1997 V25 4: pp. 631–655
REAL ESTATE
ECONOMICS

Mortgage Default in Local Markets

Dennis R. Capozza,* Dick Kazarian** and Thomas A. Thomson***

Using recent theoretical advances and an extensive panel data set on metropolitan areas, this study provides new tests of the contingent claims based model of default. The empirical modeling incorporates a full complement of variables that permit direct tests of the options-based model including the conditional effects of age and rent-to-price ratios. The role of transaction costs and trigger events is examined, and the results confirm the importance of both. The effects of aggregation and short sample periods are explored and demonstrated to affect inference in studies of mortgage default.

In recent years the confluence of theoretical advances and economic necessity has stimulated rapid advances in the understanding of the decision to default on mortgage loans. Because of the importance of the issue, a voluminous empirical and theoretical literature has evolved over the last two decades. Nevertheless, numerous unresolved issues remain—partly because the link between the theoretical models and empirical testing has only recently begun to be carefully detailed (Kau, Keenan and Kim (KKK) 1994; and Capozza, Kazarian and Thomson (CKT) 1996).

Theoretical work (*e.g.*, KKK, 1993, 1994) has typically focused on the unconditional probability of default. Unconditional probabilities are best tested using data on cumulative defaults over the entire life of mortgage loans. Empirical studies, on the other hand, most often use annual default data for seasoned loans. These data are most suited for testing the conditional probability of default, *i.e.*, the probability of default over a short horizon (CKT 1996). The interpretation of results can be quite different. For example, the unconditional effect of volatility on default is positive but the conditional effect is ambiguous with negative effects occurring at high loan-to-values (LTVs).

This study exploits the recent theoretical advances on conditional probabilities and the statistical power of an extensive panel data set on mortgage loans that originated in 64 metropolitan areas between 1977 and

* University of Michigan, Ann Arbor, MI 48109-1234 or Capozza@umich.edu.
** Lehman Brothers, New York, NY 10285-1100
*** University of Texas, San Antonio, TX 78249

1990 to refine the options-based empirical model of default and to help resolve some difficult and controversial issues. The article focuses on four areas. The first is the proper specification of an empirical model of default, including both functional form and relevant variables. The second is the role of transaction costs of default (*i.e.*, proxies for claims on other assets such as size of down payment, income and age). A third is the effect of trigger events—those events which convert the multi-period default decision to a one-period decision (including divorce, unemployment and moving rates). The final area concerns the effect of aggregation, sample size and sample period in empirical default studies.

Model Specification

Options models typically have five variables—asset price relative to exercise price, volatility, expiration date, the interest rate and the dividend yield on the asset. Well before the importance of the options to prepay and default for the analysis of mortgages was first emphasized (Findlay and Capozza 1977 and Asay 1978), empirical researchers were aware of the significance of LTV (*i.e.*, one over the asset price relative to the exercise price) and mortgage age (expiry date) in the analysis of default (von Furstenberg 1969). Additional option-based explanatory variables have been investigated more recently. For example, house price volatility was added by Foster and Van Order (1984) and found to be an important covariate. Higher interest rates affect the borrower's decision to default both directly in the option model and indirectly by reducing the effective loan to value ratio (*i.e.*, increasing a homeowner's equity). The empirical importance of interest rates was demonstrated by Vandell and Thibodeau (1985) and Foster and Van Order (1984).

The first refinement here is the inclusion of dividend or asset yield which has been ignored in the empirical literature on mortgage default. Dividend yield effects the probability of default because in equilibrium the expected drift of house prices will be lower the higher the rent yield. For a given volatility this increases the likelihood of hitting the default boundary.

Secondly, the effect of interest rates is explored in greater depth by first adjusting current LTV ratios using the Foster/Van Order (FVO) (1984) procedure and then including the change in interest rates since origination as a separate variable to see if the procedure fully captures the effect of interest rates. The evidence suggests that the FVO procedure over adjusts.

A third refinement tests whether homeowners who do not refinance fail to do so because of distress. When interest rates drop for a cohort of loans, borrowers have an incentive to refinance. Those borrowers who do not may not be able to refinance because the LTV ratio no longer meets lending guidelines. If so, the remaining loans in the cohort will default at higher rates.

Although the theoretical models clearly specify the relevant variables, they are often less clear on the relative importance of the options variables. The empirical modeling here uses the logit model which allows for simple measures of relative impact. Some variables are more important than others by as much as two or three orders of magnitude. Current LTV ratios are particularly important and dominate all other variables in relative impact.

Transaction Costs

The importance of transaction costs in default is controversial. Foster and Van Order (1984) conclude that transaction costs must account for some of the results in their option based model because borrowers do not behave "ruthlessly." Kau, Keenan and Kim (1994) point out that the default option is exercised only when house prices are well below the mortgage balance even if transaction costs are negligible. Lekkas, Quigley and Van Order (1993) and Quigley and Van Order (1995) find differences in loan loss severity and reject the hypothesis that transaction costs do not matter.

Differences in loss severity, however, are necessary, but not sufficient evidence of transaction costs having an impact. The loss severity arising with optimal default varies with all the option model variables. For example, when uncertainty (volatility) is high, the default boundary shifts to higher loan to value ratios and increases loss severity. Simulation of the conditional probabilities of default (CKT 1996) as well as unconditional probabilities (KKK 1994) confirms that transaction costs do have a large negative impact on default especially at high LTV ratios. However, to resolve this issue empirically all of the option model variables must be included in the analysis if misspecification bias is to be avoided.[1] The tests on this issue are

[1] There are other suggestive studies of the role of transaction costs. Clauretie (1987) finds that states that require judicial foreclosure or have statutory rights of redemption have lower foreclosure rates. Jones (1993), with access to data from two Canadian provinces which includes borrower characteristics, argues that "deficiency judgments are the neglected cost component that explains the low incidence of exercise of in-the-money default options reported in the literature." Jones documents that defaults were three or four times higher when deficiency judgments were not permitted. These studies are not definitive. Because they do not include the full complement of options model variables.

complementary to the indirect tests of Quigley and Van Order. Using the MSA (Metropolitan Statistical Area) level panel data and a full complement of independent variables, direct tests of the role of transaction costs are provided.[2]

Trigger Events

In the context of an option pricing model, trigger events affect the optimal default decision by converting what is normally a multi-period optimization into a single period decision. For example, suppose that five years after the origination of a 30-year loan, a borrower experiences a job transfer. With typical parameter values and a 25-year remaining life on the loan, a borrower does not default unless the LTV is about 120% even if the loan is nonrecourse. This occurs because the borrower knows that there may be a more opportune time to default in the future. The transferred borrower, on the other hand, if forced to make the decision immediately can no longer benefit if there is a more opportune time to default in the future. The optimal default boundary falls to an LTV of 100% or less for this borrower (KKK 1993).

