FHA Loans in Foreclosure Proceedings: Distinguishing Sources of Interdependence in Competing Risks

Ran Deng [†] and Shermineh Haghani *,[†]

Office of the Comptroller of the Currency, 400 7th Street S.W., Washington, DC 20219, USA;
ran.deng@occ.treas.gov
* Correspondence: shermineh.haghani@occ.treas.gov; Tel.: +1-202-649-6296
† The views in this paper are those of the authors and do not necessarily reflect those of the Office of the Comptroller of the Currency or the Department of Treasury.

Abstract: A mortgage borrower has several options once a foreclosure proceedings is initiated, mainly default and prepayment. Using a sample of FHA mortgage loans, we develop a dependent competing risks framework to examine the determinants of time to default and time to prepayment once the foreclosure proceedings is initiated. More importantly, we examine the interdependence between default and prepayment, through both the correlation of the unobserved heterogeneity terms and the preventive behavior of the individual mortgage borrowers. We find that time to default and time to prepayment are affected by several factors, such as the Loan-To-Value ratio (LTV), FICO score and unemployment rate. In addition, we find strong evidence that supports the existence of interdependence between the default and prepayment hazards through both the correlation of the unobserved heterogeneity terms and the preventive behavior of individual mortgage borrowers. We show that neglecting the interdependence through the preventive behavior of the individual mortgage borrowers can lead to biased estimates and misleading inference.

1. Introduction

A mortgage borrower is technically delinquent once a monthly mortgage payment due date is missed. Most lenders, however, give the borrower a substantial period of time (typically 90 days, but varying by lender) to bring the loan into current status by making up all the missed payments plus the associated late fees. If the borrower is still delinquent after a certain time period, the lender initiates a foreclosure proceedings. Loans that are in foreclosure proceedings are not fully terminated. In fact, some of these loans can be reinstated, prepaid or modified (extended term or other alterations to lower the monthly payment), or have other alternative outcomes. These outcomes can be considered as competing risks. In this paper, we examine the lifetime of an FHA mortgage loan from the onset of foreclosure until one of the main types of outcomes is observed. In particular, we examine the probability that an FHA mortgage loan in a foreclosure proceedings will eventually be prepaid or defaulted to Real Estate Owned (REO).[1]

---

[1] It is the U.S. Department of Housing and Urban Development (HUD) that takes ownership of any properties that complete the foreclosure process for FHA mortgage loans.

There is an extensive literature on mortgage terminations. The focus of most of the early studies was on mortgage terminations either due to prepayment (e.g., Green and Shoven 1986; Schwartz and Torous 1989; Quigley and Van Order 1990) or due to default (e.g., Cunningham and Hendershott 1984; Quigley and Van Order 1995), but not both. A number of theoretical papers emphasized the importance of jointly estimating default and prepayment (e.g., Kau and Keenan 1996; Titman and Torous 1989). The work in Foster and Van Order (1985) was among the first papers that simultaneously estimated both default and prepayment for FHA loans. The work in Schwartz and Torous (1993) applies Poisson regression to jointly estimate hazards for default and prepayment. The works in Deng et al. (1996) and Deng (1997) examine default and prepayment jointly using a competing risks model. The works in Deng et al. (2000) and Pennington-Cross (2006) also use a competing risks model, where they account for one possible type of interdependence between the default and prepayment hazards. This was through the correlation of associated unobserved heterogeneity terms for the purpose of capturing the unobservable loan-specific characteristics (such as the effect of borrowers' intentions and strategies) that affect both default and prepayment hazards. Such unobserved heterogeneity terms might induce either negative or positive interdependence between the default and prepayment hazards.

This paper adds to the existing literature by accounting for another possible type of interdependence between the default and prepayment hazards. This interdependence stems from the fact that the motives behind prepayments in the case of a mortgage for which the foreclosure proceedings is initiated are distinct from the traditional motives for prepayment. In particular, prepayments of a mortgage for which the foreclosure proceedings is initiated can be viewed as "distressed prepayments"[2] in which borrowers want to sell their homes to avoid a default outcome.[3] The reason to avoid default is its significant costs, such as legal fees and a negative credit report, that make prepayment a more attractive option to borrowers in foreclosure proceedings. Thus, if borrowers are in foreclosure proceedings and foresee themselves facing a high risk of default, they might increase their intensity to sell their homes to prepay, in order to avoid default. Such a kind of behavior by borrowers implies that higher risks of default might lead to higher probability of prepayment. This induces a positive correlation between the default and prepayment hazards.

In this paper, we specify a dependent competing risks framework to examine the interdependence between the default and prepayment hazards through both the correlation of the unobserved heterogeneity terms associated with each risk and the preventive behavior of individual mortgage borrowers. The interdependence between the hazards through the preventive behavior of individuals is referred to as the "structural" dependence in the literature (e.g., Rosholm and Svarer 2001). While the interdependence between the hazards through the correlation of the unobserved heterogeneity terms is a common practice, to our knowledge, there are no empirical studies that have examined the structural dependence in the mortgage literature. Not accounting for the structural dependence can bias the correlation; for example, if the structural dependence is not counted in the model, we might fail to detect any correlation, since the distinct driving forces of the interdependence between the default and prepayment hazards might cancel each other out. To allow for the structural dependence, we allow the default hazard, both the observable and the unobservable parts, to directly affect the prepayment hazard.

The most important finding of this paper is that default and prepayment hazards are interdependent in two distinct ways. First, we find a significant positive correlation between the unobserved heterogeneity terms. This finding suggests that there are some unobservable loan-specific characteristics that affect both default and prepayment hazards in the same direction. Second, we find a significant positive structural dependence, suggesting that higher risk of default leads to higher probability of prepayment. We show that neglecting the interdependence through both the correlation

---

2   This terminology is used by Danis and Pennington-Cross (2005).
3   The Federal Housing Administration (FHA) has a program called "Pre-Foreclosure Sales" that allows borrowers who are in foreclosure proceedings to sell their homes and to use the sales proceeds to satisfy the mortgage debt. Under this program, the debt is satisfied even if the sales proceeds are less than the loan balance owed.

of the unobserved heterogeneity terms and the preventive behavior of the individual mortgage borrowers can lead to biased estimates and misleading inference. As for the effects of covariates on the likelihood of default and prepayment, we find that loans with the following characteristics have a higher probability of default: more equity, low FICO score, high unemployment rate in the borrower's geographical area, short delinquency spells, nonjudicial states and positive interest rate spread. In addition, we find that loans with the following characteristics have higher probability to prepay: more equity, high FICO score, high unemployment rate, short delinquency spells, nonjudicial states and negative interest rate spread.

