

# **Spatial Heterogeneity in Mortgage Terminations by Refinance, Sale and Default**

by

Yongheng Deng  
University of Southern California  
650 Childs Way, RGL 328  
Los Angeles, CA 90089-0626  
[ydeng@usc.edu](mailto:ydeng@usc.edu)

Andrey D. Pavlov  
Simon Fraser University and CIRANO  
8888 University Dr.  
Burnaby, BC V5A 1S6, Canada  
[apavlov@sfu.ca](mailto:apavlov@sfu.ca)

Lihong Yang  
University of Southern California  
650 Childs Way, RGL 331  
Los Angeles, CA 90089-0626  
[lihongya@usc.edu](mailto:lihongya@usc.edu)

October 2003

*Special thanks to Nancy Wallace and the participants of the 2003 Annual AREUEA meeting for the numerous insightful comments and suggestions. Steve Cauley and the Ziman Real Estate Center at UCLA provided the real estate transaction data and invaluable suggestions on the data aspects of the project. Wayne Creadick, Jr. of Wells Fargo Mortgage provided the mortgage history data. Yongheng Deng and Lihong Yang acknowledge financial support from the Lusk Center for Real Estate at USC. Andrey Pavlov acknowledges financial support from the Social Sciences and Humanities Research Council of Canada.*

# **Spatial Heterogeneity in Mortgage Terminations by Refinance, Sale and Default**

## **Abstract**

This paper investigates the impact of spatially correlated unobservable variables on the refinancing, selling, and default decisions of mortgage borrowers. Virtually the entire mortgage literature acknowledges that borrower specific characteristics, such as culture, education, or access to information, play an important role in the mortgage termination decisions. While we do not observe these variables directly, we note that borrowers of similar background tend to cluster together in neighborhoods. We propose a method to take advantage of this information and reconcile the theoretical option-based models of mortgage terminations with the empirical experience of mortgage refinancing, sale and default.

Specifically, we combine the three-stage maximum likelihood estimation (3SMLE) approach for competing risks hazard model with random effect proposed by Deng and Quigley (2002) with the space-varying coefficient method (SVC) of Pavlov (2000) to modify the covariance structure according to the spatial distribution of the observations. Beyond a significant improvement of the model performance, this yields a number of insightful implications for mortgage termination behavior. For instance, borrowers of the affluent “West Side” of Los Angeles County both refinance and move at a higher rate than predicted by the standard maximum likelihood estimation method. At the same time, borrowers from some lower-valued neighborhoods tend to stay longer than expected with their mortgages and properties. Such findings have direct implications for mortgage pricing and have the potential to ultimately improve the equity and efficiency of the lending markets.

Key words: duration model, competing risks, space-varying coefficient, martingale residual, mortgage refinance, default, residential mobility

JEL code: G21, C31, C41, D12

## 1. Introduction

Mortgage-backed securities market has recently become the largest capital market for investors in the U.S. Not surprisingly, a large volume of literature studies mortgage borrowers' prepayment and default behavior and its impact on the pricing of mortgage-backed securities. However, due to errors in variables or limited availability of borrower characteristics, most empirical studies find a substantial discrepancy between the theoretically derived optimal behavior and the observed decisions. (See, for example, Deng, Quigley and Van Order [1996] and Stanton [1996]). This paper attempts to reconcile the theoretical option-based models of mortgage terminations with the empirical experience of mortgage terminations by refinancing, sale and default.

From a theoretical perspective, we explicitly model the borrower's costs associated with mortgage terminations and recognize that those costs vary across individuals and termination causes. Consistent with this approach, we empirically separate the three major causes of mortgage termination: refinancing, selling of the property, and default. Furthermore, since borrowers of similar characteristics (education, income, culture and ethnic background, etc.) tend to cluster together in neighborhoods, many of the omitted variables and measurement errors are spatially correlated. Recognizing this spatial correlation we empirically model the variability of the mortgage termination costs through the use of the physical location of the properties. This approach gives rise to a competing risks hazard framework with spatially correlated errors.

Consistent with the above implication, we estimate the refinancing, selling, and default probabilities using an innovative three-stage maximum likelihood estimation (3SMLE) approach for competing risks hazard model with random effect proposed by Deng and Quigley (2002). In the first stage, we estimate a competing risks model of refinance, sale and default in a conventional maximum likelihood estimation approach and collect the residuals of the estimation for each individual loan. In the second stage, we estimate the neighborhood spatial heterogeneous functions using the residuals from the first stage estimation following the space-varying coefficients method (SVC) of Pavlov (2000). In the third stage, we re-estimate the competing risks hazard model of refinance, sale and default by accounting for the consistent estimation of neighborhood

spatial heterogeneous error distributions. The 3SMLE approach allows us to account for unobserved neighborhood spatial heterogeneity using a geo-coded micro loan data and hence provides more efficient estimates.

Beyond providing a significantly better fit to the data, the proposed methodology allows us to make a number of insightful observations about the mortgage termination behavior of borrowers from different neighborhoods. For instance, we find that borrowers from the affluent West Side of Los Angeles County tend to both refinance and move at a higher rate than predicted by standard maximum likelihood estimation. On the contrary, borrowers from some lower income neighborhoods tend to stay with their mortgages and homes longer than predicted by a standard model. Since refinancing and mobility behavior influences the market value of mortgages, our findings have direct implications for mortgage pricing and have the potential to improve the efficiency and fairness of the lending industry.

In Section 2, which follows, we provide a brief review of the related mortgage termination studies. Section 3 provides a theoretical model that explicitly incorporates the individual unobservable transaction costs and their impact on termination behavior. Section 4 develops the empirical implementation. Sections 5 and 6 describe the data and provide the empirical results and Section 7 concludes.

## **2. Literature Background**

Kau and Keenan (1995) and Capone (2001) provide comprehensive reviews of the mortgage termination literature. One of the most important messages of this literature is that due to errors in variables or limited availability of borrower characteristics, most empirical studies find a substantial discrepancy between the theoretically derived optimal behavior and the observed decisions. (See, for example, Deng, Quigley and Van Order, 1996). To reconcile this discrepancy, we address the unobservable variables problem through modeling their spatial distribution. Pavlov (2001) first suggested this idea by dividing the Los Angeles Metropolitan Area into 22 exogenously determined neighborhoods and allowing the model parameters to vary across those neighborhoods.

Our first point of departure is that we incorporate the spatial distribution of the observations through the 3SMLE approach of Deng and Quigley (2002). They show that

such an approach provides for substantial efficiency gains and is better able to handle complex covariance structures.

Furthermore, we model the spatial distribution of the observations non-parametrically using the SVC method. As Pavlov (2000) shows, this approach substantially improves the estimation both in terms of out-of-sample errors and in terms of neighborhood definition. Since our goal is to model the unobservable borrower characteristics, non-parametric definition of the neighborhoods is particularly important for the economic conclusions of the model.

### 3. Mortgage Termination and Transaction Costs

Following Pavlov (2001), we employ the following three-state variable framework:

- $r$  = the instantaneous riskless rate of interest,
- $H$  = the value of the mortgaged property, and
- $G$  = the monetary equivalent gain from moving.

The dynamics of the riskless rate and home values are assumed to follow generalized stochastic processes that are potentially correlated. Following Clapp et al. (2001) and Pavlov (2001) we model the optimal and actual bundle of housing services as stochastic variables affecting the mortgage termination choices. The urban economics literature suggests that both optimal housing consumption bundle,  $h_d$ , and actual bundle of housing services provided,  $h_s$ , change over time. Turnbull (1995) suggests that housing consumption changes if income, prices, or transportation costs change. Thus, we assume that the optimal housing consumption bundle,  $h_d$ , follows a stochastic process that is derived from a process describing individual circumstances.<sup>1</sup>

We denote the gap between the optimal and the actual bundles of housing services by  $g(t) = h_d(t) - h_s(t)$ . This gap captures the desire to move and can be positive or negative. The larger, in absolute value, the difference between the optimal and the actual housing services, the more likely the benefit of moving will exceed the costs. The

---

<sup>1</sup> Note that the demand for housing services is affected by individual preferences and cannot be viewed as an exogenous process. Instead, it is derived from an exogenous process describing the individual circumstances.

monetary equivalent gain from moving,  $G(t)$ , follows a stochastic process derived from  $g(t)$  and the utility function of the agents.