In the empirical literature, the results for trigger events are mixed.[3] These mixed results are not surprising in light of the simulation results of CKT (1996). They indicate that:

> optimal default is the first choice available to the borrower. If the borrower does not default or prepay, then, an exogenous event may present itself. When an outsider observes both a default and an exogenous event, an assumption may be made that the "trigger event" caused the default, rather than the default occurring as an optimal decision in the same period

[2] The methodology and data in Quigley and Van Order (QVO) (1995) are very different but arrive at similar conclusions on the transaction cost issue. Because of computational difficulties with the highly nonlinear hazard approach, QVO are limited to considering current loan-to-value and then using indirect tests of transaction costs. In contrast, logit analysis is linear in log odds so that direct tests are possible with a full complement of relevant variables.

[3] Campbell and Dietrich (1983) find strong support for the importance of unemployment, but do not have data on house prices. When house prices are excluded, unemployment may proxy price changes. Foster and Van order (1984), on the other hand, find that adding unemployment or divorce to their options-based model adds littler explanatory power. Clauretie (1987) finds that the change in the unemployment rate is important, but that divorce rates are not. Quigley, Van Order and Deng (1994) find divorce explains higher default rates, but get mixed results regarding unemployment. Thomson (1994) finds both divorce and unemployment significant.

> as a trigger event. The default should not be credited to the trigger event unless the event occurs only because of the trigger event. The overall conclusion here is that trigger events play a minor role. If house prices are low (precipitating high current LTV ratios) defaults will be high, regardless of whether trigger events occur or not.

Trigger events are confirmed to play a minor role once other variables like LTV are appropriately controlled.

Aggregation and Specification Bias

Given the widely divergent and often contradictory empirical results on mortgage default, it is useful to explore possible sources of bias. This is done in three ways. First, alternative specifications are run on the full data sample and the coefficients compared. Second, the data are aggregated to both the regional level and the MSA level to investigate the effect of spatial aggregation. Finally, the sample is split into 1975–79 originations and 1980–83 originations to study temporal aggregation. The results suggest that spatial aggregation is more problematic than temporal disaggregation.

The next section describes the methodology. The third and fourth sections describe the data and the explanatory variables. The fifth presents the results. The final section is the conclusion.

Methodology

For probabilistic events like defaults, outcomes must fall in the [0,1] interval. The logistic transformation is a common method for imposing this restriction. The dependent variable is the observed default outcome for each loan over the past year. Because the loans are grouped into cohorts with a common set of values for the independent variables, a weighting procedure is applied in computing the likelihood function. If a cohort of 100 loans has 2 defaults, then a weight of 2 is applied to the default outcome and a weight of 98 to the nondefault outcome for that set of covariate values. More precisely, the likelihood function to be maximized can be specified as:

$$L = \prod_i P_i^{n_i}(1 - P_i)^{(N_i - n_i)} \tag{1}$$

$i = 1, 2, 3, \ldots, 45{,}986$ (MSA data)
$i = 1, 2, 3, \ldots, 4{,}455$ (regional data)
N_i = the number of loans in cohort i and
n_i = the number of loans in cohort i which defaulted.

For logistic regression:

$$P = \frac{e^{(\beta' x)}}{1 + e^{(\beta' x)}} \tag{2}$$

where:

x = a vector of covariate values and
β = a vector of model parameters.

Weighting by the number of loans in the cohort that share a common set of values for the independent variables leads to the descriptor "weighted logistic regression." Because proportions data are available, one could compute the log odds and then use weighted regression on the log odds. However, as noted earlier, most of the cohorts have a zero default rate for which a log odds is not defined. Rather than using *ad hoc* methods to adjust the data, the maximum likelihood estimation approach was chosen.

Notice that for rare events like default, $e^{\beta' x}$ is close to zero. As a result

$$P \cong e^{\beta' x}$$

and

$$\log P \cong \beta' x$$

so that

$$\frac{\partial \log P}{\partial x_j} = \frac{\frac{\partial P}{P}}{\partial x_j} \cong \beta_j.$$

That is, for rare events like default where the probabilities are small, the coefficients from a logit regression can be interpreted as the percentage effect of a change in one of the covariates on the default probability. These effects decline to zero as the probability approaches one.[4] In the empirical results that follow, an impact percentage is provided for the covariates.

[4] As $e^{\beta x}$ becomes large, $P = \frac{e^{\beta x}}{1 + e^{\beta x}} \rightarrow 1$ and $\frac{\partial \log P}{\partial x_j} \rightarrow 0$.

Impact percentage, I_j, is defined to be the percentage effect of a one standard deviation change in the covariate on default probability at the sample means:

$$I_j = \frac{P(\bar{x} + \beta_j\sigma(x_j)) - P(\bar{x})}{P(\bar{x})}$$

and for small $e^{\beta' x}$ can be approximated by

$$I_j \cong \beta_j\sigma(x_j)$$

where I_j is the impact percent of the j^{th} covariate and $\sigma(x_j)$ is the standard deviation of x_j. Impact percentage, then, is a measure of the relative importance of the explanatory variables at the sample means.

The Data

The default data arises from conventional single family mortgages that originated during the 1975–83 period in 64 MSA's and purchased by the Federal Home Loan Mortgage Corporation (Freddie Mac). The loans are tracked through 1990. Approximately 460,000 loans are represented in this database. For each loan, five pieces of information are available: the MSA of origination, the year of origination, the initial LTV, the year of termination and whether termination occurred through prepayment or default. The approximate contractual interest rate is inferred from the origination year. The loan data are then merged with panel data for the 64 MSAs. The panel includes data on house price indices, house price volatility, employment, population and divorce rates.

The strength of this database is the extensive time series and cross sectional coverage which greatly increases the statistical power of the tests—especially when the weaker effects like trigger events are being considered. The sample is not restricted to loans that terminated during the study period.

The weaknesses of this data set arise from the limited number of loans available in some locations and the lack of information regarding the borrower or property. In addition, Freddie Mac purchased seasoned loans in the early 1980s which could lead to a selection bias for loans that originated in the early part of the sample.

The individual loans are aggregated into cohorts by year of origination, MSA and LTV class. There are 64 MSA's, 9 possible years for origination and 9 LTV classes providing 5,184 possible cohorts into which a given loan may fall.

Cohorts span from 7 years for the 1983 originations, to 15 years for those originated in 1975. Some of the potential cohorts do not have any loans from the start. Other cohorts fall to zero loans prior to the end of the tracking period. The net result aggregates the 3,528,000 default opportunities into 45,986 cohort observations. The average size of a cohort is 77 loans with a range of 1 to 6,537 loans. There are 8,135 defaults with a range of 0 to 122 within a cohort of loans. The average default rate is 0.2% or 1 in 434. The number of cohort observations with a zero default rate is rather high at 42,835 (93% of the observations). Only 18 cohort observations have a 100% default rate.