The remainder of this paper is organized as follows. Section 2 presents the data and some descriptive statistics. Section 3 presents the model. Section 4 reports the empirical results. Section 5 concludes.

2. Data Description and Summary Statistics

We use a panel dataset of first-lien residential mortgage loans obtained from the OCCMortgage Metrics data (OCCMM). OCCMM includes loans serviced by seven large banks and covers monthly loan performance from January 2008 until March 2016. The dataset consists of more than 21.1 million first-lien mortgage loans with $3.6 trillion in unpaid principal balances, which make up about 38 percent of all first-lien residential mortgage debt outstanding in the U.S.[4]

For the purposes of this paper, we focus on FHA loans[5] for which the foreclosure proceedings were initiated. There is a total of 231,800 of these loans in our sample of 3,359,573 FHA loans.[6] In our analysis, we exclude loans for which: (i) foreclosure proceedings end for reasons other than default and prepayment; (ii) servicing was transferred to different servicers; and (iii) values for explanatory variables are missing. Following the outlined exclusion criteria, our final sample size is 107,627, out of which 8974 are prepaid, 84,012 are defaulted and 14,641 are still in foreclosure proceedings as of March 2016.

We measure the lifetime of an FHA loan as the number of months from the onset of foreclosure until the loan is either defaulted or prepaid. We denote a loan that is still in a foreclosure proceedings at the end of the observation period as right censored and measure its lifetime as the number of months from the onset of foreclosure until March 2016. Table 1 presents descriptive statistics of the lifetime by status. The mean lifetime of a loan is about 15 months, which is higher than that reported by Pennington-Cross (2006) for subprime loans. The mean lifetime for prepaid loans is higher than for defaulted loans. This result suggests interdependence between default and prepayment. In other words, if default and prepayment were independent, presumably no loans would make it through to prepay, as they would on average reach the default option at 14 months before the prepay at 18 months. Thus, it might be misleading to consider independency between default and prepayment.

To examine how default and prepayment rates change with age, Figure 1 shows the smooth nonparametric estimation of the hazard function.[7] The figure shows that the probability of default increases during the first 18 months after foreclosure proceedings are initiated and then rapidly decreases. In addition, the figure shows that the probability of prepayment increases as foreclosure proceedings lengthen.

---

[4]　See the quarterly OCC/OTSMortgage Metrics Reports for further details on the dataset.

[5]　FHA loans are mortgage loans on which the lender is insured against loss by the Federal Housing Administration, with the borrower paying the mortgage insurance premiums. These loans offer low down payments and generous credit score and debt to income requirements.

[6]　Ninety three percent of the FHA loans in our sample are 30-year fixed rate loans.

[7]　The estimation is based on the Nelson–Aalen estimator.

Table 1. Statistics for lifetimes of individual FHA loans by status.

Status	Mean	Median	Std.Dev.	Min	Max	Qu. (25%)	Qu. (75%)
All Loans	15.21	10.00	14.11	1.00	78.00	5.00	21.00
Defaulted Loans	14.00	9.00	12.25	2.00	78.00	5.00	19.00
Prepaid Loans	18.27	12.00	16.36	2.00	78.00	5.00	28.00

The table provides the descriptive statistics of lifetimes (in months) of all FHA loans from January 2008 to March 2016. The descriptive statistics include the mean, median, standard deviation, minimum, maximum, 25% quartile and 75% quartile. All loans refer to all defaulted, prepaid and right censored loans in the sample.
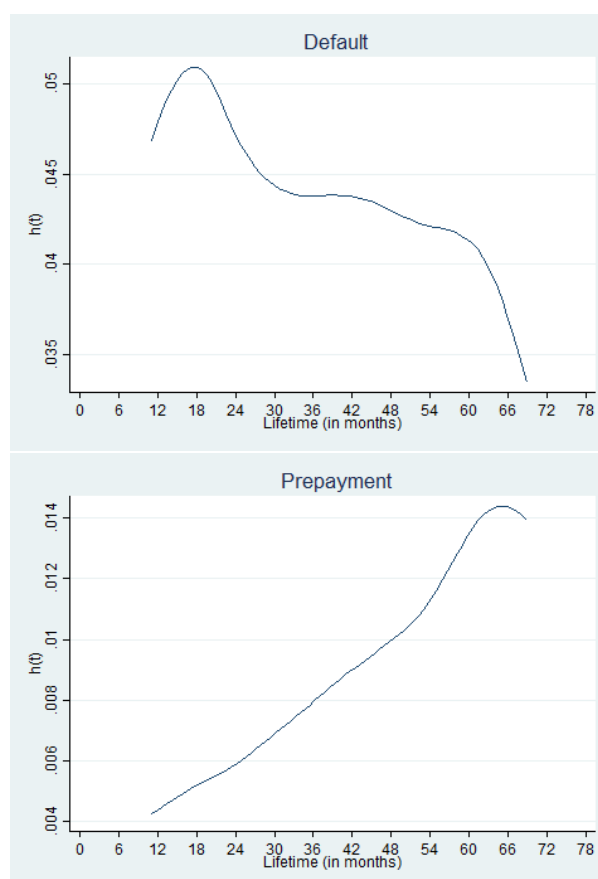

Figure 1. Smoothed nonparametric hazard Function. The figures display the smooth nonparametric estimation of default and prepayment hazard functions. The estimate is based on the Nelson–Aalen estimator. To smooth the Nelson–Aalen estimator, we specify an Epanechnikov kernel function with the default bandwidth in STATA.