We consider fully amortizing mortgages having an original principal of  $F(0)$ , and a fixed coupon rate of  $c$  for an original term to maturity of  $T$ .  $F(t)$  denotes outstanding balance at time  $t$ . Let  $M(r, H, G, t)$  denote the value to the borrower of the risky mortgage at time  $t$ . This value includes the options to prepay and default. Notice that the value of this mortgage also depends on the monetary equivalent of the difference between the optimal and the actual housing consumption bundles,  $G(t)$ . If this difference is large, the expected tenure in the property is short. If borrowers view the expected tenure in the property as the effective terms to maturity of the mortgage (i.e., the mortgage contract is terminated upon moving),  $G(t)$  affects those terms to maturity. Below we show that this has important implications for refinancing since borrowers who expect to move shortly are less likely to refinance.

In a similar setting, Pavlov (2001) derives the following boundary conditions for refinancing and moving:

$$M(r, H, G, t) - F(t) > TC_R, \quad (1)$$

$$G(t) + M(r, H, G, t) - F(t) > TC_M, \quad (2)$$

where  $TC_R$  and  $TC_M$  denote the transaction costs associated with refinancing and moving, respectively. We interpret these costs as not just the fees and/or points related to refinancing or the brokerage costs related to moving, but also the monetary equivalent of the disutility of moving or refinancing. This includes information costs, cost of time and effort, and all other drawbacks of these actions.

The refinancing boundary (1) states that a borrower should prepay as soon as the value of the risky mortgage minus the outstanding balance exceed the refinancing costs. Notice that the value of the mortgage is assumed to depend on the monetary equivalent of the gain from moving. The larger this gain, the shorter the expected holding period of the mortgage, and the smaller the benefits from refinancing.

The move boundary (2) states that a borrower will move if the monetary equivalent of the gain from moving plus the difference between the market value of the

mortgage and the outstanding balance exceed the transaction costs of moving. This allows for an interaction between the value of the mortgage and the probability of moving. In particular, if the value of the mortgage is below the outstanding balance,  $F(t)$ , the borrower may be “locked-in” into an advantageous mortgage.<sup>2</sup> In this case a borrower may delay an otherwise utility increasing move. On the other hand, if the value of the mortgage  $M(r,H,G,t)$  exceeds the outstanding balance, a borrower will be more likely to move. Furthermore, if a borrower is considering moving soon, then the market value of the mortgage will be very close to the outstanding balance and refinancing is unlikely.

Conditional on moving, borrowers have to decide whether to sell the property and prepay the mortgage or default. Absent default costs, borrowers will sell and prepay if they have positive equity in the property (i.e.,  $H(t) > F(t)$ ), and default if they have negative equity. Including default costs,  $TC_d$ , conditional on moving, the borrowers will:

- Sell and prepay, if  $H(t) > F(t) - TC_d$ , or
- Default, otherwise.

This condition allows for some borrowers to optimally avoid default even if they have negative equity in their homes.

The above treatment of default is distinct from the traditional approach (Kau et al., 1995, Schwartz and Torous, 1989, among others). We assume that borrowers will default only if they have made the decision to move described above. In other words, we assume that a decline in the value of the property alone is not enough to justify default. This is consistent with the empirical observation that a very large number of homeowners with negative equity do not default. Cauley (1996) reports that up to 44% of homes purchased between 1989 and 1991 in Los Angeles County had negative equity in 1995, yet the increase of defaults was negligible.

This assumption makes the option to default quite different from a typical put option on an asset as it is not exercised the way a financial put option is. In terms of the model, this assumption eliminates the boundary condition that borrowers will default if the value of the property falls below the outstanding balance. This, however, does not preclude the equity in the property and the costs of default to enter the decision to move.

---

<sup>2</sup> See Quigley (1987) for a discussion on the “lock-in” effect in residential mobility.

Figure 1 summarizes the above mortgage termination model.

[Figure 1 is about here]

Since the desire to move,  $G$ , and the transaction costs, as defined above, are unobservable and individual specific, we can analyze the probability of violation of boundary conditions (1) or (2). Therefore, we model moving and refinancing as a hazard function that depends on the value of the mortgage and on the outstanding mortgage balance. Furthermore, conditional on moving, we analyze the probabilities of default versus selling of the property as a logit choice function which depends on the available equity. Specifically, we define the following two hazard functions:

$$h_R = h_R(r, H, G, t), \quad (3)$$

$$h_M = h_M(r, H, G, t), \quad (4)$$

where  $h_R$  is the conditional probability at time  $t$  that the refinancing condition (1) is reached, and  $h_M$  is the conditional probability at time  $t$  that the move condition (2) is reached.

#### 4. Empirical Implementation

The hazard functions corresponding to the two competing risks of mortgage termination, refinancing and moving, can be described as:

$$h_j(t_j | z_{j,t}(r, H, G)) = \lim_{\Delta t \rightarrow 0} \frac{P(t_j \leq T_j < t_j + \Delta t | T_j > t_j)}{\Delta t}, \quad j = R, M, \quad (5)$$

where  $j$  indicates one of the two conditions and  $z_j(r, H, G)$  is a covariate vector which is a function of the three-state variables:  $r$ , the instantaneous riskless rate of interest,  $H$ , the value of the mortgaged property, and  $G$ , the monetary equivalent gain from moving.  $z_j(r, H, G)$  may contain time-varying as well as time-invariant covariates.



Reaching each of the conditions corresponding to refinancing or relocation results in termination of the mortgage. Thus, assuming there are no terminations due simultaneously to more than one cause, (i.e., a borrower who simultaneously refinances and moves), the overall instantaneous termination probability is:

$$h(t_R, t_M | z_{j,t}(r, H, G)) = \sum_j h_j(t_j; z_{j,t}(r, H, G)), \quad j = R, M. \quad (6)$$

The joint survivor function conditional on  $r$ ,  $H$ , and  $G$  can be expressed in the following form:

$$\begin{aligned} S(t_R, t_M | r, H, G, \xi_R, \xi_M, \theta) \\ = \exp \left\{ -\xi_R \sum_{k=1}^{t_R} \exp(\gamma_{R,k} + \beta'_R z_{R,k}(r, H, G)) \right. \\ \left. - \xi_M \sum_{k=1}^{t_M} \exp(\gamma_{M,k} + \beta'_M z_{M,k}(r, H, G)) \right\}, \end{aligned} \quad (7)$$

where  $\xi_R$  and  $\xi_M$  are spatially correlated unobserved error terms associated with the hazard functions for refinance and move respectively.  $\theta$  is a vector of parameters (e.g.,  $\gamma$  and  $\beta$ ) of the hazard function.  $\gamma_{j,k}$  are parameters of the baseline hazard function.