To obtain the regional data, the 64 MSAs are aggregated into the 5 Freddie Mac regions. The MSA data is weighted by population to compute regional measures. The maximum number of cohorts in the regional aggregation is 405. The actual number of cohort observations is 4,455 over the span of the data.

Since there is considerable debate over the merits of using median or repeat sales data, both the National Association of Realtors median house price and the Freddie Mac repeat sales data are investigated. The repeat sales series is available at the regional level but not the MSA level.[5] The median and the repeat sales price series exhibit similar overall patterns, but there are timing differences, especially in the Northeast. The Pearson product-moment correlations of the price changes are .59, .73, .89, .88 and .92 for the Northeast, Southeast, Northcentral, Southwest and West regions, respectively. The correlations of first differences suggest there should not be a large difference in empirical estimates between the two data sources.

Neither the median nor the repeat sales data are fully quality adjusted. The upward quality drift in the median prices is about 2% per year (Hendershott and Thibodeau 1990) and occurs both because new houses of above average quality are added and because existing houses are renovated. Much of the quality drift arises from renovations. Repeat sales data include only existing houses so that only the drift from renovations applies. Since typical repeat sales procedures attempt to exclude or adjust for houses that increase in size, the quality drift is mitigated. Many existing houses are renovated soon after purchase. For the purpose of measuring house value after loan origination,

[5] This price series is also available for about one-third of the cities used in the MSA level analysis. It was not used in the MSA level analysis because it did not cover enough of the MSA's which were included.

it is the value before quality adjustment that is relevant for loan default. Borrowers base their default decision on the renovated value not the quality adjusted value. Therefore, for studies of loan default the quality drift in median sales prices is not the liability that it might be for other applications.

The Explanatory Variables

The unit of observation of the independent variables varies from national to loan specific. Most of the data, where available or appropriate, is at the metropolitan level. Mortgage age is loan specific; interest rate data is based on national levels. Most other data is metropolitan level except for the divorce rate and the unemployment rate where state level data were substituted for a few MSAs when the local data was unavailable. The variable definitions and sources are given below. The hypothesized signs appear in parentheses.

Frictionless Option Pricing Based Variables

As indicated earlier, there are five variables relevant to the option pricing model—asset price relative to exercise price, volatility, expiration date, the interest rate and the dividend yield. Since exclusion of a relevant variable biases the coefficients of the included variables, proxies are needed for all five. In addition, the option model is highly nonlinear in many of the variables so that consideration must be given to functional form (CKT 1996).

- *The Current Loan-to-Value (CLTV) Index* (+) = (Current mortgage value)/(Current house value index).

The CLTV Index is the proxy for asset price relative to the exercise price. The denominator, current house price, is computed by assuming the house value has changed the same amount as the house price index (median or repeat sales) for the MSA since loan origination.[6] To obtain the numerator, the current mortgage value, the remaining payments are discounted at current

---

[6] As noted, recent empirical studies have been based on option pricing models and stress the importance of CLTV in analyzing default. Alternative measures of CLTV were considered because there is an ongoing debate over the merits of available approaches to measuring house price indices. The median price results is emphasized because it has broader coverage and, as discussed earlier, the lack of quality adjustment is an advantage. Consistent with these considerations, the fit is better than for the other measures. The Haurin, Hendershott and Kim (1991) house price series was also tried in the denominator for the 1982-90 period. The results were similar.

interest rates using the FVO (1984) assumption that the loan will be repaid after 40% of its remaining life.[7]

Because the effect of CLTV may not be linear in the logit regression, a quadratic term and a cross product term with age are included. Since, at most, 100% of the loans in a cohort can default, the effect of CLTV on default probability should decline at high CLTV levels. The logistic function guarantees this diminishing effect.

- *Age* (+/−) = The number of years since the mortgage was originated. *Age* equals 30 minus the time to expiration since the sample includes only 30-year mortgages.

In option models the unconditional effect of age on optimal default increases for the first four or five years and then declines (KKK 1994). Conditional on CLTV, age does not have much direct effect on default. Instead, the impact is indirect through the stochastic process for house prices (CKT 1996). As the mortgage ages, the distribution of prices around the point estimate of CLTV increases. For any given CLTV, the percentage above the default boundary will increase as age increases. Therefore, both linear and quadratic terms in age are included in the regressions.

Two variables refine the interest rate effect:

- *Spread* (+/−) = Current mortgage interest rate minus the coupon rate at which the mortgage was originated.
- *Maxdrop* (+) = Max[0, Mortgage contract rate − Lowest observed interest rate since mortgage origination].

Spread is an attempt to determine whether the adjustment for current interest rates in CLTV fully tracks borrower perceptions of the value of the debt. The sign is indeterminate because the FVO assumption is arbitrary and may cause the CLTV measure to under or over account for the effect of a change in interest rates on the value of the mortgage.[8]

---

[7] The outstanding loan balance without adjustment was also tried but was found to have lower power than the FVO specification and is not reported.

[8] Two other methods for measuring the impact of the interest rate spread were also evaluated. They are—measuring interest rate differences in ratio form, that is current interest rate divided by coupon rate, and an indicator variable that takes the value one when the current interest rate is 200 or more basis points higher than the coupon rate. These alternate measures of the interest rate effect provided lower explanatory power and are not reported.

Maxdrop is the maximum drop in interest rates that has been observed since the mortgage was originated. The purpose is to assess previous opportunities to refinance. If a homeowner has negative equity, refinancing may not be possible. Failure to refinance when it is optimal to do so should signal distress. Thus, this variable provides a further refinement of CLTV. If interest rates have stayed the same or increased over time, this variable takes the value of zero.

- *Sigma* (+/−) = The time series volatility of house prices.

Sigma is computed as the standard deviation of the time series of percentage price changes from 1975 to 1989. It attempts to measure the degree to which individual houses will stray from the CLTV Index. The expected sign for this variable is ambiguous. Unconditionally, the higher the volatility, the higher the expected number of defaults (KKK 1994). Conditional on the current equity position, however, the effect can be negative since the higher the volatility, the higher the value of the option to delay default, and the lower the house price default boundary (CKT 1996).

- *Rent-to-Price Ratio* (+/−) = Home rental cost divided by home price.

Rent-to-Price is a measure of the "dividend yield." Observations at the decadal census are interpolated to compute this ratio for each year by MSA. Unconditionally, the higher the dividend rate, the lower the house price drift. Thus, it is more likely that house prices will fall to the default boundary.[9] Conditional on the current equity position, this sign is ambiguous since there are two offsetting effects (CKT 1996). The higher the rent-to-price ratio, the higher the value of delaying default, and the lower the house price default boundary.

Transaction Costs Variables

Since there are no direct measures of transaction costs, proxies must be used. Three variables measure transaction costs:

- *Personal Income Index* (−) = (Current MSA per capita income)/(Per capita income at loan origination). (Source: Bureau of Economic Analysis.)

