A Transformed Mixed Multinomial Logit Model with an Application to Residential Mortgage Terminations

Ran An

Ph.D., Economics, Syracuse University

Wenzhen Lin

Department of Economics

Syracuse University

wlin18@syr.edu

Jan Ondrich

Center for Policy Research

and

Department of Economics

Syracuse University

jondrich@maxwell.syr.edu

Abstract

In this paper we apply the Ding-Tian-Yu-Guo transformation to multinomial logit duration models with and without non-parametric random effect controls for over-dispersion in the data due to unobserved heterogeneity. Ding, Tian, Yu, and Guo (2012) analyze transformations of the binomial logit duration model for which the results are an exact binomial logit duration model for one value of a shape parameter and an interval-censored proportional hazard model for the limiting value of the shape parameter. We analyze the simultaneous mortgage-termination risks of 90-day delinquency and prepayment for single-family 30 year fixed-rate mortgages securitized by Fannie Mae using the Fannie Mae Public Use Data. We show that the transformation can control for over-dispersion in the data and performs better in both cases than the corresponding models without the transformation.
Keywords: transformation, over-dispersion, mortgage termination, 90-day delinquency, prepayment

1 Introduction

Ding, Tian, Yu, and Guo (2012) introduce a discrete-time transformation family of the binary logit duration model and apply it to bankruptcy probability prediction with time-varying covariates. The transformation model family contains the Shumway (2001) model and Cox proportional hazards model (Cox 1972). In this paper, we construct a way to apply the Ding-Tian-Yu-Guo transformation to multinomial logit duration models, which allow for termination into multiple states, and add controls for unobserved heterogeneity. Our empirical application uses the Fannie Mae Public Use Data to analyze the simultaneous mortgage-termination risks of 90-day delinquency and prepayment for single-family 30 year fixed rate mortgages. Both in sample and out-of-sample validation statistics show the models with the Ding-Tian-Yu-Guo transformation improves the performance over the correspond-

ing models without the transformation.

As of the first quarter of 2017, there was about $14.4 trillion in total outstanding US mortgage debt, more than half of which was held in mortgage-backed securities (MBS's), mortgage bundles sold to investment banks or other investors. MBS's allow millions of Americans to own homes by effectively connecting the needs of investors and borrowers. However, the borrowers make individual decisions based on current circumstances that cannot be accurately predicted by MBS investors. Each month borrowers decide whether to prepay, go into delinquency, or remain current. Prepayment occurs when the borrower pays off the loan before the maturity date. It can be viewed as the exercise of a financial option to buy the mortgage (call option). Delinquency occurs when the borrower fails to make a payment; 90-day delinquency will always be the first step in a default. Default can be viewed as the exercise of a financial option to sell the mortgage (put option).

Kau, Donald, Walter, and James (1992) use option-pricing theory to rationally price mortgages for both default and prepayment risks, and show that the decisions to prepay or default are substitutes. Default behavior for residential mortgages has been studied by, among others, Ambrose, Capone, and Deng (2001), Lacour-Little and Malpezzi (2003), Agarwal, Deng, and He (2014), and An, Deng, and Gabriel (2019). Follain, Ondrich, and Sinha (1997) and Archer, Ling, and McGill (2002) studied residential prepayment behavior.

Lancsater (1979), in a study of unemployment duration distribution, shows that parameter estimates may be inconsistent when there are omitted variables, even when the omitted variables are orthogonal to the included variables. He proposes a mixed duration model that incorporates a parameteric random effect to control for the unobserved heterogeneity. Heckman and Singer (1984) develop a method for estimating a mixed model with a non-parametric random effect to overcome the "over-parametrization" inherent in parametric forms. An alternative method, suggested by Trussell and Richards (1985), is to keep increasing the number of mass

points until the likelihood improvement becomes insignificant. Deng, Quigley, and van Order (2000) introduce mass-point corrections for unobserved heterogeneity in a competing risks proportional hazard model for prepayment and default. Clapp, Deng, and An (2006) compare such corrections for unobserved heterogeneity across proportional hazard and multinomial logit models. They find that the proportional hazard versions of their models outperform the multinomial logit versions both in sample and out of sample. Other studies that use a competing risks framework include Calhoun and Deng (2002) and Pennington-Cross (2003).