The following explanatory variables are used to examine the determinants of default and prepayment hazards and their interdependence:

- LTV: To measure equity remaining in the property, we calculate Loan-To-Value ratio (LTV) using the current balance of the loan in each month and the estimated property value.[8]
- FICO score: To proxy for the overall borrower's creditworthiness, we use the borrower's FICO score in each month.
- Unemployment rate: To proxy for financial instability, we use the seasonally-adjusted monthly unemployment rate lagged by six months in the state where the property is located.[9]

---

[8] The estimated property value is obtained from the Lender Processing Services (LPS) Home Price Index (HPI).

[9] The seasonally-adjusted monthly unemployment rate is obtained from the Bureau of Labor Statistics (BLS).

- Delinquency spell: To measure delinquency behavior, we calculate the fraction of months in delinquency prior to the beginning of a foreclosure proceedings.
- Judicial status: To examine state foreclosure laws, we use an indicator variable equal to one if the state is a judicial foreclosure state, and zero otherwise.[10]
- Interest rate spread: To measure the change in the market interest rate, we use an indicator variable equal to one if the current interest rate is higher than the current 30-year fixed rate, and zero otherwise.[11,12].

Table 2 provides summary statistics of the explanatory variables. A quick comparison shows that the average characteristics are different between default and prepayment. In particular, on average, defaulted loans have less equity, lower FICO score and higher unemployment rate.

Table 2. Statistics of the explanatory variables by status. LTV, Loan-To-Value ratio.

	LTV			
Status	**Mean**	**Std.Dev.**	**Min**	**Max**
All Loans	95.48	6.14	80.91	100.58
Defaulted Loans	99.47	4.40	88.97	103.51
Prepaid Loans	75.71	27.17	5.38	91.64
	FICO Score			
Status	**Mean**	**Std.Dev.**	**Min**	**Max**
All Loans	540.20	18.31	516.51	567.38
Defaulted Loans	538.76	17.57	516.54	564.87
Prepaid Loans	556.64	23.27	523.40	590.85
	Unemployment Rate			
Status	**Mean**	**Std.Dev.**	**Min**	**Max**
All Loans	7.84	0.56	6.97	8.74
Defaulted Loans	8.20	0.54	7.38	9.06
Prepaid Loans	7.70	0.62	6.70	8.72
	Delinquency Spell			
Status	**Mean**	**Std.Dev.**	**Min**	**Max**
All Loans	0.41	0.23	0.02	1
Defaulted Loans	0.41	0.23	0.02	1
Prepaid Loans	0.40	0.22	0.03	1
	Judicial Status			
Status	**Mean**	**Std.Dev.**	**Min**	**Max**
All Loans	0.41	0.49	0	1
Defaulted Loans	0.37	0.48	0	1
Prepaid Loans	0.37	0.48	0	1
	Interest Rate Spread			
Status	**Mean**	**Std.Dev.**	**Min**	**Max**
All Loans	0.94	0.04	0.88	0.97
Defaulted Loans	0.95	0.03	0.91	0.98
Prepaid Loans	0.89	0.07	0.79	0.96

The table provides the descriptive statistics for LTV, FICO score, unemployment rate, delinquency spell, judicial states, and interest rate spread from January 2008 to March 2016. The descriptive statistics include the mean, standard deviation, minimum and maximum. The reported descriptive statistics for time-dependent variables are averaged over loans. All loans refer to all defaulted, prepaid and right censored loans in the sample.

---

[10] We identified judicial states using RealtyTrac.com and FindLaw.com.
[11] The 30-year fixed rate is obtained from primary mortgage market survey.
[12] About 94 percent of the loans are 30-year fixed rate mortgages.

3. Econometric Methodology

In this section, we propose a dependent competing risks duration model that is capable of incorporating time-varying covariates and censored observations easily. More importantly, the model controls for unobserved covariates, allows for estimating the default and prepayment hazards jointly and accounts for the interdependence of these hazards through both the correlation of the unobserved heterogeneity terms and the preventive behavior of individual mortgage borrowers. We first describe the specification of the model and then derive the likelihood function.

3.1. Model Specification

There are two main options available to the mortgage borrowers once foreclosure proceedings are initiated; namely, default (D) and prepayment (P). The prepayment (P) option can be viewed as "distressed prepayments" since borrowers in the foreclosure proceedings want to sell their homes to avoid a default outcome. Suppose that nonnegative random variables T_D and T_P are the potential lifetimes from the onset of foreclosure until default (D) and prepayment (P), respectively. In the competing risks framework, only the shortest lifetime is actually observed; that is $T = \min[T_D, T_P]$ and the corresponding actual event type, $J \in \{D, P\}$. Let $x(t)$ be a vector of observable covariates at time t and $v = (v_D, v_P)$ be a vector of unobservable covariates. The advantage of introducing two unobservable covariates (also called unobserved heterogeneity terms or frailties) is the possibility of exploring the dependence between the default and prepayment hazards, whenever v_D and v_P are positively or negatively correlated. In particular, this specification avoids using a restrictive one-factor model (e.g., Flinn and Heckman 1982; Clayton and Cuzick 1985; Heckman and Walker 1990) and so does not restrict the sign of dependence when a sufficiently flexible class of joint distributions is chosen for the unobserved heterogeneity terms.

Before elaborating the model specification, we list the regularity assumptions where index $i, i = 1, \ldots, n$, denotes individual mortgage loans:

Assumption 1. *(a) The unobserved heterogeneity terms are time invariant and depend on the individual mortgage loans i. (b) The individual heterogeneities* $(v_{iD}, v_{iP}), i = 1, \ldots, n$, *are independent and have the same distribution* $G(v_D, v_P)$.[13]

Assumption 2. *The potential lifetimes* T_{iD} *and* $T_{iP}, i = 1, \ldots, n$, *are independent conditional on the observable covariate histories* $X_i = \{x_i(t), t \in \mathcal{N}\}, i = 1, \ldots, n$, *and on heterogeneities* $(v_{iD}, v_{iP}), i = 1, \ldots, n$.