[9] If the mortgage payment is less than the rent on an alternative living space, then the borrower will defer default regardless of the equity position in a house since the option can be kept alive costlessly.

- *Pct25-34* (+) = The percentage of the MSA population in the age 25–34 cohort during 1990. (Source: Bureau of the Census.)
- *Pct35-44* (−) = The percentage of the MSA population in the age 35–44 cohort during 1990. (Source: Bureau of the Census.)

As personal income rises, the costs of default are expected to increase since the financial consequences of a negative credit rating and the threat of deficiency judgments will increase. The hypothesized sign, therefore, is negative.

The young age group, *Pct25-34*, is more mobile and has fewer assets, leading to the hypothesized positive sign, since transaction costs are lower for individuals in this age cohort. The middle age group, *Pct35-44*, is likely to have school age children and more financial assets which will increase the transaction costs of default, leading to the hypothesized negative sign.

Trigger Event Variables

Three types of trigger events are considered:

- *Unemployment* (+) = The unemployment rate in the MSA. (Source: Bureau of Economic Analysis.)
- *Divorce* (+) = The divorce rate in the MSA. (Source: National Center for Health Statistics.)
- *Move75* (+) = The proportion of people who changed residence during the 1975-80 period.[10] (Source: Bureau of the Census.)

As unemployment increases, borrowers encounter ability to pay problems leading to higher default rates. Divorce can also lead to ability to pay problems as joint resources are separated to support two households. All these events convert the multi-period optimal default decision into a one-period decision and accelerate the optimal time to default. The transaction costs of default may also fall since a move may occur simultaneously as the

[10] This variable could be considered endogenous since some people move in response to foreclosure. The number of movers, however, greatly exceeds the number of defaulters: Thus, most moves must be exogenous. The 1990 census data for moving rates over the 1985–90 period was also used as an empirical covariate. The two data series are highly correlated; thus, it is reasonable to use only one. The *Move75* variable provides the better fit.

result of a divorce or job transfer. All are hypothesized to have positive signs.[11]

Results

Descriptive Statistics

Table 1 presents descriptive statistics for all the variables at both the metropolitan and regional levels. Mortgage age is loan specific, and interest rate data is national. Thus, these variables are not affected by aggregation. For other variables, however, Table 1 shows that aggregation reduces the dispersion in the independent variables. This is illustrated by the last column, which provides the ratio of the range of the regional data to the range of the MSA data. Aggregation to the regional level often reduces the dispersion of the variables by half or more. The effect is most pronounced for the trigger event and especially the transaction cost variables. Therefore, one expects aggregation to have the greatest adverse effect on these variables.

Initial Specification

In option models of default, homeowner equity (*CLTV*) and loan age are the most important variables. Rational default occurs only when home equity is negative. Age is important because it takes time for the stochastic house price process to reach the default boundary. The initial specification of the base model appears in Table 2 and focuses on these two variables. The full model appears in Table 3 and includes all the options-related variables as well as the transaction costs and trigger event variables.

Linearity. Since the options model is highly nonlinear, examination of functional form is essential. The logistic function is linear in log odds of the default rate ($log(P/(1 - P)) = B'x$). Plotting the log odds versus a covariate indicates whether the logistic will be sufficient to linearize the relationship. The plots for *CLTV* and *AGE* are roughly linear, but a quadratic model fits well and is parsimonious. The plot versus age also suggests that an age 1 dummy may be appropriate. Therefore, the independent variables include Age, Age^2, $CLTV$, $CLTV^2$ and $CLTV*Age$.

[11] Divorce rates were available only through 1987. The divorce rates for the final three years are estimated using linear regression predictions from the first 12 years. State level divorce rates are used where MSA data were not available.

Table 1 ■ Descriptive statistics measured at the metropolitan and regional levels.

	Metropolitan Data				Regional Data				
	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min.	Max.	Ratio of Reg. to MSA Range
Base Model Variables									
Age	5.4	3.4	1.0	15.0	5.4	3.4	1.0	15.0	1.0
Age^2	40.0	44.2	1.0	225.0	40.0	44.2	1.0	225.0	1.0
CLTV Index (median price)	0.5	0.2	0.1	1.2	0.5	0.2	0.1	1.0	0.8
CLTV^2	0.3	0.2	0.0	1.5	0.2	0.2	0.0	0.9	0.6
*CLTV*Age*	2.2	1.2	0.3	8.6	2.1	1.1	0.3	7.0	0.8
Other Options-related Variables									
Spread (%)	1.5	2.7	−6.1	6.6	1.5	2.7	−6.1	6.6	1.0
Maxdrop	0.5	1.1	0.0	6.1	0.5	1.1	0.0	6.1	1.0
Sigma (median price) (%)	7.0	2.0	3.0	9.0	6.0	1.0	4.0	7.0	0.5
Rent-to-Price Ratio (%)	4.5	0.8	2.5	7.8	4.5	0.6	3.7	6.0	0.4
Transaction Cost Variables									
Personal Income Index	1.6	0.4	1.0	3.8	1.6	0.4	1.0	3.4	0.9
Pct25-34	19.5	1.7	13.9	23.5	19.1	1.0	17.4	20.4	0.3
Pct35-44	14.0	1.0	11.0	19.5	13.8	0.3	13.0	14.1	0.1
Trigger Event Variables									
Unemployment (%)	5.4	2.0	1.5	14.9	5.4	1.5	3.0	10.1	0.5
Divorce (%)	5.3	1.2	2.5	9.5	5.3	0.8	3.2	7.1	0.6
Move75	52.8	6.4	32.5	66.2	51.8	5.7	40.2	55.6	0.5
Other									
CLTV Index (repeat sales data)	NA				0.5	0.2	0.1	0.9	
Sigma (repeat sales data) (%)	NA				6.0	1.0	3.0	7.0	

This table provides descriptive statistics for the variables used in the empirical analysis. Statistics are provided both for aggregation to the MSA level and to the regional level. The total number of observations = 3,527,814. The data are weighted by the number of loans in each cohort. *Age* is the number of years since origination. *CLTV* is the current loan to value ratio using current estimated house prices and the FVO adjustment for interest rates. *Interest Rate Spread* is the difference between current mortgage rates and the rate at origination. *Maxdrop* is the maximum drop in interest rates since origination. *Sigma* is the standard deviation of the percentage price changes from 1975-89. *Rent-to-Price Ratio* is the ratio of average rents to average house prices from Census data. *Personal Income Index* is an index of income with the year of origination set equal to one. *Pct25-34* is the percentage of the population in the 25-34 year old cohort in 1990. *Pct35-44* is the percentage in the 34-44 year old cohort in 1990. *Move75* is the proportion of the population that changed residence from 1975–80.