In this paper, we estimate models that predict prepayment and 90-day delinquency rates given economic scenario, loan information, and borrower characteristics, such as FICO score, the number of units, original loan-to-value ratio, original debt-to-income ratio, the loan size, and the interest rate.

The remainder of this paper is organized as follows. Section 2 provides the summary of the data and definitions of explanatory variables. Section 3 discusses the models with and without non-parametric random effect controls for over-dispersion, and for each of these, with and without the Ding-Tian-Yu-Guo transformation . Section 4 presents the empirical results. Section 5 presents the conclusions.

2 Data

The data used in this paper is the public use Fannie Mae single-family loan performance data for 30-year fixed rate loans. Our data are randomly selected from 466,700 mortgages originated from January 1st, 2000 to December 31, 2018, for Miami. We chose Miami because it had high prepayment and default rates simultaneously. We randomly choose 2.4 percent of the loans from our dataset. The exact algorithm is as follows: first, using a uniform distribution, we randomly assigned a number between 0 and 1 to each loan, then, we chose the loans whose number is between 0.558 and 0.582.

3 Model

Our application uses annual data on mortgages. In any year the mortgage can be one of three possible states: A^p, A^d, and C. A^p represents the act of prepayment, A^d represents the act of going into 90-day delinquency, and C represents the act of remaining current.

We sometimes use a state $A = A^p \cup A^d$, which means that either prepayment or 90-day delinquency happens in that year.

3.1 Multinomial Logit Model

The binary and multinomial logit are among the most widely used discrete choice models. These models imply proportional substitution across alternatives. The independence of irrelevant alternative (IIA) property of the multinomial logit model means that for any two alternatives i and j, the ratio of the probabilities does not depend on any alternatives other than i and j. Chipman (1960) and Debreu (1960) point out that the IIA property is clearly inappropriate in many choice situations.

The probability functions for a multinomial logit model are

$$P(A^p_{mt}) = \pi_{pmt} = \frac{exp(z'_{pmt}\beta)}{1+exp(z'_{pmt}\beta)+exp(z'_{dmt}\beta)} \quad ,$$
$$P(A^d_{mt}) = \pi_{dmt} = \frac{exp(z'_{dmt}\beta)}{1+exp(z'_{pmt}\beta)+exp(z'_{dmt}\beta)} \quad , \text{ and}$$
$$P(C_{mt}) = \pi_{cmt} = \frac{1}{1+exp(z'_{pmt}\beta)+exp(z'_{dmt}\beta)}$$

The likelihood function for the multinomial logit has the form:

$$l(\beta) = \prod_{m=1}^{M}\prod_{t=1}^{T_m} \pi_{pmt}^{\delta^p_{mt}}\pi_{dmt}^{\delta^d_{mt}}\pi_{cmt}^{\delta^c_{mt}}$$
$$=\prod_{m=1}^{M}\prod_{t=1}^{T_m}\left(\frac{exp(z'_{pmt}\beta)}{1+exp(z'_{pmt}\beta)+exp(z'_{dmt}\beta)}\right)^{\delta^p_{mt}}\left(\frac{exp(z'_{dmt}\beta)}{1+exp(z'_{pmt}\beta)+exp(z'_{dmt}\beta)}\right)^{\delta^d_{mt}}\left(\frac{1}{1+exp(z'_{pmt}\beta)+exp(z'_{dmt}\beta)}\right)^{\delta^c_{mt}}$$

where m is the indicator of the mortgage, t is the age of the mortgage, T_m is the age of mortgage m at termination, and the π's are the probabilities, and the δ_{mt}'s are the outcome indicator variables.