Following Deng, Quigley and Van Order (2000), the probability of termination by refinance, move or censoring<sup>3</sup> can be expressed as following:

$$\begin{aligned} F_R(t_i | \xi_{R,i}, \xi_{M,i}) &= S(t_i, t_i | \xi_{R,i}, \xi_{M,i}) - S(t_i + 1, t_i | \xi_{R,i}, \xi_{M,i}) \\ &\quad - \frac{1}{2} \left\{ S(t_i, t_i | \xi_{R,i}, \xi_{M,i}) + S(t_i + 1, t_i + 1 | \xi_{R,i}, \xi_{M,i}) \right. \\ &\quad \left. - S(t_i, t_i + 1 | \xi_{R,i}, \xi_{M,i}) - S(t_i + 1, t_i | \xi_{R,i}, \xi_{M,i}) \right\}, \end{aligned} \quad (8)$$

---

<sup>3</sup> Censoring refers to the situation where no event of either refinancing or moving has occurred by the end of data collection period.

$$\begin{aligned}
F_M(t_i | \xi_{R,i}, \xi_{M,i}) &= S(t_i, t_i | \xi_{R,i}, \xi_{M,i}) - S(t_i, t_i + 1 | \xi_{R,i}, \xi_{M,i}) \\
&\quad - \frac{1}{2} \left\{ S(t_i, t_i | \xi_{R,i}, \xi_{M,i}) + S(t_i + 1, t_i + 1 | \xi_{R,i}, \xi_{M,i}) \right. \\
&\quad \left. - S(t_i, t_i + 1 | \xi_{R,i}, \xi_{M,i}) - S(t_i + 1, t_i | \xi_{R,i}, \xi_{M,i}) \right\},
\end{aligned} \tag{9}$$

and

$$F_C(t_i | \xi_{R,i}, \xi_{M,i}) = S(t_i, t_i | \xi_{R,i}, \xi_{M,i}), \tag{10}$$

and the log likelihood function of the competing risks model is given by

$$\log L = \sum_{i=1}^N \delta_{R,i} \log(F_R(T_i)) + \delta_{M,i} \log(F_M(T_i)) + \delta_{C,i} \log(F_C(T_i)), \tag{11}$$

where  $N$  is the sample size and  $\delta_{j,i}$ ,  $j = R, M, C$ , are indicator variables that take the value of one if the  $i^{\text{th}}$  loan is terminated by refinance, move, or censoring, respectively, and zero otherwise.

The log likelihood function specified by equations (7) through (11) depends on the pair of random effects,  $(\xi_{R,i}, \xi_{M,i})$ , with unknown distribution. Deng and Quigley (2002) has noted that consistent evaluation of equations (7) through (11) can be obtained by substituting a consistent estimate  $(\hat{\xi}_{R,i}, \hat{\xi}_{M,i})$  for the unobserved  $(\xi_{R,i}, \xi_{M,i})$ .

The development of counting process theory and martingale residuals in survival models provides a useful instrument for the unobserved  $(\xi_{pi}, \xi_{di})$ . Following Barlow and Prentice (1988), and Therneau et al. (1990), the martingale residual for the  $i$ th individual is defined as

$$\begin{aligned}
\hat{M}_i(t) &= N_i(t) - \hat{E}_i(t) \\
&= N_i(t) - \int_0^t Y_i(s) \exp(X_i(s) \hat{\beta}) d\hat{h}_0(s),
\end{aligned} \tag{12}$$

where  $N_i(t)$  takes a value 1 at time  $t$  if individual  $i$  has experienced the event of interest and 0 otherwise;  $Y_i(s)$  is a censor indicator that takes value 1 if individual  $i$  has survived up to time  $s$ , and 0 otherwise;  $h_0(s)$  is the baseline hazard function. The martingale residuals can be interpreted as the excess number of events observed in the data but not predicted by the model.

Following Deng and Quigley (2002), we estimate the log likelihood function using following a three-stage approach that accounts for unknown random effects

$$(\xi_{R,i}, \xi_{M,i}):$$

First, we estimate the competing risks hazard model of mortgage refinance and move specified by equations (7) through (11), ignoring the unobserved error components,  $(\xi_{R,i}, \xi_{M,i})$ , and obtain the martingale residuals,  $(M_{R,i}, M_{M,i})$ , for the refinance and move functions, respectively.

Second, the martingale residuals  $(M_{R,i}, M_{M,i})$  obtained from the first stage estimation are analyzed using the space-varying coefficients method (SVC) following Pavlov (2000). This method combines non-parametric estimation of the error structure with respect to the longitude and latitude coordinates of each property and parametric components. In our application, we use a second degree polynomial to model the effect of the time since origination. Further, we separate the non-termination (censored) and termination errors. The method allows for any number of independent variables to enter the parametric function.

For each observation we estimate a modified error based on weighted linear regression which *does not include* the observation under consideration itself. The weights are computed from the physical distance between the observations. Excluding the observation of interest from the estimation guarantees that the procedure does not over fit the data and retains the forecasting properties of the estimation.<sup>4</sup>

Third, the log likelihood function specified by equations (7) through (11) is re-estimated by correcting the random-effect errors, i.e., simply replace  $(\xi_{R,i}, \xi_{M,i})$  in

---

<sup>4</sup> For a more detailed description of the SVC method see Pavlov (2000). Note that a number of other non-parametric or parametric methods can be used to model the covariance structure.

equation (7) by the exponential of the predicted martingale residuals  $(\exp(\hat{M}_{R,i}), \exp(\hat{M}_{M,i}))$ , such that

$$\begin{aligned}
& S(t_{R,i}, t_{M,i} | \gamma_i, z(r_i, H_i, G_i), \xi_{R,i}, \xi_{M,i}, \theta) \\
&= \exp \left\{ -\exp(\hat{M}_{R,i}) \sum_{k=1}^{t_{R,i}} \exp(\gamma_{R,k} + \beta'_R z_{R,k}(r_i, H_i, G_i)) \right. \\
&\quad \left. - \exp(\hat{M}_{M,i}) \sum_{k=1}^{t_{M,i}} \exp(\gamma_{M,k} + \beta'_M z_{M,k}(r_i, H_i, G_i)) \right\}.
\end{aligned} \tag{13}$$

## 5. Data

The mortgage loan data comes from *Wells Fargo Mortgage* and contains private-label securitized mortgage loan originations, spanning a period from 1988 to 2001. The data is in high quality and is consistent through time. The second data set used in this analysis contains all residential real estate transactions in Los Angeles County since 1988. The underlying data comes from the County Recorders of Los Angeles and is provided by *DataQuick*.

To separate the moving from the refinancing decision, we merge the loan and transaction data sets. There are four common fields in the two data sets: transaction date, transaction value, mortgage loan size, and postal zip code. While the values of those variables match exactly for most loans, we allow for small variations as follows:  $-1/+3$  months on the transaction date,  $\pm 1\%$  of property value and origination loan balance. Such small discrepancies are justified because of potentially different accounting for points, fees, and taxes, as well as mortgage closing versus transaction date records.

Once a loan is matched, we can determine whether the property in question was sold within a short period of the time of mortgage termination. If a mortgage was prepaid without a contemporaneous transaction, then we assume it has been refinanced. If there is a contemporaneous transaction, then the mortgage was prepaid due to the move. After the match we are left with 3,665 usable observations, 2,569 of which refinanced, 857 are still active, 166 moved, and 73 defaulted.

## 6. Empirical Results

Below we describe the explanatory variables used in the empirical estimation. Most of those variables are standard in the mortgage termination literature.

### 6.1 Explanatory Variables: Refinancing

The framework outlined in Section 3, and especially boundary condition (1), suggests that, conditional on not relocating, the explanatory variables related to refinancing should measure two components:

- the benefits of refinancing, (i.e., the difference between the value of the mortgage and the outstanding balance), and
- the related refinancing costs, broadly defined as not only the direct fees but also the information costs and the costs of time and effort necessary to refinance.

#### 6.1.1 Benefit of Refinancing

*Interest Rate Ratio.* Following Richard and Roll (1989) we consider the ratio of the two rates:

$$Rate\ Ratio(t) = \frac{Prevailing\ Rate(0)}{Prevailing\ Rate(t)}, \quad (14)$$

where *Prevailing Rate* indicates the prevailing Fixed Rate Mortgage (FRM) rates on the market, 0 indexes origination time, and  $t$  indexes for the time span since the mortgage loan was originated. We choose the prevailing mortgage rates since they already incorporate the market valuation of the prepayment options discussed above. Notice that both components are prevailing market rates, not coupon rates.

Holding everything else constant, an increase of the interest rate ratio would increase the value of the mortgage relative to the outstanding balance. According to boundary condition (1), this increase would make refinancing more likely. Therefore, we expect the rate ratio to have positive impact on refinancing.

Distinct from many studies on mortgage termination, we consider the change in prevailing mortgage rates in Equation (14) rather than the difference between the

individual origination mortgage coupon rate and the current prevailing rates. The individual rate typically contains a risk premium or discount specific to that borrower or property. Since we do not observe this premium or discount through time or at termination, we cannot include it at origination either. In other words, it is unrealistic to assume that all borrowers would be able to obtain the prevailing mortgage rate. Instead, our specification of the interest rate ratio (14) captures changes in the macroeconomic environment. Below we consider two additional variables that capture the changes of interest rates available to the specific borrower, i.e., capture the changes in the risk premium or discount for that borrower.