Fit. Goodness of fit in Tables 2–5 is assessed in three ways. First, for each of the independent variables, the p-value is indicated. Second, the Akaike Information Criterion (AIC), a measure of the predictive usefulness of this model, is included.[12] The lower the AIC, the better the model should predict.

[12] The AIC provided in the SAS output and reproduced in the tables is 2*log of the likelihood +2k, where k is the number of covariates, including the intercept term.

Table 2 ■ The base model of default.

	MSA Level Data			Regional with Median Prices			Regional with Repeat Sales Prices		
Explanatory Variables	Coefficient Estimate	Impact Percentage	p	Coefficient Estimate	Impact Percentage	p	Coefficient Estimate	Impact Percentage	p
Intercept	−19.39		**	−20.81		**	−21.18		**
Dummy Age 1	−0.94		**	−0.92		**	−0.88		**
Age	0.71		**	0.93		**	1.09		**
Age^2	−0.02		**	−0.03		**	−0.03		**
CLTV Index	25.21	763	**	26.99	872	**	27.57	745	**
CLTV^2	−11.02		**	−11.11		**	−11.33		**
*CLTV*Age*	−0.09		0.02	−0.24		**	−0.46		**
AIC	93,443			94,070			95,429		
Corr^2	0.31			0.58			0.49		
Base Default Rate (%)	0.09			0.09			0.09		

Dependent Variable = Annual default rate. Weighted logistic regression estimates. The *impact percentage* is the change in the dependent variable (in percentage terms) if the given variable is increased by one standard deviation, or for a binary variable, if it takes the value 1, when other variables are at their means, and binary variables are at 0.

* Indicates $0.0095 > p > 0.0001$.
** Indicates $p < 0.0001$.

Finally, because this is a nonlinear regression model, R^2 cannot be used as a measure of goodness of fit. Maddala (1988) suggests alternatives for logistic regression. The simplest is the correlation squared (*Corr*^{2}) between the predicted and actual results (which equals the R^2 in a linear regression). The *Corr*^{2} is computed by cohort, weighted by the number of loans in each cohort. The AIC for the base model in Table 2 is 93443 versus an AIC of 115049 for simply using the mean default rate as the predicted default rate. The *Corr*^{2} measure is 0.31.

Economic Impact. As indicated earlier, assessment of the impact of an individual covariate in a nonlinear regression is not as straight forward as for a linear regression. To implement impact percentage, the percentage change in the predicted default rate is calculated for a one standard deviation increase in a covariate, while continuous variables are set to their mean values and binary variables are set to 0. For binary variables, the impact percentage shows the effect of the covariate taking the value 1.

When all variables are at their means, and binary variables are at zero, the base level default prediction is 0.89%, or 1 in 1120, which is less than the 1 in 434 that represents the overall database. In other words, the average mortgage does not default at the average rate, but rather at a lower rate.

Table 3 ■ The full model.

	MSA Level Data			Regional Level Data		
Explanatory Variables	Coefficient Estimate	Impact Percentage	*p*	Coefficient Estimate	Impact Percentage	*p*
Intercept	−22.98		**	−36.11		**
Dummy LTV0-49	1.46		**	2.51		**
Dummy Age 1	−0.83		**	−0.79		**
Base Model Variables						
Age	1.24		**	1.78		**
Age^2	−0.02		**	−0.03		*
CLTV Index	36.50	1066	**	50.79	1834	**
CLTV^2	−16.58		**	−24.24		**
*CLTV*Age*	−0.71		**	−1.24		**
Other Options-related Variables						
Interest Rate Spread	0.31	128	**	0.39	181	**
Maxdrop	0.52	82	**	0.60	99	**
Sigma	11.00	19	**	−18.32	−17	0.05
Rent-to-Price Ratio	−0.3	−22	**	−0.97	−41	*
Transaction Cost Variables						
Personal Income Index	−1.25	−41	**	−0.80	−28	*
Pct25-34	0.01	1	0.44	0.22	24	**
Pct35-44	−0.1	−9	**	0.57	19	0.01
Trigger Event Variables						
Unemployment	0.06	12	**	0.12	20	**
Divorce	0.06	7	**	0.38	36	**
Move75	0.01	5	0.02	−0.08	−36	**
AIC	91,426			92,268		
Corr^2	0.36			0.70		
Base Default Rate (%)	0.07			0.06		

This table reports the estimates of the full model with the complete set of option pricing variables as well as the transaction cost and trigger event variables. The dependent variable in all cases is the annual default rate. The models are estimated using weighted logistic regression. Both equations use the median price data to calculate the *CLTV Index*. *Age* is loan age since origination. *CLTV Index* is the index of current loan-to-value ratio. *Pct25-34* is the percentage of the population in the 25-34 year old age cohort. *Pct35-44* is the percent in the 35-44 age cohort. *Move75* is the proportion of the population who changed residence from 1975–1980. The *impact percentage* is the percentage change in the dependent variable if the given variable is increased by one standard deviation, or for a binary variable, if it takes the value 1, when other variables are at their means, and binary variables are at 0.

* Indicates $0.0095 > p > 0.0001$.
** Indicates $p < 0.0001$.

If the CLTV increases by one standard deviation above its average value, the projected default rate, using the values in Table 2, increases to 0.77%, (1 in 130) which is an increase of 763% over the base level; hence, the impact percent shown for CLTV is 763%. CLTV, therefore, strongly impacts default. Because of the nonlinear functional form, impact percentage is not constant but declines as CLTV increases.

Figure 1 illustrates the effect of CLTV on conditional probability of default using the estimates in Table 2. The concavity of the log P curve (Panel B) reflects the declining percentage impact of CLTV. Notice that conditional default probabilities are only 9%, even at high average MSA CLTVs. This occurs because not all high CLTV loans default in a given year. Also, since CLTV is an MSA average, roughly half the loans have CLTV below the average.

The direct effect of age on conditional default probability is small (CKT 1996). As indicated earlier, most of the effect is indirect through the effect of time since origination on the distribution of CLTVs around the point estimate. Aging increases the size of the default tail for a given CLTV. If house prices follow a log-normal diffusion, the standard deviation of the distribution is a concave function of time since origination. The estimates in Table 2 are consistent with this conditional default interpretation.[13]

Regional Data. Table 2 also provides two regional regressions: one with median house prices and another with the repeat sales data. Predictably, with the aggregation to regions, the overall explanatory power increases from a $Corr^2$ of .31 to .58 (median prices). This arises because aggregation reduces the variation in the dependent variable. Qualitatively, the base model is little affected by aggregation since all the signs remain the same. If age proxies for unobserved heterogeneity in CLTV in the regional regressions, then the magnitude of the age effect may change. This in fact occurs and is most pronounced for the *CLTV*Age* cross product term.