Assumption 3. *The individual heterogeneities are independent of the covariate histories.*

Assumption 4. *The variables* T_{iD} *(resp.* $T_{iP}), i = 1, \ldots, n$, *have identical conditional distributions given the individual covariate histories and the individual unobserved heterogeneities.*[14]

Assumption 5. *The type-specific hazard functions conditional on* $(x_i(t), v_{iD}, v_{iP}), i = 1, \ldots, n$, *are mixed proportional hazard functions:*

$$h_D(t; x(t), v_D) = h_{0D}(t) \exp(x(t)' \beta_D) \exp(v_D),$$
$$h_P(t; x(t), v_P, v_D) = h_{0P}(t) \exp(x(t)' \beta_P) \exp(v_P) \exp(\gamma . \log(h_D(t; x(t), v_D))). \tag{1}$$

---

[13] This assumption is commonly imposed in microeconomic studies, and it indicates that the focus of the analysis is on individual omitted heterogeneity. It implies that individual heterogeneities that depend on both individual loans and time are excluded. This allows us to assume away the moral hazard phenomena (e.g., Gourieroux and Jasiak 2004) and the omitted dynamic variables. The omitted time-dependent variables could be loan-specific or common to all loans. The analysis of these unobserved variables is left for further research.

[14] Assumptions 2 to 4 are standard.

where β_D and β_P are type-specific regression coefficients' vectors and h_{0D} and h_{0P} are the type-specific baseline hazard functions. The parameter γ captures the structural dependence of the prepayment hazard rate on the default probability.

Equation (1) accounts for interdependence between default and prepayment hazards in two ways, through: (1) the correlation of the unobserved heterogeneity terms; default and prepayment hazards might share similar or distinctive unobserved loan-specific characteristics that are identified by the negative or positive correlation of v_D and v_P; and (2) the structural dependence; borrowers who are in foreclosure proceedings and foresee themselves facing a high risk of default might increase their intensity to sell their homes to prepay in order to avoid default. If this hypothesis is true, we should expect γ to be significant and positive.

The model defined by Equation (1) nests three restricted models that are generally used in applied studies. The first restriction can be imposed to the general model by specifying $\gamma = 0$, which eliminates the structural dependence of the default and prepayment hazards and allows the interdependence between the hazards only through unobserved heterogeneity terms. The next restriction can be applied by assuming that the unobserved heterogeneity terms v_D and v_P are independent (i.e., $v_D \perp v_P$). This is a common assumption in empirical competing risks studies.[15] The last restriction can completely ignore unobserved heterogeneity terms. To illustrate the advantage of the model defined by Equation (1) and the potential bias of the restricted models, all four models are estimated.

Assumption 6. *The baseline hazard functions follow an expo-power distribution:*

$$h_{0j}(t) = \alpha_j t^{\alpha_j - 1} exp(\theta_j t^{\alpha_j}),\tag{2}$$

where $j = D, P, \alpha_j > 0, -\infty < \theta_j < +\infty$.

This parametric specification was introduced by Saha and Hilton (1997). It can represent a variety of patterns of the hazard function, including constant, monotonically increasing, monotonically decreasing, U-shaped, inverted U-shaped or display humps. It includes as a special case the Weibull hazard function for $\theta = 0$, which is monotone. For $\theta \neq 0$, the hazard function has a turning point at $[(1 - \alpha_j)/(\alpha_j \theta_j)]^{1/\alpha_j}$.

Conditional on the observable covariate histories, the distributions of the uncensored and right censored observations are characterized by the probabilities $Pr(t \leq T < t + \Delta t, J = j|X(t))$ and $Pr(T > c|X(c))$, respectively. These probabilities are obtained by integrating out v_D and v_P:

---

[15]   One of the main reasons for these studies to make the independence assumption, in addition to computational convenience, is the common misunderstanding that dependent competing risks' specifications are not identifiable. This non-identifiability property is studied in detail by Tsiatis (1975), who proves that for any joint survival function with arbitrary dependence between the competing risks, one can find a different joint survival function with independent competing risks. If that is the case, then there is no point in complicating the model with the dependence assumption because the data cannot test for it anyway. However, Tsiatis's argument is valid only if the sample is homogenous. Thus, the problem of non-identifiability can be resolved by introducing heterogeneity through the variation of the observed covariates, as discussed at length by Heckman and Honore (1989), Abbring and Van den Berg (2003) and Colby and Rilstone (2004).

$$Pr(t \leq T < t + \Delta t, J = j|X(t)) = \int_{v_D} \int_{v_P} Pr(t \leq T < t + \Delta t, J = j|X(t), v_D, v_P) dG(v_D, v_P)$$

where $j = D, P$

and $Pr(t \leq T < t + \Delta t, J = D|X(t), v_D, v_P)$

$$= h_D(t; x(t), v_D) \exp\left(-\int_0^t h_D(u; x(u), v_D) du\right) \exp\left(-\int_0^t h_P(u; x(u), v_P, v_D) du\right) \Delta t.$$

and $Pr(t \leq T < t + \Delta t, J = P|X(t), v_D, v_P)$

$$= h_P(t; x(t), v_P, v_D) \exp\left(-\int_0^t h_D(u; x(u), v_D) du\right) \exp\left(-\int_0^t h_P(u; x(u), v_P, v_D) du\right) \Delta t. \qquad (3)$$

This quantity depends on the covariate histories up to time t only. In addition, we have:

$$Pr(T > c|X(c)) = \int_{v_D} \int_{v_P} Pr(T > c|X(c), v_D, v_P) dG(v_D, v_P),$$

where $Pr(T > c|X(c), v_D, v_P) = \exp\left(-\int_0^c h_D(u; x(u), v_D) du\right) \exp\left(-\int_0^c h_P(u; x(u), v_P, v_D) du\right). \quad (4)$

This quantity depends on the covariate history up to time c only.

