure 2. Quarterly dummies for the hedonic appraisal presented in Table 1, Model 1.