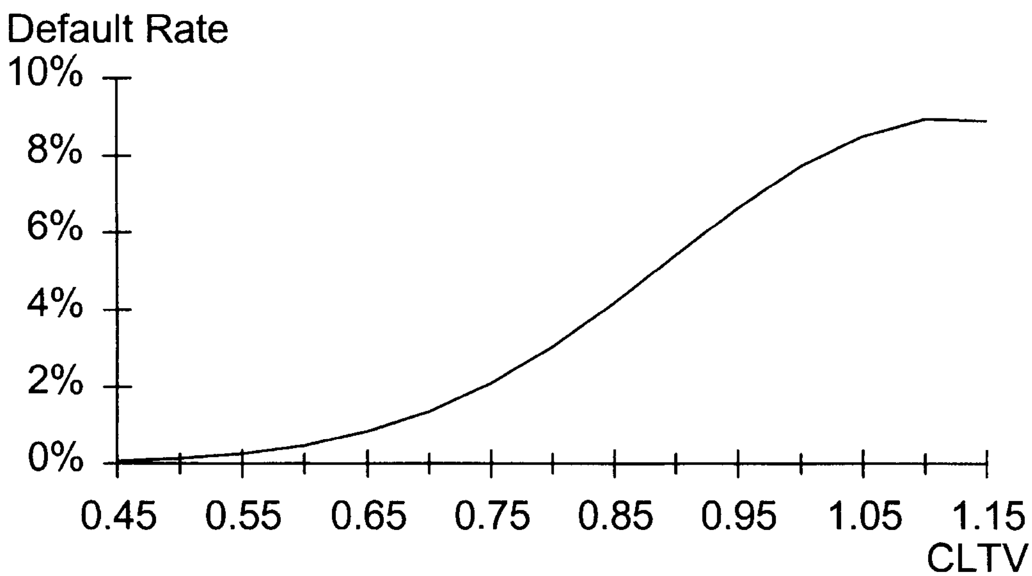
The two regional regressions in Table 2 provide additional insights on the effect of quality drift. The conjecture, as discussed earlier, is that for default modeling, repeat sales data underestimate price changes while median prices overestimate price changes. This arises because homeowners often renovate or improve their properties after purchase and loan origination. Full quality adjustment, as is attempted with repeat sales data, does not capture the resulting increase in equity. A review of the coefficients of the two regional regressions in Table 2 indicates that there is remarkable similarity between the two regressions. However, the impact percentage is higher for the median price regression than for the repeat sales regression. This is consistent with the conjecture since the impact of CLTV should be underestimated (overestimated) if the price index understates (overstates) price changes. The impact percentages differ primarily in the coefficient on the interaction term,

[13] The unconditional effect of age is hump shaped rather than monotonic so that the results are not consistent with the unconditional interpretation.

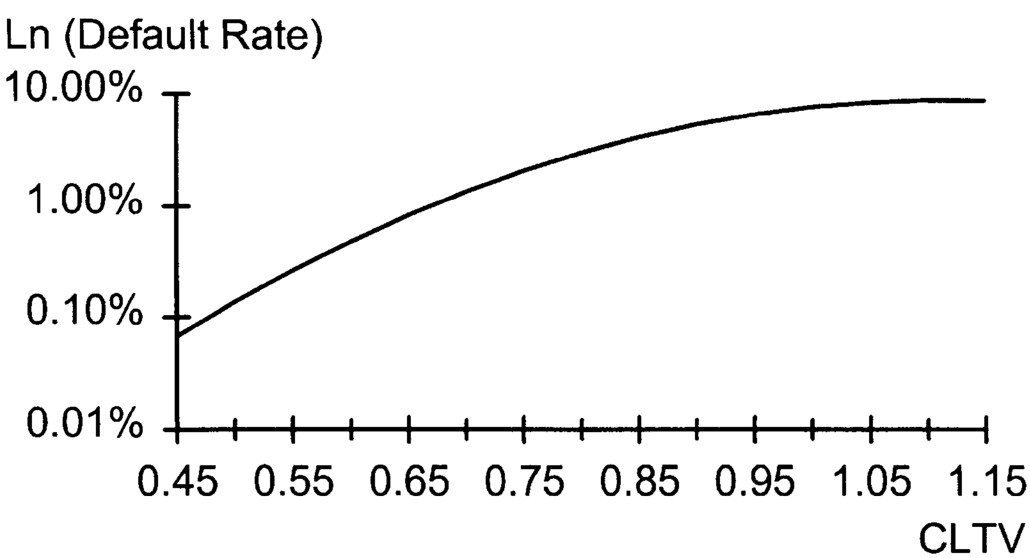
Figure 1 ■ The effect of CLTV on the default probability.

This figure shows the effect of CLTV on default. The graph is based on the MSA model in Table 2. The graph shows the nonlinear relationship between CLTV and default and the diminishing effect of CLTV at high CLTVs. In Panel A the vertical axis is the actual default rate and the graph illustrates the logistic function. In Panel B the vertical axis is the log of the default rate which shows the diminishing percentage impact of CLTV on the default rate.

Panel A



Panel B



*CLTV*Age*, which is exactly what is expected if the two series are similar except for the extent of quality drift. Which set of coefficients is less biased cannot be determined from these data; however, it is reassuring that the effect of quality drift appears to be small and largely mitigated by the inclusion of the interaction term.

Comparing the overall goodness of fit in the two regional regressions reveals that fit is better when using the median prices. The AIC is lower and the $Corr^2$ is higher. This suggests that the median house price data provide a slightly better empirical foundation for computing the CLTV Index than do the repeat sales house price data. Only the median house price data results are reported in the remaining tables.

The Full Model

Table 3 displays the results for the full model. It includes the complete option specification plus transaction cost and trigger event variables. Relative to the base model the fit improves significantly with $Corr^2$ rising to .36 for the MSA data and .70 for the regional data.

The coefficients on *CLTV* and *Age* are difficult to interpret because of the nonlinearity of the logit model. Plotting the implied function for the full model (not reported) similar to Figure 1 reveals that the curves for the two models differ primarily at high CLTVs where the full model curve plots as much as 2% or more below the base model when all other variables are at their means. Stated differently, *excluding some of the relevant variables increases the loading on CLTV and overstates the effect of this variable on defaults.*

The additional option model variables in Table 3 are significant but of lower economic impact than *CLTV*. *Spread*, the difference between the current interest rate and the coupon rate on the mortgage, is positive suggesting the FVO adjustment for mortgage value overcompensates. Recall that the FVO adjustment assumes that an increase in mortgage rates will be seen by the borrower as the same as a decrease in the balance on their loan since the present value of their payments is less than the contractual balance. If this adjustment to CLTV is accurate, *Spread* should not be significant. The positive and significant coefficient indicates that homeowners assess that the effect of a rise in interest rates on the value of the mortgage balance is less than that computed in the CLTV Index.[14]

---

[14] If the Foster-Van Order adjustment to CLTV had not been made, the sign on the spread variable would be negative. The variable would then capture the effect of increasing interest rates on the present value of the mortgage payments directly.