In practice, the model has to be completed by specifying the joint distribution of the unobserved heterogeneity terms. In this subsection, we use an extension of the approach of Heckman and Singer (1984) (see also Nickell 1979; Van den Berg et al. 2004) and assume the following:

Assumption 7. *The joint distribution of the unobserved heterogeneity terms is bivariate discrete in which v_D and v_P can only take two values. Let v_D^1 and v_D^2 denote the values of v_D and v_P^1 and v_P^2 denote the values of v_P. Conditional on covariate histories, the set of individual mortgage loans can be divided into four classes that correspond to (v_D^1, v_P^1), (v_D^1, v_P^2), (v_D^2, v_P^1) and (v_D^2, v_P^2), respectively. The sizes of these classes are unknown a priori and will be approximated by means of their associated probability estimates. Under Assumption 7, the joint distribution of v_D, v_P is characterized by the following elementary probabilities:*

$$Pr(v_D = v_D^1, v_P = v_P^1) = p_{11}, \quad Pr(v_D = v_D^1, v_P = v_P^2) = p_{12},$$
$$Pr(v_D = v_D^2, v_P = v_P^1) = p_{21}, \quad Pr(v_D = v_D^2, v_P = v_P^2) = p_{22}.$$

with $0 \leq p_{kl} \leq 1$ and $\sum_{k=1}^2 \sum_{l=1}^2 p_{kl} = 1$ for $k, l = 1, 2$.[16,17]

Under Assumption 7, the characteristics of the uncensored and right censored distributions become:

$$Pr(t \leq T < t + \Delta t, J = j|X(t)) = \sum_{k=1}^2 \sum_{l=1}^2 Pr(t \leq T < t + \Delta t, J = j|X(t), v_D^k, v_P^l) p_{kl}, \qquad (5)$$

---

[16] To ensure that the probabilities lie between $[0, 1]$ and sum up to one, we apply the logistic transformation, i.e.,

$$p_{kl} = \frac{\exp(q_{kl})}{\sum_{k=1}^2 \sum_{l=1}^2 \exp(q_{kl})}.$$

where $-\infty < q_{kl} < +\infty$, for $k, l = 1, 2$.

[17] The covariance of v_D and v_P can be derived as (see Van den Berg et al. 1994): $Cov(v_D, v_P) = (p_{11}p_{22} - p_{12}p_{21})(v_D^1 - v_D^2)(v_P^1 - v_P^2)$. Therefore, the correlation between v_D and v_P becomes:

$$\rho(v_D, v_P) = \frac{p_{11}p_{22} - p_{12}p_{21}}{\sqrt{(p_{11} + p_{12})(p_{11} + p_{21})(p_{22} + p_{12})(p_{22} + p_{21})}}.$$

Variables v_D and v_P will be perfectly correlated if either $p_{12} = p_{21} = 0$, or $p_{11} = p_{22} = 0$. Further, v_D and v_P are independent if and only if $p_{11}p_{22} - p_{12}p_{21} = 0$.

and

$$Pr(T > c|X(c)) = \sum_{k=1}^{2} \sum_{l=1}^{2} Pr(T > c|X(c), v_D^k, v_P^l) p_{kl}. \tag{6}$$

3.2. The Likelihood Function

We derive the likelihood function as follows:

$$
\begin{aligned}
\ell \quad & (\beta, \alpha, \theta, v, p, \gamma) \\
\propto & \prod_{i \in W_{11}} \left[\sum_{k=1}^{2} \sum_{l=1}^{2} (h_D(t_i; x_i(t_i), v_D^k) \exp\left(-\int_0^{t_i} h_D(u; x_i(u), v_D^k) du\right) \exp\left(-\int_0^{t_i} h_P(u; x_i(u), v_P^l, v_D^k) du\right) p_{kl} \right] \\
\times & \prod_{i \in W_{12}} \left[\sum_{k=1}^{2} \sum_{l=1}^{2} (h_P(u; x_i(u), v_P^l, v_D^k) \exp\left(-\int_0^{t_i} h_D(u; x_i(u), v_D^k) du\right) \exp\left(-\int_0^{t_i} h_P(u; x_i(u), v_P^l, v_D^k) du\right) p_{kl} \right] \\
\times & \prod_{i \in W_2} \left[\sum_{k=1}^{2} \sum_{l=1}^{2} \exp\left(-\int_0^{c_i} h_D(u; x_i(u), v_D^k) du\right) \exp\left(-\int_0^{c_i} h_P(u; x_i(u), v_P^l, v_P^l) du\right) p_{kl} \right]
\end{aligned} \tag{7}
$$

where W_{11} is the set of 84,012 uncensored loans that are defaulted (D), W_{12} is the set of 8974 loans that are prepaid (P) and W_2 is the set of 14,641 right-censored loans that are still in foreclosure proceedings as of March 2016.

There are three important points that should be noted about the likelihood function: (1) In order to avoid identification problems, we assume no constant covariates; that is, no intercept in the proportionality term. The levels of the intensities are captured by means of the values $v_D^1, v_D^2, v_P^1, v_P^2$, which are left unconstrained; (2) The likelihood function is valid when the covariates are continuously observed since the foreclosure proceedings is initiated. This condition is automatically satisfied by covariates x_i, which depend on individuals only. However, the covariates that depend on time are usually observed in discrete time. In this case, the likelihood function has to be approximated by assuming that the covariates are constant between two consecutive observation dates; (3) There is no closed-form expression for the integration of the prepayment hazard function. Thus, the integral is evaluated using the trapezoidal rule.[18]

4. Empirical Analysis

Here, we report and discuss the maximum likelihood estimates of the general model and its associated nested models. The general model, Model (1), is the unrestricted model introduced in the previous section. Model (2) is the model in which there is no structural dependence between the default and prepayment hazards, i.e., $\gamma = 0$. Model (3) is the model in which the unobserved heterogeneity terms v_D and v_P are assumed independent. This independence assumption is equivalent to the condition $p_{11}p_{22} - p_{12}p_{21} = 0$, whenever $v_D^1 \neq v_D^2$ and $v_P^1 \neq v_P^2$. Under Model (3), the two competing risks are independent conditional on the observed covariates. Finally, Model (4) is the model without unobserved heterogeneity terms. Tables 3 and 4 provide estimation results for Model (1), Model (2), Model (3) and Model (4), respectively.[19] The intercepts are set equal to zero in all models with unobserved heterogeneity terms (that is Models (1), (2) and (3)) since the intercepts cannot be distinguished from multiplicative constants in unobserved heterogeneities.