3.2 The Transformed Mutinomial Logit Model

Ding, Tian, Yu, and Guo (2012) analyze the risk of bankruptcy by constructing transformed binary logit models. In this paper, we use one of their two transformation functions. This transformation function gives as the probability of bankruptcy in a given year

$$\pi_{mt} = \begin{cases} 1 - \frac{1}{\left(1+\text{cexp}\left(z'_{bmt}\beta\right)\right)^{1/c}} & , \quad c > 0 \\ 1 - \exp\left(-\exp\left(z'_{bmt}\beta\right)\right) & , \quad c = 0 \end{cases} .$$

The probability of continuing another year without bankruptcy is $1 - \pi_{mt}$.

The effect of c in the present case is similar to the effect of transformation parameters in the Box-Cox model (see Box and Cox 1962). In the present case, when c limits to 0 from the right, the discrete time survival model is the interval-censored proportional hazard model. When c limits to 1, the discrete time survival model is a binary logit.

We incorporate the Ding-Tian-Yu-Guo transformation into the multinomial logit model in section 3.2.1 and present full information maximum likelihood estimation. In section 3.2.2, we present a two-step method that maybe useful in finding the starting values for the parameters.

3.2.1 Full Information Maximum Likelihood

In the transformed multinomial logit model, the probabilities for each year are

$$\text{P}(A^p_{mt}|A_{mt}) = \frac{exp(z'_{pmt}\beta)}{exp(z'_{pmt}\beta)+exp(z'_{dmt}\beta)} \quad ,$$
$$\text{P}(A^d_{mt}|A_{mt}) = \frac{exp(z'_{dmt}\beta)}{exp(z'_{pmt}\beta)+exp(z'_{dmt}\beta)} \quad ,$$

$$P(A_{mt}) = 1\text{-}\frac{1}{(1+cexp(I^A_{mt}))^{\frac{1}{c}}} \quad \text{, and}$$
$$P(C_{mt}) = \frac{1}{(1+cexp(I^A_{mt}))^{\frac{1}{c}}} \quad ,$$

where I^A_{mt} is called inclusive value for event A at age t of mortgage m. Its formular is given by:

$$I^A_{mt} = log(exp(z'_{pmt}\beta) + exp(z'_{dmt}\beta)) \quad .$$

The marginal probabilities of A^p_{mt} and A^d_{mt} are given by

$$P(A^p_{mt}) = P(A^p_{mt}|A_{mt})P(A_{mt}) = \frac{exp(z'_{pmt}\beta)}{\exp(I^A_{mt})}(1 - \frac{1}{(1+cexp(I^A_{mt}))^{\frac{1}{c}}}) \quad \text{, and}$$
$$P(A^d_{mt}) = \frac{exp(z'_{dmt}\beta)}{\exp(I^A_{mt})}(1 - \frac{1}{(1+cexp(I^A_{mt}))^{\frac{1}{c}}}) \quad .$$

Therefore the likelihood function is

$$\begin{aligned} l(\beta, c) = \prod_{m=1}^{M} \prod_{T=1}^{T_m} & (\frac{exp(z'_{pmt}\beta)}{\exp(I^A_{mt})}(1 - \frac{1}{(1+c\exp(I^A_{mt}))^{\frac{1}{c}}}))^{\delta^p_{mt}} \\ & (\frac{exp(z'_{dmt}\beta)}{\exp(I^A_{mt})}(1 - \frac{1}{(1+c\exp(I^A_{mt}))^{\frac{1}{c}}}))^{\delta^d_{mt}} \\ & (\frac{1}{(1+cexp(I^A_{mt}))^{\frac{1}{c}}})^{1-\delta^p_{mt}-\delta^d_{mt}} \quad . \end{aligned}$$

3.2.2 A Two-Step Estimation Method

Amemiya (1978) shows that the two-step estimation of a multivariate logit model can be considerably simpler than full information maximum likelihood estimation, especially for the model with many dependent variables. Moreover, the two-step estimation model may be more helpful in finding the starting values.