*Slope of the Yield Curve.* The yield curve is included in the estimation to capture the expectations about future interest rates. It is computed as the difference between the yields of the 30-year Treasury bond and the 3-month Treasury bill.

*Change in the LTV ratio.* Our specification of the interest rate ratio, Equation (14), focuses on the macroeconomic influences on the mortgage rate available to the specific borrower. To capture changes in the risk premium or discount for that borrower we consider the change in the LTV ratio through time. Using a hedonic transaction – based approach developed by Pavlov (2000) we estimate the value of each property through time. To accommodate the non-linear impact of the LTV ratio on mortgage rates, we consider only increases of the LTV ratio above 80%:

$$LTV(0,t) = \begin{cases} LTV(t) - LTV(0), & \text{if } LTV(t) > 80\% \quad \text{or} \quad \text{if } LTV(0) > 80\% \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

Holding the prevailing rate constant, an increase in the LTV ratio would result in higher borrowing rate for the homeowner. This has an effect of reducing the value of the mortgage,  $M$ , which, in turn, makes refinancing less likely.

*High origination LTV ratio.* To further capture the risk premium or discount for the specific borrower, we include a variable measuring the extent to which the *original* LTV ratio was above 80%:

$$HighLTV = \begin{cases} LTV(0) - 80\%, & \text{if } LTV(0) > 80\% \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

The inclusion of this variable is driven by data and empirical considerations, and, therefore, our theory provides no prediction on the direction of its impact.

### 6.1.2 Transaction Costs of Refinancing

As discussed above, we interpret transaction costs to include both the monetary expenses and the non-monetary costs related to the process. This definition requires the inclusion of several proxies for the refinancing costs. According to boundary condition (1), high refinancing costs have a negative impact on the probability of refinancing.

*Refinancing History.* Foregoing numerous refinancing opportunities indicates high transaction costs or other constraints. To capture this effect we include the “woodhead” measured as the number of months the refinancing rates have been lower than the coupon rate and the borrower does not refinance the loan.<sup>5</sup> A large number of missed refinance opportunity is perhaps an indication of high transaction costs.

*Outstanding Balance.* Since the benefits of refinancing are proportional to the outstanding balance, while many of the costs are fixed, higher outstanding balance would make refinancing more likely.

*Documentation.* Borrowers able to obtain financing with limited documentation may face lower transaction costs and be more willing to refinance.

## 6.2 Explanatory Variables: Moving

### 6.2.1 Value of the mortgage

The terminal condition for moving, Equation (2), suggests that all variables related to the value of the mortgage affect the probability of moving. Therefore, we include the interest rate ratio, changes in the LTV ratio, high original LTV ratio, outstanding balance, and slope of the yield curve in the move estimation. Nonetheless,

---

<sup>5</sup> The “woodhead” effect in mortgage refinancing refers to the observation that some financially less astute mortgage borrowers repeatedly ignore profitable refinance opportunities. See Deng and Quigley (2002) for more discussion.

we expect this effect to be small relative to the difference between the actual and desired bundle of housing services.

As discussed above, the interest rate ratio has a positive impact on the difference between the market value of the mortgage and the outstanding balance,  $M - F$ , which, in turn, makes moving more likely. Furthermore, an increase in the LTV ratio reduces the value of the mortgage relative to the outstanding balance, which has a negative impact on moving. High original LTV and the slope of the yield curve are empirically motivated variables and our theory does not predict their impact on moving.

The impact of the outstanding balance on moving, however, is unclear. Both the monetary equivalent of the gain of moving,  $G$ , and the moving costs,  $TC_M$ , are independent of the outstanding balance. Therefore, higher outstanding balance would make the difference between the market value of the mortgage and the outstanding balance relatively more important. Since it is not clear whether this difference is positive or negative, however, we are unable to make a prediction on the impact of the outstanding balance on moving.

### **6.2.2 Desire to move**

The desire to move is perhaps one of the most difficult variables to proxy. We use the tenure in the home as a proxy for the desire to move. Furthermore, if a borrower anticipates moving, i.e., the monetary equivalent of the gain from moving,  $G$ , is large, then he or she would forego otherwise advantageous refinancing opportunities. Therefore, “woodhead” in refinancing, as described above, is a measure of the desire to move and we expect it to have a positive impact on the probability to move.

### **6.2.3 Selling or Default: The Two Vehicles of Moving**

The framework outlined in Section 3 suggests that, conditional on the decision to move, a borrower will choose between selling the property and defaulting on the loan. The two variables entering this decision are the LTV ratio and the costs of default. We approximate the costs of default by the level of documentation, origination date, outstanding balance, and “woodhead” measure.



### 6.3 Parameter Estimates

Tables 1 and 2 list all variables described above and provide descriptive statistics of the sample. We report the values of all time-varying variables at origination and at termination. The observations we use were originated between 1988 and 1996, although, as mentioned above, we tracked the event history of each individual loan up to 2002. The origination dates are approximately equally spread between 1988 and 1996.

[Table 1 is about here]

[Table 2 is about here]

Of the time-varying covariates, perhaps the most important one is the interest rate ratio. By construction the interest rate ratio is 1 at origination but largely varies at termination. This effect is largest for the refinanced mortgages. The other variables are consistent with previous literature.

Table 3 provides the parameter estimates from three models we examine. Model 1 is estimated by a standard maximum likelihood approach. All parameter estimates in both the refinancing and the moving estimation are as expected and consistent with the previous literature.

[Table 3 is about here]

Model 2 in Table 3 includes exogenous neighborhood definitions in the estimation. Following Pavlov (2001), we consider the 22 regions of Los Angeles County. These regions are designed to be of roughly equal size and yet overlap with the important neighborhoods of the metropolitan area. Many of the dummy variables associated with these regions are significant and the log likelihood is significantly higher relative to Model 1.<sup>6</sup>

Finally, Model 3 in Table 3 presents the result of the 3SMLE with non-parametric modeling of the spatial distribution of the random effects. While we retain the parametric

---

<sup>6</sup> The difference between the two log-likelihood values is well above the 99<sup>th</sup> percentile critical value for a likelihood ratio test.

neighborhood variables, notice that all except four of them lose their significance.<sup>7</sup> None of the other covariates change sign or significance level relative to Model 2 with the exception of the outstanding balance in the move estimation, which loses significance. The log likelihood increases substantially with a likelihood ratio test value of 1,666 (compared to the 99% critical value of 9.21). Recall that the space-varying coefficients estimation for each modified error was performed *excluding* the observation itself. Thus, the substantially better fit is *not* driven by over fitting the data but reflects the spatial distribution of the unobserved spatially correlated random effects.

Beyond a substantially better fit, non-parametric modeling of the random effects yields further insights into the mortgage termination behavior of borrowers. Figure 2 depicts the refinancing residuals for the terminations due to refinancing, and Figure 3 depicts the refinancing residuals for the terminations due to moving or censoring.<sup>8</sup> Similarly, Figure 4 depicts the moving residuals for termination cause of moving, and Figure 5 depicts the moving residuals for termination causes of refinancing or censoring.<sup>9</sup>

[Figures 2-5 are about here]

First, notice the substantial spatial correlation of those residuals. Furthermore, we are in a position to identify endogenously determined neighborhoods where the residuals are unusually high or low. For instance, borrowers from the affluent West Side of Los Angeles both refinance and move at a higher rate than predicted by Model 1. This is evident both from the termination and censored residuals. At the same time, borrowers from the lower price neighborhoods both tend to refinance and move at a lower rate.

Those figures provide a very clear intuition why inclusion of spatial modeling of the residuals vastly improves the estimation. As we discussed already, borrowers of similar characteristics tend to cluster together in neighborhoods. Given enough data, we should use prepayment history of only similar borrowers in modeling their termination behavior. Since we need a substantially larger data set to reliably estimate low probability events such as sale or default, we are forced to draw observations from remote

---

<sup>7</sup> Out of the 44 parameter estimates in the two models only 4 remain significant.