Based on the likelihood ratio tests, all the restricted models are rejected in favor of Model (1) (see Appendix A for details on comparing Models (1), (2), (3) and (4) based on the likelihood ratio tests). Thus, it can be concluded that unobserved dependent heterogeneities, as well as the structural dependence of the prepayment hazard rate on the default probability exist. In particular, the results confirm that neglecting the structural dependence can lead to overestimation of the correlation of the

---

18   The mathematical details of the estimation of the likelihood function will be provided upon request.

19   The numbers in parentheses are the standard errors for the estimated coefficients. *, ** and *** indicate that the coefficients are statistically significant at the 10%, 5% and 1% levels, respectively. The standard errors reported for the $p = (p_{11}, p_{12}, p_{21}, p_{22})$ and ρ are estimated using the delta method. The mathematical details of the estimation will be provided upon request.

unobserved heterogeneity terms. This can be seen by comparing Model (1) to Model (2), where the magnitude of the correlation parameter, ρ, decreased, but remained significant, and the coefficient of the structural dependence, γ, is positive and statistically significant.

Table 3. Dependent competing risks' estimates: Models (1) and (2).

	Model (1)				Model (2)			
	Default		Prepayment		Default		Prepayment	
LTV	-0.205 (0.108)	*	-3.223 (0.182)	***	-0.183 (0.122)		-3.373 (0.274)	***
FICO Score	-0.181 (0.046)	***	0.299 (0.082)	***	-0.194 (0.043)	***	0.131 (0.086)	
Unemployment Rate	1.038 (0.116)	***	0.369 (0.162)	**	1.024 (0.145)	***	0.818 (0.212)	***
Delinquency Spell	-0.753 (0.063)	***	-0.671 (0.187)	***	-0.755 (0.071)	***	-1.28 (0.136)	***
Judicial States	-0.859 (0.030)	***	-0.430 (0.030)	***	-0.86 (0.025)	***	-0.43 (0.025)	***
Interest Rate Spread	0.366 (0.167)	**	-0.530 (0.211)	**	0.307 (0.080)	***	-0.424 (0.071)	***
α	2.956 (0.059)	***	1.109 (0.098)	***	2.974 (0.058)	***	2.617 (0.179)	***
θ	-7.595 (0.523)	***	-1.139 (0.081)	***	-8.349 (0.029)	***	1.288 (0.029)	***
v_D^1	6.553 (0.137)	***			6.709 (0.063)	***		
v_D^2	0.911 (0.146)	***			1.069 (0.057)	***		
v_P^1	0.524 (0.172)	***			5.899 (0.228)	***		
v_P^2	-0.404 (0.067)	***			0.665 (0.257)	***		
P_{11}	0.353 (0.013)	***			0.427 (0.014)	***		
P_{12}	0.216 (0.010)	***			0.046 (0.012)	***		
P_{21}	0.000 (0.000)	***			0.002 (1.91×10^{-4})	***		
P_{22}	0.430 (0.015)	***			0.525 (0.171)	***		
ρ	0.624 (0.013)	***			0.907 (0.017)	***		
γ	0.421 (0.055)	***						
Log-Likelihood	$-107,855.160$				$-108,080.63$			

The table provides the maximum likelihood estimates for Model (1) and Model (2). Model (1) is the model with dependent unobserved heterogeneities and structural dependence, and Model (2) is the model with dependent unobserved heterogeneities. The numbers in parentheses are the standard errors for the estimated coefficients. *, ** and *** indicate that the coefficients are statistically significant at the 10%, 5% and 1% levels, respectively.

Table 4. Independent competing risks' estimates: Models (3) and (4).

	Model (3)				Model (4)			
	Default		**Prepayment**		**Default**		**Prepayment**	
Constant					0.373 (0.330)		−1.053 (0.788)	
LTV	−0.236 (0.188)		−2.444 (0.341)	***	−0.542 (0.145)	***	−3.663 (0.272)	***
FICO Score	−0.196 (0.050)	***	0.261 (0.170)		−0.153 (0.049)	***	0.112 (0.091)	
Unemployment Rate	1.050 (0.234)	***	0.481 (0.671)		0.763 (0.179)	***	1.041 (0.473)	**
Delinquency Spell	−0.696 (0.203)	***	−1.491 (0.656)	**	−1.331 (0.159)	***	−1.749 (0.229)	***
Judicial States	−0.858 (0.030)	***	−0.433 (0.030)	***	−0.87 (0.023)	***	−0.42 (0.023)	***
Interest Rate Spread	0.321 (0.194)	*	−0.599 (0.365)		0.525 (0.149)	***	−0.256 (0.304)	
α	2.977 (0.076)	***	2.681 (0.277)	***	1.231 (0.031)	***	0.925 (0.214)	***
θ	−8.400 (0.800)	***	1.218 (0.275)	***	−1.829 (0.113)	***	3.328 (0.018)	***
v_D^1	6.749 (0.209)							
v_D^2	1.073 (0.190)							
v_P^1	6.756 (0.493)							
v_P^2	−0.283 (0.343)							
P_{11}	0.039 (0.008)							
P_{12}	0.511 (0.018)							
P_{21}	0.032 (0.006)							
P_{22}	0.418 (0.018)							
Log−Likelihood	−109,291.210				−192,769.690			

The table provides the maximum likelihood estimates for Model (3) and Model (4). Model (3) is the model with independent unobserved heterogeneities, and Model (4) is the model with no unobserved heterogeneities. The numbers in parentheses are the standard errors for the estimated coefficients. *, ** and *** indicate that the coefficients are statistically significant at the 10%, 5% and 1% levels, respectively.

In the following, we focus on the results of Model (1) to analyze the effects of covariates. The results in Table 3 show that the higher the equity in the property (evidenced by LTV), the higher the probabilities of default and prepayment. These results suggest that lenders seek to own properties that have more equity to lower their loss rate, and borrowers like to sell the properties that have more equity to lower their mortgage debt. In terms of credit scores, the results indicate that loans with higher FICO scores are less likely to default and more likely to prepay. Unemployment rate is used as a proxy for financial instability and suggests that a higher unemployment rate increases the

probabilities of default and prepayment. The share of months in which the loan was delinquent prior to a foreclosure proceedings affects the likelihood of default and prepayment. In particular, loans with long delinquency spells are less likely to default and to prepay. Judicial states have lower probabilities of default and prepayment than nonjudicial states, suggesting that the foreclosure process lasts longer for states in which foreclosure is processed through the state court system. In terms of interest rates, an increase in the interest rate spread increases the probability of default and decreases the probability of prepayment.