In the first step, we use only the year observations for which there is prepayment or 90-day delinquency. We run a binary logit for P($A^p|A$) versus P($A^d|A$). The first stage estimating P($A^p|A$) and P($A^d|A$) is a binary logit, so we have

$$\mathrm{P}(A^d|A) = \frac{exp(z'_{mt}\beta_1)}{1+exp(z'_{mt}\beta_1)}$$
$$\mathrm{P}(A^p|A) = \frac{1}{1+exp(z'_{mt}\beta_1)}$$

where z_{mt} contains all the variables, β_1[1] are the estimated coefficient in the step 1, and the likelihood function can be written as

[1] $\beta_1 = \beta_d - \beta_p$, because we set prepayment as the reference group.

$$l(\beta) = \prod_{m=1}^{M} \prod_{T=1}^{T_m} \left(\frac{exp(z'_{mt}\beta_1)}{1+exp(z'_{mt}\beta_1)}\right)^{\delta^d_{mt}} \left(\frac{1}{1+exp(z'_{mt}\beta_1)}\right)^{\delta^p_{mt}} \quad .$$

In the second step, we bring in the result from the first stage and calculate the marginal probabilities of A and C.[2]

$$\begin{aligned} \mathrm{P}(A) &= 1 - \frac{1}{(1+cexp(I^A))^{\frac{1}{c}}} \\ \mathrm{P}(C) &= \frac{1}{(1+cexp(I^A))^{\frac{1}{c}}} \end{aligned}$$

where I^A is the inclusive value $log(exp(z'_{pmt}\beta) + exp(z'_{dmt}\beta))$, and the likelihood can be written as

$$l(\beta, c) = \prod_{m=1}^{M} \prod_{T=1}^{T_m} \left(1 - \frac{1}{(1+cexp(I^A))^{\frac{1}{c}}}\right)^{\delta^A_{mt}} \left(\frac{1}{(1+cexp(I^A))^{\frac{1}{c}}}\right)^{\delta^C_{mt}} \quad .$$

3.3 Models with Unobserved Heterogeneity

Heckman and Singer (1984) suggest using finite mixture random effects in a continuous time duration model, Follmann and Lambert (1989) discuss generalizing panel binary logistic regression using non-parametric mixing. In the present case, we use non-parametric mixing for panel multinomial logit regression across mortgage age. In Follmann and Lambert, the number of trials is fixed, whereas in our case the number of trials is determined by the number of sample years for each mortgage. Follman and Lambert argue that over-dispersion relative to the binomial distribution is possible if the trials are positively correlated, perhaps because an important covariate is omitted.

[2]In the second step, we estimate the coefficients for the common variables: FICO, loan size, multiple units dummy, DTI, and age dummies for prepayment. The coefficient for delinquency for those common variables equal the coefficient (β_1) from step 1 plus the coefficient for prepayment. For the coefficients of the variables which exist in only one of the delinquency or prepayment risks, we use the corresponding coefficient or negative of the corresponding coefficient from β_1 in step 1.

3.3.1 Multinomial Logit Model with Unobserved Heterogeneity

The probability functions for mass point j of the P distinct mass points in the finite mixture distribution of random effects are:

$$\pi_{pmt,j} = \frac{\eta_{p,j} exp(z_{pmt}^{'}\beta)}{1+\eta_{p,j} exp(z_{pmt}^{'}\beta)+\eta_{d,j} exp(z_{dmt}^{'}\beta)}$$
$$\pi_{dmt,j} = \frac{\eta_{d,j} exp(z_{dmt}^{'}\beta)}{1+\eta_{p,j} exp(z_{pmt}^{'}\beta)+\eta_{d,j} exp(z_{dmt}^{'}\beta)}$$
$$\pi_{cmt,j} = \frac{1}{1+\eta_{p,j} exp(z_{pmt}^{'}\beta)+\eta_{d,j} exp(z_{dmt}^{'}\beta)}$$

where $\eta_{p,j}$ is the scale parameter associated with the risk of prepayment for the j^{th} mass point, and $\eta_{d,j}$ is the scale parameter associated with the risk of 90-days delinquency for the j^{th} mass point.