As hypothesized, the sign on *Maxdrop* is positive and has an important impact on defaults. Borrowers who do not refinance when the opportunity presents itself often refrain from doing so because of financial distress. These borrowers are then more likely to default in the future. *Sigma,* the house price volatility measure, has a positive effect in the MSA data. This is consistent with other empirical studies where most data is drawn from low LTV cohorts. Theoretically, the conditional effect of volatility (CKT 1996) is ambiguous but the negative effect occurs only at LTVs above 100%. This variable reverses sign and loses significance in the regional data.

The rent-to-price ratio is a novel addition to empirical default studies. The effect is negative and statistically significant but the economic impact is small. This is consistent with the conditional probability simulations in CKT (1996).

Transaction Cost. The three transaction costs variables in Table 3 are correctly signed. All but *Pct25-34* are statistically significant in the MSA sample. In the regional sample, all are significant but *Pct35-44* reverses sign. Unlike the indirect evidence in Quigley and Van Order (1995), these results provide *direct evidence that transaction costs are influencing the default decisions of borrowers.* Two of the three transaction cost variables are highly significant, but the economic impact is definitely secondary to the frictionless option model variables.

Trigger Events. The trigger event variables in Table 3 include the unemployment rate, the divorce rate and the moving rate. They are hypothesized to increase defaults since they reduce the transaction costs of default and convert the multi-period optimal default decision to a one period decision. In the MSA equation, all the trigger event variables are significant and correctly signed. In the regional equation *Move75* reverses sign. These results provide direct evidence that trigger events influence the default decisions of borrowers. Again, however, the economic impact is secondary to the standard option model variables.

Split Sample

Since many of the results differ markedly from earlier studies using smaller samples or more aggregated data, it is illuminating to explore the extent to which the power of the larger and more disaggregated sample is influencing the results. Table 4 splits the MSA sample by loan origination period, 1975–79 originations in the first regression and 1980–83 originations in the second regression. The originations from the earlier period have a much lower default rate. Nevertheless, there is remarkable similarity in the signs

Table 4 ■ Split sample regressions using MSA data.

	1975–79 Originations			1980–83 Originations		
Explanatory Variables	Coefficient Estimate	Impact Percentage	*p*	Coefficient Estimate	Impact Percentage	*p*
Intercept	−24.54		**	−20.06		**
Dummy Age 1	−0.91		**	−0.12		.22
Dummy LTV0-49	2.41		**	−0.18		.66
Base Model Variables						
Age	1.39		**	1.37		**
Age²	−0.02		**	−0.05		**
CLTV Index	42.75	1132	**	24.41	381	**
CLTV²	−21.23		**	−9.49		**
*CLTV*Age*	−0.80		**	−0.55		**
Other Options-related Variables						
Interest Rate Spread	0.40	136	**	0.44	167	**
Maxdrop	0.55	23	**	0.66	223	**
Sigma	12.44	13	**	8.96	17	**
Rent-to-Price Ratio	−0.03	−3	.30	−0.45	−32	**
Transaction Cost Variables						
Personal Income Index	−1.25	−42	**	−0.28	−6	.34
Pct25-34	0.13	24	**	−0.03	−6	.03
Pct35-44	−0.26	−22	**	−0.09	−9	*
Trigger Event Variables						
Unemployment	0.04	8	*	0.04	8	*
Divorce	0.03	3	.23	0.12	15	**
Move75	−0.04	−20	**	0.04	33	**
AIC	40855			49454		
Corr²	.50			.35		
Base Default Rate (%)	0.03			0.25		

This table reports the estimates of the full model over samples split by origination year and aggregated to the MSA level. The dependent variable in all cases is the annual default rate. The models are estimated using weighted logistic regression. Both equations use the median price data to calculate the *CLTV Index*. *Age* is loan age since origination. *CLTV Index* is the index of current loan-to-value ratio. *Pct25-34* is the percentage of the population in the 25-34 year old age cohort. *Pct35-44* is the percentage in the 35-44 age cohort. *Move75* is the proportion of the population who changed residence from 1975–1980. The *Impact Percentage* is the percentage change in the dependent variable if the given variable is increased by one standard deviation, or for a binary variable, if it takes the value 1, when other variables are at their means, and binary variables are at 0.

* Indicates $.0095 > p > .0001$.
** Indicates $p < .0001$.

of the variables between the two periods so that the inference is similar. Sign reversals occur only for *Pct25-34* and *Move75*. The magnitude of the coefficients places more economic importance on the additional option related variables in the 1980–83 period and less on the transaction cost variables. The overall conclusion is that *sample period does not have a major impact on the empirical results when using MSA level data.*

Table 5 ■ Split sample regressions with regional level data.

	1975–79 Originations			1980–83 Originations		
Explanatory Variables	Coefficient Estimate	Impact Percentage	*p*	Coefficient Estimate	Impact Percentage	*p*
Intercept	−63.60		**	−6.80		.12
Dummy Age 1	−0.31		.07	−0.06		.52
Dummy LTV0-49	3.23		**	0.12		.77
Base Model Variables						
Age	1.48		**	1.60		**
Age^2	−0.01		.01	−0.07		**
CLTV Index	47.34	1639	**	25.58	479	**
CLTV^2	−24.94		**	−9.03		**
*CLTV*Age*	−0.54		*	−0.69		**
Other Options-related Variables						
Interest Rate Spread	0.42	148	**	0.45	170	**
Maxdrop	0.18	7	*	0.72	255	**
Sigma	−146.40	−77	**	89.70	174	**
Rent-to-Price Ratio	−1.82	−62	**	−0.16	−9	.14
Transaction Cost Variables						
Personal Income Index	−2.53	−65	**	0.50	10	.43
Pct25-34	0.46	53	**	−0.15	−15	.01
Pct35-44	3.62	197	**	−1.48	−38	**
Trigger Event Variables						
Unemployment	0.14	23	**	−0.14	−23	**
Divorce	−0.08	−6	.51	0.58	66	**
Move75	−0.12	−47	**	−0.02	−13	.17
AIC	41984			49225		
Corr^2	.73			.74		
Base Default Rate II (%)	0.03			0.23		

This table reports the estimates of the full model over samples split by origination year and aggregated to the regional level. The dependent variable in all cases is the annual default rate. The models are estimated using weighted logistic regression. Both equations use the median price data to calculate the *CLTV Index*. *Age* is loan age since origination. *CLTV Index* is the index of current loan to value ratio. *Pct25-34* is the percentage of the population in the 25–34 year old age cohort. *Pct35-44* is the percentage in the 35–44 age cohort. *Move75* is the proportion of the population who changed residence from 1975–1980. The *Impact Percentage* is the percentage change in the dependent variable if the given variable is increased by one standard deviation, or for a binary variable, if it takes the value 1, when other variables are at their means, and binary variables are at 0.