<sup>8</sup> There is no distinction between moving and censoring in the refinancing model.

<sup>9</sup> There is no distinction between refinancing and censoring in the moving model.

neighborhoods whose residents might face very different refinancing or moving costs. Spatial modeling of the residuals is one way to reconcile these opposing forces and produce more statistically reliable and economically meaningful estimates.

Finally, we consider the sale versus default decision. Recall that we model sale and default as two vehicles to achieve a move. Therefore, we estimate a logit model of sale versus default, conditional on the decision to move. Out of the 239 move observations in our data set, 166 are sales and 73 are defaults.

Table 4 reports the parameter estimates of the logit model. All parameter estimates are consistent with our expectations. Perhaps most importantly, higher LTV ratio increases the default risk. Loans originated more recently have lower default rates than those originated in late 80's (between 1988 to 1990). As we would expect, the other parameter estimates are statistically insignificant, but in right sign, i.e. higher outstanding loan balance leads to higher default risks, and those who missed many profitable refinancing opportunities may face higher transaction costs and hence are less likely to default.

## **7.0 Conclusion**

In this paper we address the issue of spatially correlated unobservable variables influencing borrowers' mortgage termination behavior. We incorporate a semi-parametric model of the estimation errors into the three-stage maximum likelihood estimation methodology proposed by Deng and Quigley (2002). This approach yields two main benefits: it provides a closer fit to the data, and it allows for a number of interesting observations about the neighborhood effect of borrower behavior.

We find that, holding everything else constant, borrowers from the affluent West Side of Los Angeles County are more likely to refinance or move than predicted by the base estimation. Following our theoretical model, we conclude that those borrowers face lower refinancing or moving transaction costs. Furthermore, borrowers from lower-income neighborhoods tend to keep their mortgages for substantially longer periods, again holding everything else constant. This indicates that those borrowers face relatively higher mobility and refinancing transaction costs.

Such findings are important because refinancing and mobility costs have a substantial impact on the market value of the mortgages. Higher transaction costs for the borrower increase the value of the mortgage. Competitive lending markets would ultimately pass this higher value to the borrowers in terms of lower mortgage rates. Detailed analysis of the spatial distribution of the unobserved variables related to mortgage termination would allow for such savings and ultimately improve the efficiency as well as the equity of the lending markets.

## References

- Barlow, W. E., and R. L. Prentice (1988): "Residuals for Relative Risk Regression," *Biometrics* 75, 65-74.
- Capone, Charles A. (2001): "Introduction to the Special Issue on Mortgage Modeling," *Journal of Real Estate Finance and Economics* 23(2), 131-137.
- Cauley, S. (1996): "California Real Estate Market," *Quarterly UCLA-Anderson Business Forecast*.
- Clapp, John M., Gerson M. Goldberg, John P. Harding and Michael LaCour-Little (2001): "Movers and Shuckers: Interdependent Prepayment Decisions," *Real Estate Economics* 29(3) 411-450.
- Deng, Yongheng, and John M. Quigley (2002): "Woodheads Behavior and the Pricing of Residential Mortgages," *University of Southern California, Lusk Center Working Paper Series* 2003-1005.
- Deng, Yongheng, John M. Quigley, and Robert Van Order (1996): "Mortgage Default and Low Downpayment Loans: The Cost of Public Subsidy," *Regional Science and Urban Economics* 26(3-4) 263-285.
- Deng, Yongheng, John M. Quigley, and Robert Van Order (2000): "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options," *Econometrica* 68(2), 275-307.
- Kau, J.B., and D.C. Keenan (1995): "An Overview of the Option-Theoretic Pricing of Mortgages," *Journal of Housing Research* 6(2), 217-244.
- Kau, J.B., D.C. Keenan, W.J. Muller, and J.F. Epperson (1995): "The Valuation at Origination of Fixed Rate Mortgages with Default and Prepayment," *Journal of Real Estate Finance and Economics* 11(1), 5-36.
- Pavlov, A. (2000): "Space Varying Regression Coefficients: A Semi-Parametric Approach Applied to Real Estate Markets," *Real Estate Economics* 28(2), 249-283.
- Pavlov, A. (2001): "Competing Risks of Mortgage Termination: Who Refinances, Who Moves, and Who Defaults?" *Journal of Real Estate Finance and Economics* 23(2), 185-211.
- Quigley, John M. (1987): "Interest Rate Variations, Mortgage Prepayments and Household Mobility," *Review of Economics and Statistics* 69(4), 636-643.
- Richard, S. F. and R. Roll (1989): "Prepayments on Fixed Rate Mortgage-Backed Securities," *Journal of Portfolio Management* 15, 73-82.

Schwartz, E.S., and W.N. Torous (1989): "Prepayments on Fixed Rate Mortgage-Backed Securities," *Journal of Finance* 44(2), 357-392.

Stanton, R. (1996): "Unobservable Heterogeneity and Rational Learning: Pool Specific vs. Generic Mortgage-Backed Security Prices," *Journal of Real Estate Finance and Economics* 12, 243-263.

Therneau, T. M., P. M. Grambsch, and T. R. Fleming (1990): "Martingale Based Residuals for Survival Models," *Biometrika* 77, 147-160.

Turbull, G. (1995): *Urban Consumer Theory*. Washington, DC: Urban Institute Press.

TABLE 1—DISTRIBUTION OF TIME-CONSTANT CATEGORICAL VARIABLE

Variables	All Loans	Refinance	Move	Active *
Full Document	1,960 (53.48)	1,288 (50.14)	162 (67.78)	510 (59.51)
Reduced Document	1,705 (46.52)	1,281 (49.86)	177 (32.22)	347 (40.49)
Originated in 1988~1990	541 (14.76)	521 (20.28)	17 (7.11)	3 (0.35)
Originated in 1991~1992	1358 (37.05)	1210 (47.10)	57 (23.85)	91 (10.62)
Originated in 1993	1,099 (29.99)	449 (17.48)	111 (46.44)	539 (62.89)
Originated in 1994~1996	667 (18.20)	389 (15.14)	54 (22.59)	224 (26.14)
Number of Loans	3,665	2,569	239	857

*Note:* Percentage of the population for each category is in parentheses.

\*Active refers to those still outstanding at the end of the observation period.

TABLE 2—MEANS AND STANDARD ERRORS OF COVARIATES AT ORIGINATION AND TERMINATION

Variables	At Origination				At Termination	
	All Loans	Refinance	Move	Active*	Refinance	Move
Interest Rate Ratio	1.0000 (0.000)	1.0000 (0.000)	1.0000 (0.000)	1.0000 (0.000)	1.1529 (0.111)	1.0605 (0.138)
Yield Curve Slope	2.8525 (1.158)	2.7560 (1.264)	2.9518 (1.063)	3.1139 (0.730)	2.5858 (1.460)	0.9902 (0.895)
Outstanding Balance (in million dollars)	0.3224 (0.132)	0.3303 (0.135)	0.2828 (0.096)	0.3097 (0.128)	0.3210 (0.132)	0.2668 (0.092)
Woodhead (measure of foregone profitable refinance opportunity)	0.0000 (0.000)	0.0000 (0.000)	0.0000 (0.000)	0.0000 (0.000)	5.0868 (12.224)	7.4184 (16.502)
Difference between current and original LTV	0.4773 (1.232)	0.4674 (1.281)	0.5197 (1.487)	0.4953 (0.979)	3.7414 (5.326)	1.6999 (8.064)
Original High LTV	1.6851 (3.704)	1.4185 (3.44`)	4.0787 (4.904)	1.8167 (3.832)	1.4185 (3.44`)	4.0787 (4.904)
Number of Observations	3,665	2,569	239	857	2,569	239

*Note:* Standard Errors are in parentheses.

\*Active includes those outstanding at the end of the observation period.