Then, the log-likelihood function for the mixed multinomial logit model is:

$$\begin{aligned} L(\beta, p, \eta) &= \sum_{m=1}^{M} log(\sum_{j=1}^{P} p_j (\prod_{t=1}^{T_m} (\pi_{pmt,j})^{\delta_{mt}^{p}} (\pi_{dmt,j})^{\delta_{mt}^{d}} (\pi_{cmt,j})^{\delta_{mt}^{c}})) \\ &= \sum_{m=1}^{M} log(\sum_{j=1}^{P} p_j (\prod_{t=1}^{T_m} (\frac{\eta_{p,j} exp(z_{pmt}^{'}\beta)}{1+\eta_{p,j} exp(z_{pmt}^{'}\beta)+\eta_{d,j} exp(z_{dmt}^{'}\beta)})^{\delta_{mt}^{p}} \\ &\qquad (\frac{\eta_{d,j} exp(z_{dmt}^{'}\beta)}{1+\eta_{p,j} exp(z_{pmt}^{'}\beta)+\eta_{d,j} exp(z_{dmt}^{'}\beta)})^{\delta_{mt}^{d}} \\ &\qquad (\frac{1}{1+\eta_{p,j} exp(z_{pmt}^{'}\beta)+\eta_{d,j} exp(z_{dmt}^{'}\beta)})^{\delta_{mt}^{c}})) \end{aligned}$$

where p_j is the probability of the j^{th} mass point.

3.3.2 The Transformed Multinomial Logit Model with Unobserved Heterogeneity

The probability functions for mass point j are:

$$\pi_{pmt,j} = \frac{\eta_{p,j} exp(z_{pmt}^{'}\beta)}{\exp(I_{mt,j}^{A})} (1 - \frac{1}{(1+c exp(I_{mt,j}^{A}))^{\frac{1}{c}}})$$
$$\pi_{dmt,j} = \frac{\eta_{d,j} exp(z_{dmt}^{'}\beta)}{\exp(I_{mt,j}^{A})} (1 - \frac{1}{(1+c exp(I_{mt,j}^{A}))^{\frac{1}{c}}})$$
$$\pi_{cmt,j} = \frac{1}{(1+c exp(I_{mt,j}^{A}))^{\frac{1}{c}}}$$

where

$I_{mt,j}^{A} = log(\eta_{p,j} exp(z_{pmt}^{'}\beta) + \eta_{d,j} exp(z_{dmt}^{'}\beta))$.

The log-likelihood function has the form

$$L(\beta, p, \eta) = \sum_{m=1}^{M} log(\sum_{j=1}^{P} p_j(\prod_{t=1}^{T_m}(\pi_{pmt,j})^{\delta_{mt}^p}(\pi_{dmt,j})^{\delta_{mt}^d}(\pi_{cmt,j})^{\delta_{mt}^c}))$$

$$= \sum_{m=1}^{M} log(\sum_{j=1}^{P} p_j(\prod_{t=1}^{T_m}(\frac{\eta_{p,j}exp(z_{pmt}^{'}\beta)}{\exp(I_{mt,j}^A)}(1 - \frac{1}{(1+cexp(I_{mt,j}^A))^{\frac{1}{c}}}))^{\delta_{mt}^p}$$