Table 3 also lists the estimated parameters of expo-power distribution. Using these estimates from Model (1), Figure 2 presents the baseline hazards for default and prepayment. As shown in Figure 2, the baseline hazard for default appears to be inverted U-shaped. That is, the likelihood of default increases in the first months, reaches a peak and then decreases. The baseline hazard for prepayment features an initial increase followed by a gradual decrease. Note that at all time points, the baseline hazard for prepayment is higher than the baseline hazard for default. This means that, in the absence of covariates, the chance of prepayment for loans in foreclosure proceedings is higher than the chance of default.
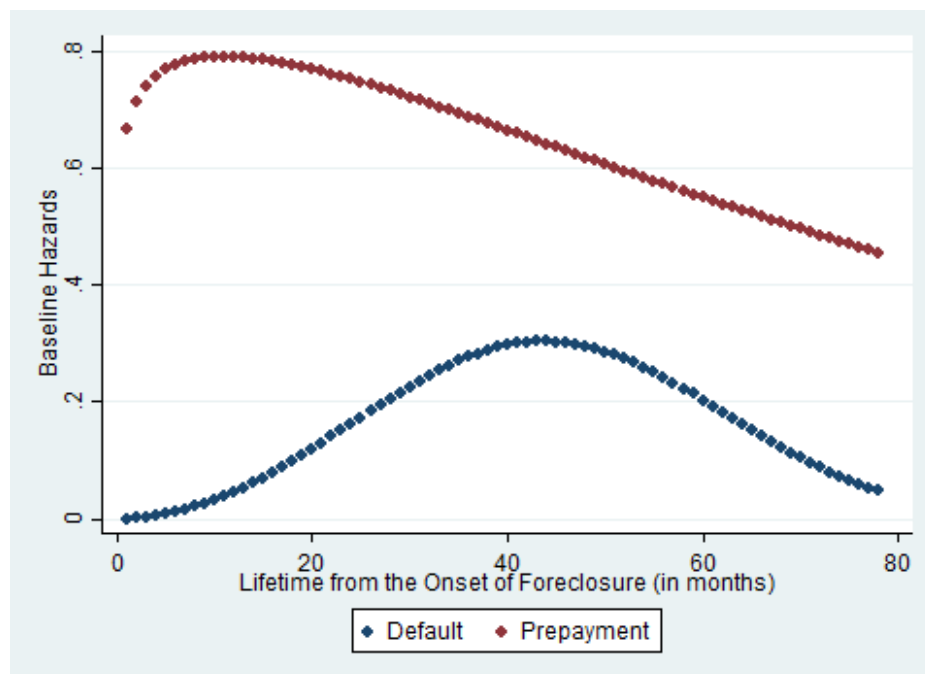

Figure 2. Baseline hazards for default and prepayment. The figure displays the estimates of the baseline hazards for default and prepayment. The estimate of the baseline hazards for event type j ($j = D, P$) is obtained using the maximum likelihood estimates of α_j and θ_j ($j = D, P$) from Model (1) and the lifetimes of loans from the onset of foreclosure.

$\rho(v_D, v_B)$ and γ in Table 3 denote interdependence between the default and prepayment hazards through the correlation of the unobserved heterogeneity terms and through the preventive behavior of individual mortgage borrowers, respectively. The positive and significant sign of the estimated correlation between the unobserved heterogeneity terms suggests that there are some unobservable loan-specific characteristics that affect both default and prepayment hazards in the same direction. The positive and significant sign of γ implies that the higher risk of default leads to a higher probability of prepayment. The result supports the hypothesis of structural dependence induced by the preventive behavior of individual mortgage borrowers.

5. Conclusions

Using a panel data of FHA mortgage loans, we specify a dependent competing risks framework to examine the determinants of the default and prepayment hazards once the foreclosure proceedings is initiated. More importantly, we examine the interdependence between default and prepayment, through both the correlation of the unobserved heterogeneity terms and the preventive behavior of the individual mortgage borrowers. We incorporate interdependence between the default and prepayment hazards through both the correlation of the unobserved heterogeneity terms associated with each risk and the preventive behavior of individual mortgage borrowers.

Our most important empirical finding here is that default and prepayment hazards are interdependent in two distinct ways. First, we find a significant positive correlation between the unobserved heterogeneity terms. This finding suggests that there are some unobservable loan-specific characteristics that affect both default and prepayment hazards in the same direction. Second, we find a significant positive structural dependence, suggesting that higher risk of default leads to higher probability of prepayment. We show that neglecting the interdependence through the correlation of the unobserved heterogeneity terms and through the preventive behavior of the individual mortgage borrowers can lead to biased estimates and misleading inference. As for the effects of covariates, we find that equity, FICO score, unemployment rate, delinquency spells, judicial states and interest rate spread are affecting the default and prepayment hazards.

Author Contributions: All authors have contributed in this paper. All authors have read and approved the final paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Comparing Models (1), (2), (3) and (4) Based on the Likelihood Ratio Tests

By comparing Models (3) and (4), we can assess whether unobserved independent heterogeneity terms exist in the default and prepayment hazards. The test statistics for the presence of v_D and v_P are independent under the null since the likelihood function can be factorized into the product of likelihood functions for default and prepayment. However, the likelihood ratio test for the null hypothesis $H_{0D} = \{v_D^1 = v_D^2\}$ and for the null hypothesis $H_{0P} = \{v_P^1 = v_P^2\}$ are nonstandard, since fewer parameters are identified under the null hypothesis than under the alternative. For instance, the probabilities p_{11}, p_{12}, p_{21} and p_{21} are not identifiable if $v_D^1 = v_D^2$ and $v_P^1 = v_P^2$. A careful analysis of this problem is out of the scope of our analysis and would require either assumptions on the local alternatives of interest or some prior restrictions on the parameter domain to avoid difficulties (e.g., Andrews and Ploberger 1994). It has been widely assumed in the literature that the critical value of the chi-square distribution with two degrees of freedom is a conservative test to use. Thus, in our analysis, we compare the likelihood ratio test with the critical value of the χ_2^2 distribution. For Model (3), the log-likelihood values for default and prepayment are $-61{,}931.37$ and $-28{,}668.87$, respectively; and for Model (4), the log-likelihood values for default and prepayment are $-135{,}705.99$ and $-35{,}007.83$, respectively. The calculated values of the likelihood ratio test are larger than the critical value of χ_2^2 at the five percent level. Thus, a significant improvement of Model (3) over Model (4) is concluded.