TABLE 3—MLE AND 3SMLE ESTIMATES FOR COMPETING RISKS  
OF RESIDENTIAL MORTGAGE REFINANCE AND MOVE

	Model 1		Model 2		Model 3	
	Refinance	Move	Refinance	Move	Refinance	Move
Interest Rate Ratio	7.866 (36.18)	3.594 (4.74)	7.940 (36.21)	3.656 (4.60)	7.608 (31.71)	3.626 (4.36)
Yield Curve Slope	0.494 (23.52)	-0.516 (5.80)	0.496 (23.35)	-0.511 (5.51)	0.414 (17.19)	-0.531 (5.58)
Outstanding Balance (in million dollars)	0.779 (6.35)	-1.524 (2.06)	0.500 (3.52)	-1.846 (2.17)	0.397 (2.67)	-1.625 (1.76)
Woodhead (measure of foregone profitable refinance opportunity)	-0.018 (7.01)	0.008 (1.44)	-0.018 (6.55)	0.008 (1.47)	-0.015 (4.21)	0.008 (1.30)
Difference between current and original LTV	-0.057 (14.85)	-0.037 (3.60)	-0.056 (14.42)	-0.039 (3.65)	-0.063 (15.67)	-0.046 (4.31)
Original High LTV	-0.030 (5.04)	0.076 (4.58)	-0.025 (4.09)	0.078 (4.49)	-0.021 (2.89)	0.063 (3.19)
Full Document	0.042 (1.01)	0.048 (0.28)	0.057 (1.36)	0.057 (0.32)	0.060 (1.30)	0.032 (0.16)
Originated in 1991~1992	0.934 (18.42)	0.152 (0.44)	0.921 (17.85)	0.209 (0.57)	0.950 (17.50)	0.217 (0.55)
Originated in 1993	1.050 (11.03)	0.426 (0.99)	1.031 (10.81)	0.478 (1.04)	1.307 (12.08)	0.443 (0.89)
Originated in 1994~1996	1.034 (11.15)	0.176 (0.42)	1.001 (10.70)	0.230 (0.52)	1.105 (10.14)	0.210 (0.44)
Neighborhood 1 (dummy variable)	- -	- -	0.344 (2.86)	0.287 (0.70)	0.136 (1.05)	0.241 (0.53)
Neighborhood 2 (dummy variable)	- -	- -	-0.015 (0.10)	0.686 (1.33)	-0.095 (0.56)	0.525 (0.88)
Neighborhood 3 (dummy variable)	- -	- -	0.101 (0.62)	0.383 (0.84)	0.031 (0.17)	0.361 (0.71)
Neighborhood 4 (dummy variable)	- -	- -	0.291 (2.02)	0.383 (0.72)	0.135 (0.88)	0.361 (0.62)
Neighborhood 5 (dummy variable)	- -	- -	0.261 (2.07)	0.784 (2.26)	0.115 (0.84)	0.665 (1.66)
Neighborhood 6 (dummy variable)	- -	- -	-0.245 (0.86)	0.170 (0.23)	-0.074 (0.22)	-0.045 (0.05)

TABLE 3—Continued.

	Model 1		Model 2		Model 3	
	Refinance	Move	Refinance	Move	Refinance	Move
Neighborhood 7 (dummy variable)	-	-	0.361 (1.49)	0.051 (0.07)	0.248 (0.90)	0.020 (0.02)
Neighborhood 8 (dummy variable)	-	-	0.483 (4.16)	-0.390 (0.70)	0.341 (2.72)	-0.407 (0.68)
Neighborhood 9 (dummy variable)	-	-	0.480 (3.70)	0.806 (1.76)	0.283 (2.05)	0.654 (1.26)
Neighborhood 10 (dummy variable)	-	-	0.449 (4.14)	0.664 (1.85)	0.278 (2.38)	0.539 (1.33)
Neighborhood 11 (dummy variable)	-	-	0.169 (0.91)	0.501 (1.18)	0.122 (0.57)	0.420 (0.86)
Neighborhood 12 (dummy variable)	-	-	0.107 (0.78)	0.790 (1.81)	0.101 (0.69)	0.643 (1.28)
Neighborhood 13 (dummy variable)	-	-	0.615 (3.90)	-0.510 (0.67)	0.516 (2.89)	-0.524 (0.65)
Neighborhood 14 (dummy variable)	-	-	0.157 (1.17)	0.428 (0.91)	0.123 (0.81)	0.301 (0.55)
Neighborhood 15 (dummy variable)	-	-	0.292 (2.41)	0.154 (0.34)	0.226 (1.69)	0.088 (0.18)
Neighborhood 16 (dummy variable)	-	-	-0.015 (0.08)	0.009 (0.01)	-0.108 (0.57)	0.039 (0.06)
Neighborhood 17 (dummy variable)	-	-	0.266 (2.41)	0.379 (1.09)	0.159 (1.33)	0.297 (0.75)
Neighborhood 18 (dummy variable)	-	-	0.333 (2.99)	0.303 (0.81)	0.230 (1.95)	0.248 (0.59)
Neighborhood 19 (dummy variable)	-	-	0.047 (0.32)	0.301 (0.61)	0.032 (0.18)	0.284 (0.53)
Neighborhood 20 (dummy variable)	-	-	0.227 (1.78)	0.997 (2.76)	0.135 (1.00)	0.818 (1.98)
Neighborhood 21 (dummy variable)	-	-	0.159 (1.12)	0.680 (1.61)	0.139 (0.92)	0.525 (1.06)
Log Likelihood	-13,328		-13,285		-12,452	

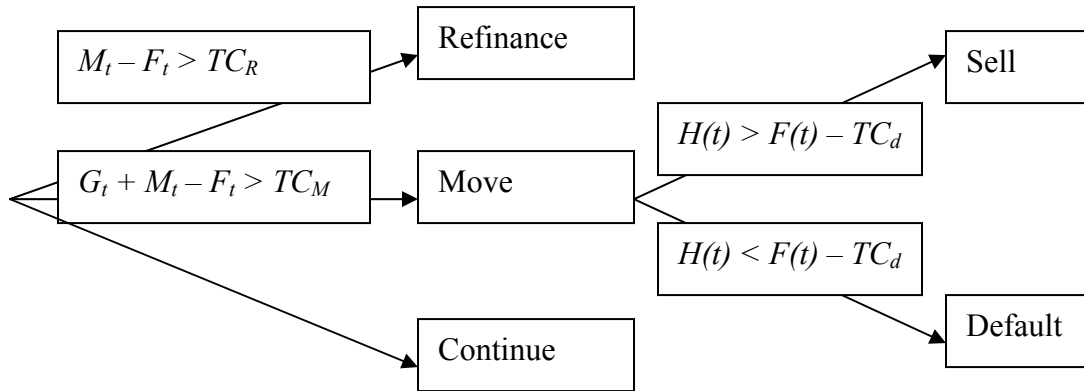
Note: t-ratios are in parentheses. Models 1 and 2 are estimated by MLE approach. Model 3 is estimated using a 3SMLE approach that accounts for spatially correlated neighborhood random effects during the maximum likelihood estimation. Refinance and move functions are considered as correlated competing risks and they are estimated jointly. Baseline hazard function for refinance and move are defined in our estimation following the Public Securities Association (PSA) specification. The PSA curve specifies the baseline hazard curve begins at 0.2 percent annual constant hazard rate in the first month and increases by 0.2 percent in each successive month until month 30, when the series levels out at 6 percent per year until maturity.

TABLE 4 – MLE ESTIMATES OF LOGIT MODEL FOR DEFAULT VS SELL

Intercept	0.153 (0.16)
Outstanding Balance (in million dollars)	1.384 (0.83)
Woodhead (measure of foregone profitable refinance opportunity)	-0.002 (0.18)
Difference between current and original LTV	0.131 (4.24)
Original High LTV	0.062 (1.58)
Full Document	-0.521 (1.27)
Originated in 1991~1992	-1.739 (2.19)
Originated in 1993	-1.735 (1.97)
Originated in 1994~1996	-1.904 (2.07)
Log Likelihood	-118

*Note:* t-ratios are in parentheses. Estimation results are based on a sample of 239 loans conditioning on move, among which 166 loans were prepaid due to sale of the property and 73 loans were in default.