$$(\frac{\eta_{d,j}exp(z_{dmt}^{'}\beta)}{\exp(I_{mt,j}^A)}(1 - \frac{1}{(1+cexp(I_{mt,j}^A))^{\frac{1}{c}}}))^{\delta_{mt}^d}$$

$$(\frac{1}{(1+cexp(I_{mt,j}^A))^{\frac{1}{c}}})^{\delta_{mt}^c})) \quad .$$

4 Empirical Results

To clearly assess the effect of Ding-Tian-Yu-Guo transformation, we present the transformed and untransformed empirical results in a single table. Table 1 compares the untransformed multinomial logit model (MNL) with the transformed multinomial logit model, estimated respectively without unobserved heterogeneity in models 1 and 2 and with unobserved heterogeneity in models 3 and 4. Trussell and Richards (1985) suggest adding mass points until there is no significant increase in the log-likelihood. Using this approach, we stop at three mass points. We present the results of four mass points in the Appendix B.

Table 1 shows that the likelihood increases from the untransformed model to the transformed model. Moreover, both forms of model are improved by incorporating unobserved heterogeneity, given the log likelihood increases so much.

Table 1: Multinomial Logit with and without Transformation

(Standard deviations are in parentheses.)

Estimate	Model 1	Model 2	Model 3	Model 4
Delinquency				
Baseline Intercept (Group 1)			2.246	4.283
			(0.224)	(1.076)
Baseline Intercept (Group 2)			0.177	-0.783
			(0.106)	(1.164)
Baseline Intercept (Group 3)			-1.902	1.283
			(0.209)	(1.128)
Recession Indicator	1.017	1.248	0.923	1.142
	(0.108)	(0.107)	(0.114)	(0.125)
Negative Equity	4.405	5.767	6.425	7.292
	(0.287)	(0.308)	(0.362)	(0.305)
Negative Equity Square	-2.022	-2.838	-2.635	-3.286
	(0.214)	(0.228)	(0.227)	(0.229)
Negative Equity * Recession	-0.361	-0.038	-0.751	-0.426
	(0.222)	(0.221)	(0.255)	(0.265)
FICO	-1.057	-1.329	-1.461	-1.713
	(0.057)	(0.061)	(0.046)	(0.088)
Log Loan Size	1.926	0.303	0.970	-0.389
	(1.731)	(0.837)	(0.207)	(2.161)
Dummy Units>1	0.124	0.168	0.101	0.033
	(0.230)	(0.030)	(0.053)	(0.194)
Debt-to-Income	1.052	1.238	1.404	1.485
	(0.277)	(0.200)	(0.128)	(0.310)
Unemployment Rate	0.211	0.561	0.547	0.747
	(0.144)	(0.158)	(0.064)	(0.199)
Original Loan-to-Value	2.468	3.081	3.127	3.542
	(0.240)	(0.139)	(0.074)	(0.295)
Prepayment				
Baseline Intercept (Group 1)			-1.735	-1.296
			(0.240)	(0.538)
Baseline Intercept (Group 2)			-5.096	-2.576
			(0.244)	(0.550)
Baseline Intercept (Group 3)			-2.765	-6.121
			(0.220)	(0.857)
Call Option	5.822	8.275	7.144	8.929
	(0.174)	(0.307)	(0.043)	(0.320)
FICO	-0.059	-0.176	-0.163	-0.273
	(0.024)	(0.028)	(0.022)	(0.040)
Log Loan Size	3.088	4.540	1.835	2.749
	(0.575)	(0.690)	(0.293)	(0.809)
Dummy Units>1	-0.416	-0.599	-0.575	-0.733
	(0.105)	(0.131)	(0.030)	(0.144)
Debt-to-Income	-0.663	-0.699	-0.619	-0.619
	(0.112)	(0.132)	(0.111)	(0.167)
Mass Point (Group 1)			0.345	2.367
			(0.106)	(0.258)
Mass Point (Group 2)			-1.426	1.790
			(0.113)	(0.269)
c		5.765		3.980
		(0.494)		(0.345)
Log Likelihood	-22181.48	-21974.54	-21947.47	-21922.38

Notes: The four models 1-4 are multinomial logit model, multinomial logit model with transformation, multinomial logit with unobserved heterogeneity, and multinomial logit model with transformation and unobserved heterogeneity.