* Indicates $.0095 > p > .0001$.
** Indicates $p < .0001$.

Table 5 provides the same split of origination periods but for the regional sample. With the regional data there is a large difference between the two sets of originations. There are six sign reversals between the two periods. In other cases where there is no sign reversal, there is a large change in the coefficient, for example with *Maxdrop* and *Rent-to-Price*. The implication is that the loss of statistical power from *aggregation to the regional level does have a significant impact on the empirical results* and helps to explain the divergent results in the literature.

Conclusion

This study exploits recent theoretical advances on the probability of default and the statistical power of an extensive panel data set on metropolitan areas to investigate some difficult issues in mortgage default. The empirical model includes a full complement of relevant option model variables and avoids the specification bias that arises when the relevant variables are only partially specified. Strong evidence on a number of issues is provided. First, similar to other empirical studies of default, LTV is an important determinant of default. However, the importance of LTV can be overstated if other relevant variables are excluded from the empirical specification.

Second, the options based model of default includes five variables. Measures of all five are included in the empirical specification and all are significant. The results are consistent with the effect of mortgage age arising indirectly through the stochastic process for house prices. Age increases the distribution around the point estimate of CLTV and increases the conditional probability of default. Most notable, however, is analysis of the rent-to-price ratio or dividend yield which is usually excluded from empirical models. A negative relationship with defaults is found. There are two possible interpretations for this finding. An ability-to-pay interpretation is that when rents are high the alternative to owning is less attractive. The options interpretation, on the other hand, would be that when the dividend yield is high the drift of the house price is smaller or may even be negative. With less positive drift the upside potential is smaller, and conditional on a given LTV, it is worthwhile to default earlier.

Transaction costs have become the most contentious area. Transaction costs are included directly in the empirical model along with the full complement of options based variables. Transaction costs matter even after controlling for all the other options variables. These results complement those of Quigley and Van Order (1995) who find indirect evidence of the effects of transaction costs from loss severity data.

The search for evidence of an economic effect of trigger events, like unemployment and divorce, on default has yielded mixed results in the past. Trigger events affect default by converting the multi-period decision into a one-period decision. Stated differently, trigger events shorten the expiration date of the option and increase defaults conditionally. The evidence here strongly supports this hypothesis. Both unemployment and divorce rates have a statistically significant effect on default.

Transaction cost and trigger event variables have much less impact on default, as measured by the impact percentage, than the options related variables. Since these effects are weak, they are also the most difficult to verify empirically and are sensitive to specification. The strong evidence of these effects appears only in the MSA level data.

Finally, because the literature includes many diverse findings, evidence is provided on the effect of sample size and aggregation. Splitting the sample into two periods by origination does not greatly affect the results if the analysis is at the MSA level. On the other hand, when the data are aggregated to the regional level the results are highly variable, both in sign and magnitude between the two periods. This is evidence that aggregation is more destructive of statistical power than sample size.

Because of the importance of loan defaults to the lending industry, considerable research effort has and will continue to be devoted to understanding the causes and consequences of mortgage default. No one empirical study can provide the final evidence on this issue. The results here, however, do point to some of the paths that need to be pursued in future research. For example, the full nature of the interaction between LTV and other determinants of default needs to be explored in greater depth. The limitations of both the hazard model and logit approaches leave much work to be done on functional form. In addition, the role of borrower and property characteristics in the context of the options model is still little examined.

The authors thank Jan Brueckner, Pat Hendershott, Jim Shilling, Robert Van Order the reviewers and the participants at a seminar at Ohio State University for helpful comments.

References

Asay, M. 1978. A Theory of the Rational Pricing of Mortgage Contracts. Ph.D. Dissertation. University of Southern California: Los Angeles, CA.

Campbell, T.S. and J.K. Dietrich. 1983. The Determinants of Default on Insured Conventional Residential Mortgage Loans. *Journal of Finance* 38(5): 1569–1581.

Capozza, D.R., D. Kazarian and T.A. Thomson. 1996. The Conditional Probability of Default. Working paper. University of Michigan Business School: Ann Arbor, MI.

Clauretie, T.M. 1987. The Impact of Interstate Foreclosure Cost Difference and the Value of Mortgages on Default Rates. *Journal of the American Real Estate and Urban Economics Association* 15(3): 152–167.

Findlay, M.C. and D.R. Capozza. 1977. The Variable Rate Mortgage and Risk in the Mortgage Market. *Journal of Money, Credit, and Banking* 9(2): 356–364.

Foster, C. and R. Van Order. 1984. An option-based model of mortgage default. *Housing Finance Review* 3(4): 351–372.

Haurin, D.R., P.H. Hendershott and D. Kim. 1991. Local House Price Indexes: 1982–1991. *Journal of the American Real Estate and Urban Economics Association* 19(3): 451–472.

Hendershott, P.H. and T. Thibodeau. 1990. The Relationship Between Median and Constant Quality House Prices: Implications for Setting FHA Loan Limits. *Journal of the American Real Estate and Urban Economics Association* 18(3): 323–334.

Jones, L.D. 1993. Deficiency Judgments and the Exercise of the Default Option in Home Mortgage Loans. *Journal of Law and Economics* 36(1): 115–138.

Kau, J.B., D.C. Keenan and T. Kim. 1993. Transaction Costs, Suboptimal Termination and Default Probabilities. *Journal of the American Real Estate and Urban Economics Association* 19(3): 247–263.

———. 1994. Default Probabilities for Mortgages. *Journal of Urban Economics* 35(3): 278–296.

Lekkas, V., J. Quigley and R. Van Order. 1993. Loan Loss Severity and Optimal Mortgage Default. *Journal of the American Real Estate and Urban Economics Association* 21(4): 353–371.

Maddala, G.S. 1988. Introduction to Econometrics. MacMillan Publishing Company: New York, NY.

Quigley, J.M., R. Van Order and Y. Deng. 1994. The Competing Risks for Mortgage Termination by Default and Prepayment. A Minimum Distance Estimator. Presented at the NBER Summer Institute: Cambridge, MA.

Quigley, J.M. and R. Van Order. 1995. Explicit Tests of Contingent Claims Models of Mortgage Default. *The Journal of Real Estate Finance and Economics* 11(2): 99–117.

SAS Institute Inc. 1989. SAS/STAT User's Guide, Version 6, Fourth edition, Volume 2. SAS Institute Inc., Cary, NC.

Thomson, T.A. 1994. A Metropolitan Analysis of Mortgage Loan Defaults. (Abstract.) *Journal of Finance* 49(3): 1097–1098.

Vandell, K.D. and T. Thibodeau. 1985. Estimation of Mortgage Defaults Using Disaggregate Loan History Data. *Journal of the American Real Estate and Urban Economics Association* 13(3): 292–316.

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* 24(3): 459–477.