**Figure 1: Mortgage Termination Model**



**Figure 2: Refinancing Residuals: Termination cause = Refinancing**

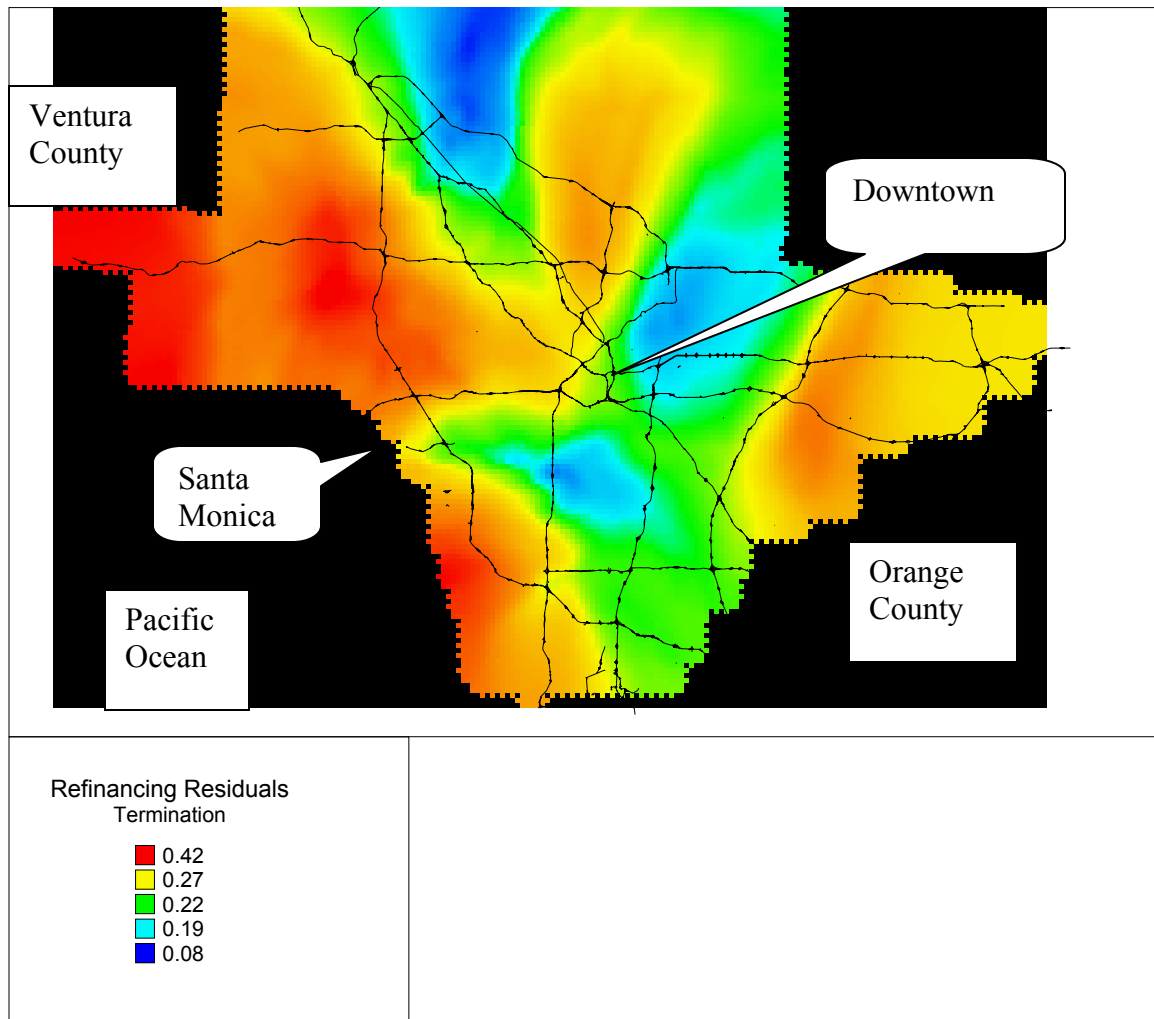


Figure 2 depicts the refinancing residuals from the second stage of our three-stage SMLE approach for loans that terminated because of refinancing. Above average residuals (red areas) indicate that borrowers from those regions tend to refinance more readily than predicted by a standard model.

**Figure 3: Refinancing residuals: Termination cause = Moving or Censored**

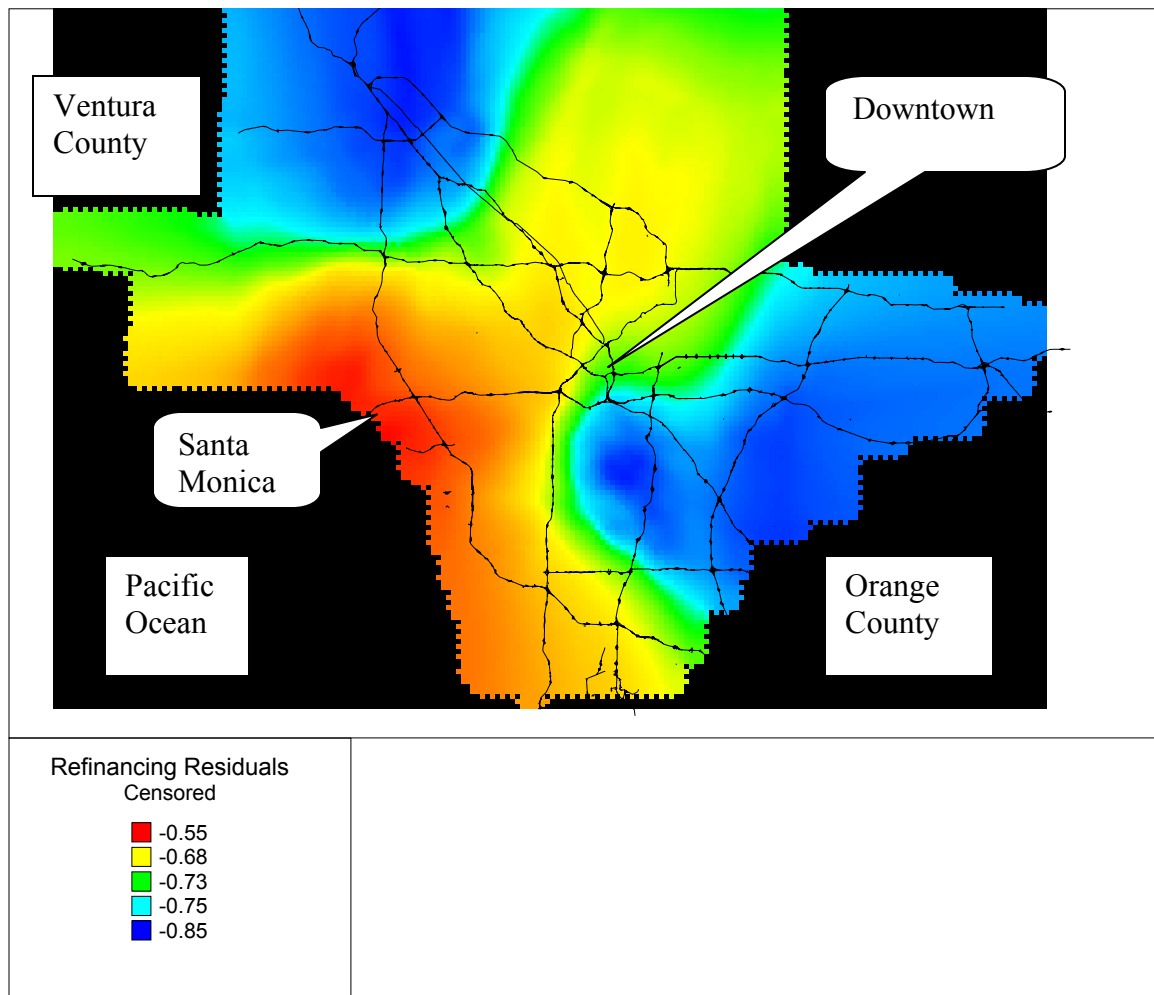


Figure 3 depicts the refinancing residuals from the second stage of our three-stage SMLE approach for loans that terminated because of moving or were censored. Above average residuals (red areas) indicate that borrowers from those regions tend to refinance more readily than predicted by a standard model.