By comparing Models (2) and (3), we can test whether the unobserved heterogeneity terms are dependent. Testing for independence between v_D and v_P is equivalent to testing the null hypothesis $H_0 = \{p_{11}p_{22} - p_{12}p_{21} = 0\} = \{\rho = 0\}$. Under the null hypothesis, the likelihood ratio test is distributed as a chi-square with one degree of freedom. The calculated value of the likelihood ratio test is larger than the critical value of χ_1^2 at the five percent level. Hence, we conclude that the unobserved heterogeneity terms are dependent, and Model (2) is improved significantly over Model (3). Equivalently, we can consider the significance of the maximum likelihood estimate of correlation ρ. We reject the null hypothesis that the estimate of the correlation parameter, ρ, equals zero.

By comparing Models (1) and (2), we can test whether the structural dependence of the prepayment hazard rate on the default probability exists or not. The calculated value of the likelihood

ratio test is larger than the critical value of χ_1^2 at the five percent level, which supports the existence of interdependence between the default and prepayment hazards through both the correlation of the unobserved heterogeneities and the preventive behavior of individual mortgage borrowers.

References

Abbring, Jaap H., and Gerard J. Van den Berg. 2003. The Identifiability of the Mixed Proportional Hazards Competing Risks Model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65: 701–10.

Andrews, Donald W. K., and Werner Ploberger. 1994. Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative. *Econometrica* 62: 1383–414.

Clayton, David, and Jack Cuzick. 1985. Multivariate Generalization of the Proportional Hazards Model. *Journal of Royal Statistical Society Series A* 148: 82–117.

Colby, Gordana, and Paul Rilstone. 2004. Nonparametric Identification of Latent Competing Risks Models. *Econometric Theory* 20: 883–90.

Cunningham, Donald F., and Patric Hendershott. 1984. Pricing FHA Mortgage Default Insurance. *Housing Finance Review* 13: 373–92.

Danis, Michelle A., and Anthony Pennington-Cross. 2005. *A Dynamic Look at Subprime Loan Performance*; Working paper; St. Louis: Federal Reserve Bank of St. Louis.

Deng, Yongheng. 1997. Mortgage Termination: An Empirical Hazard Model with Stochastic Term Structure. *The Journal of Real Estate Finance and Economics* 14: 309–31.

Deng, Yongheng, John M. Quigley, Robert Van Order, and Freddie Mac. 1996. Mortgage Default and Low Down-payment Loans: The Cost of Public Subsidy. *Regional Science and Urban Economics* 26: 263–85.

Deng, Yongheng, John M. Quigley, and Robert Van Order. 2000. Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options. *Econometrica* 68: 275–307.

Flinn, Christopher, and James Heckman. 1982. New Methods for Analyzing Structural Models of Labor Force Dynamics. *Journal of Econometrics* 18: 115–68.

Foster, Chester, and Robert Van Order. 1985. FHA Terminations: A Prelude to Rational Mortgage Pricing. *Journal of the American Real Estate and Urban Economics Association* 13: 273–91.

Green, Jerry R., and John B. Shoven. 1986. The Effect of Interest Rates on Mortgage Prepayments. *Journal of Money, Credit and Banking* 18: 41–50.

Gourieroux, Christian, and Joann Jasiak. 2004. Heterogeneous INAR(1) Model with Application to Car Insurance. *Insurance: Mathematics and Economics* 34: 177–92.

Heckman, James J., and Bo E. Honore. 1989. The Identifiability of the Competing Risks Model. *Biometrika* 76: 325–30.

Heckman, James, and Burton Singer. 1984. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica* 52: 271–320.

Heckman, James J., and James R. Walker. 1990. Estimating Fecundability from Data on Waiting Times to First Conception. *Journal of American Statistical Association* 85: 283–94.

Kau, James B., and Donald C. Keenan, 1996. *Patterns of Rational Default*; Working paper; Athens: The University of Georgia.

Nickell, Stephen. 1979. Estimating the Probability of Leaving Unemployment. *Econometrica* 47: 1249–66.

Pennington-Cross, A. 2006. *The Duration of Foreclosure on the Subprime Mortgage Market: A Competing Risks Model with Mixing*; Working paper; St. Louis: Federal Reserve Bank of St. Louis.

Quigley, John M., and Robert Van Order. 1990. Efficiency in the Mortgage Market: The Borrower's Perspective. *Journal of the American Real Estate and Urban Association* 18: 237–52.

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

Rosholm, Michael, and Michael Svarer. 2001. Structurally Dependent Competing Risks. *Economic Letters* 73: 169–73.

Saha, Atanu, and Lynette Hilton. 1997. Expo-power: A Flexible Hazard Function for Duration Data Models. *Economics Letters* 54: 227–33.

Schwartz, Eduardo S., and Walter N. Torous. 1989. Prepayment and the Valuation of Mortgage-Backed Securities. *The Journal of Finance* 44: 375–92.

Schwartz, Eduardo S., and Walter N. Torous. 1993. Mortgage Prepayment and Default Decisions: A Poisson Reegression Approach. *Journal of the American Real Estate and Urban Economics Association* 21: 431–49.

Titman, Sheridan, and Walter Torous. 1989. Valuing Commercial Mortages: An Empirical Investigation of the Contingent Claims Approach to Pricing Risky Debt. *The Journal of Finance* 44: 345–73.

Tsiatis, Anastasios. 1975. A Nonidentifiability Aspect of the Problem of Competing Risks. *Proceedings of the National Academy of Sciences* 72: 20–22.

Van den Berg, Gerard J., Marten Lindeboom, and Geert Ridder. 1994. Attrition in Longitudinal Panel Data and the Empirical Analysis of Dynamic Labour Market Behaviour. *Journal of Applied Econometrics* 9: 421–35.

Van den Berg, Gerard J., Bas Van der Klaauw, and Jan C. Van Ours. 2004. Punitive Sanctions and the Transition Rate from Welfare to Work. *Journal of Labor Economics* 22: 211–41.