**Figure 4: Moving residuals: Termination cause = Moving**

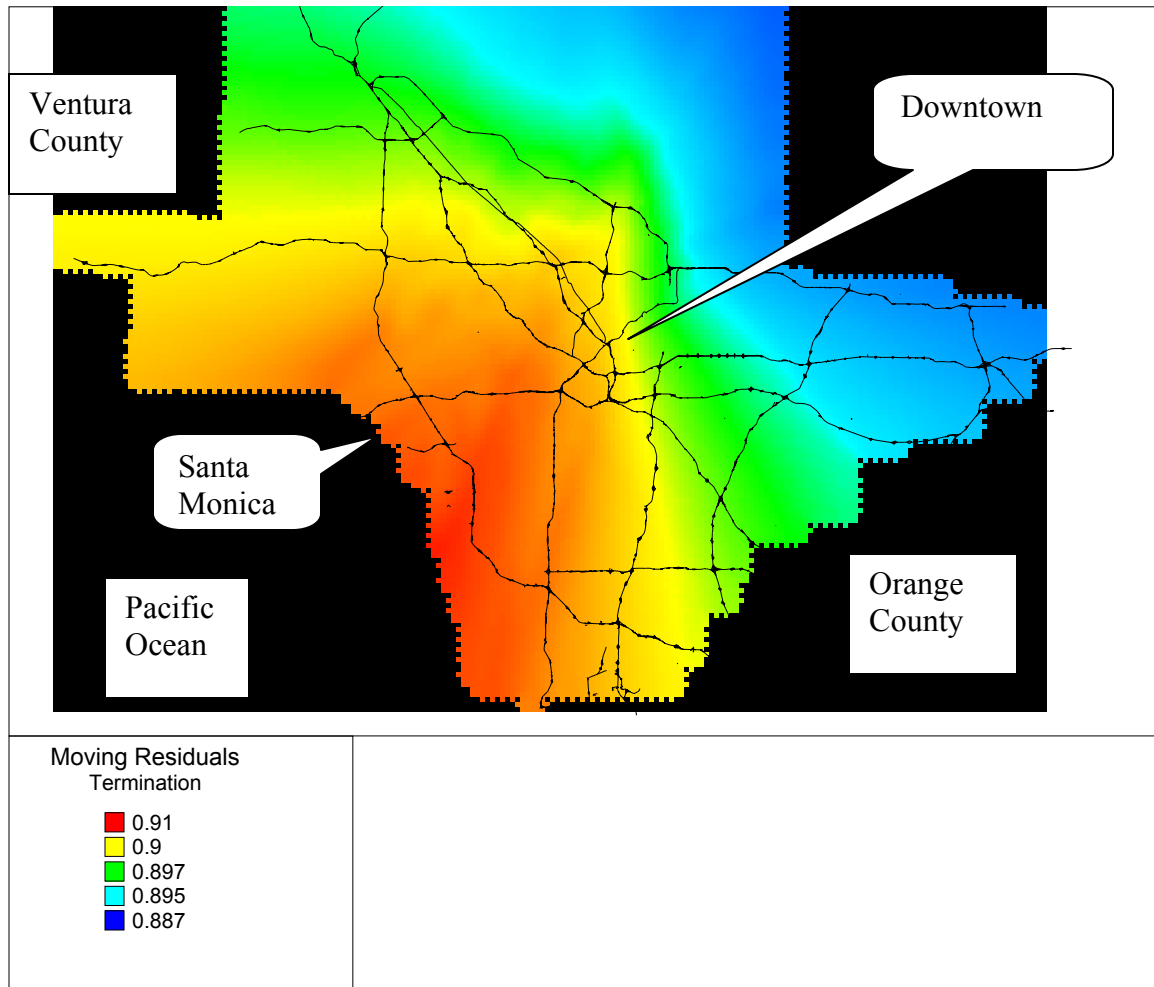


Figure 4 depicts the moving residuals from the second stage of our three-stage SMLE approach for loans that terminated because of moving. Above average residuals (red areas) indicate that borrowers from those regions tend to move more readily than predicted by a standard model.

**Figure 5: Moving residuals: Termination cause = Refinancing or Censored**

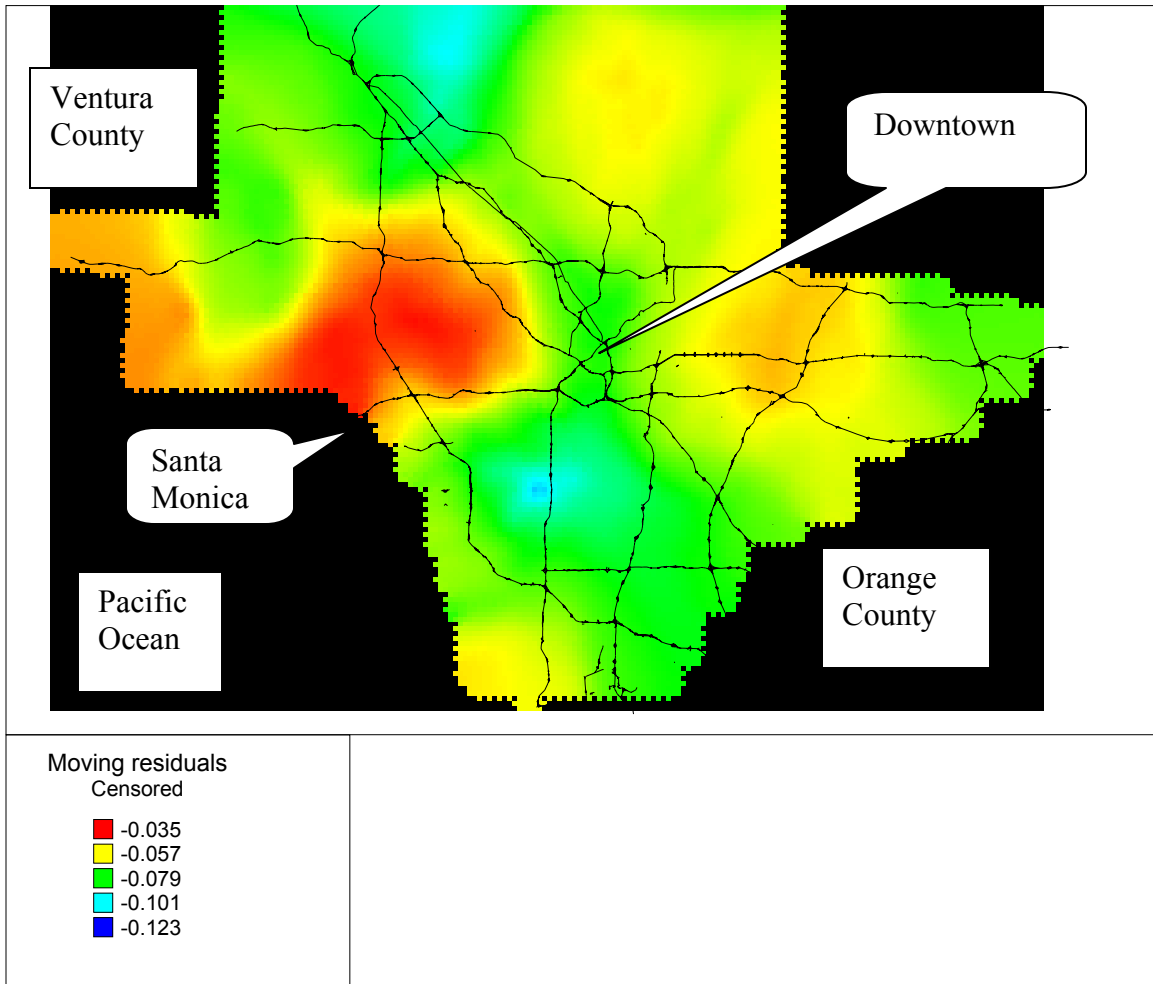


Figure 5 depicts the moving residuals from the second stage of our three-stage SMLE approach for loans that terminated because of refinancing or were censored. Above average residuals (red areas) indicate that borrowers from those regions tend to move more readily than predicted by a standard model.