The results for the coefficient estimates are uniform across models and consistent with the predictions of option theory. The probability of prepayment increases when the call option is positive (in the money); similarly higher negative equity increases the risk of 90-day delinquency. The results suggest that the behavior of borrowers may be affected by other factors. The estimates show that the 90-day delinquency risk is associated with a higher original loan-to-value ratio. The estimates also show that a higher unemployment rate will increase the 90-day delinquency risk. The debt-to-income ratio is positive in the 90-day delinquency risk and negative in the prepayment risk. This means that for the same monthly mortgage payment, a borrower with a higher income is more likely to prepay and less likely to become delinquent. Owning more than one unit negatively affects the prepayment risk. A higher FICO score negatively affects both termination risks. The FICO coefficient estimate in the 90-day delinquency risk is 10 times larger than the corresponding coefficient estimate in the prepayment risk.

4.1 Comparison of the Specifications

The model with both unobserved heterogeneity and the Ding-Tian-Yu-Guo transformation have the largest absolute coefficient values and lowest p-values for several important variables, among them the negative equity variables and the unemployment rate in the delinquency risk, the call option in the prepayment risk, and the FICO score, the debt-to-income ratio, and the loan-to-value ratio in both risks. Moreover, the model controlling for unobserved heterogeneity does better than the model without controls for unobserved heterogeneity and the transformed model does better than the untransformed model. We conjecture that this is because mixing and transforming the model are alternative and complementary ways to control for to over-dispersion. The model that is both mixed and transformed works best because the transformation controls for residual mixing not captured by the Trussell and Richards procedure. In Appendix A we show that for any multinomial outcome,

except for the special case, there exists a valuable transformation parameter that increases the variance of the outcome and therefore controls for over dispersion.

The multinomial logit specifications with and without unobserved heterogeneity both have a larger log likelihood with the Ding-Tian-Yu-Guo transformation than without it. McFadden, Train and Tye (1977), Hausman and McFadden (1984), and McFadden (1987) develop tests for the IIA property. A Wald test for c equals 1 in a transformed model also tests IIA. The estimated value of the transformation parameter c is 5.765 for the transformed multinomial logit without unobserved heterogeneity and 3.980 for the transformed multinomial logit with unobserved heterogeneity. The null hypothesis that c equals 1 is rejected in the both cases at the 1 percent level.

4.2 Out-of-Sample Predictive Accuracy

For the cross-sample validation we use a method similar to that of Clapp, Deng, and An (2006), but instead of holding back a 10 percent sub-sample for validation, we hold back a 50 percent sub-sample. For both sub-samples we use the estimation results to predict the delinquency and prepayment probabilities in the final year for each mortgage. For each sub-sample and each risk we then regress the indicators for delinquency and prepayment on their respective predicted probabilities and compare R-squares.

Table 1 presents the results. The models with the Ding-Tian-Yu-Guo transformation provide a better fit than corresponding untransformed models for both risks. Moreover, the models without unobserved heterogeneity are out-preformed by the corresponding unobserved heterogeneity specifications.

An alternative criterion is McFadden's Pseudo R-square, which compares an estimated unrestricted log likelihood from a specification with both covariates and loan age dummies to an estimated restricted log likelihood with loan age dummies

only:

$$\rho = 1 - LL_U/LL_R \quad ,$$

where LL_U is the unrestricted log likelihood and LL_R is the restricted log likelihood. The Pseudo R-squares do not change very much across sub-samples within models. The results are presented in Table 3 and are qualitative similar to the results in Table 2. For each of the four models in sample and out of sample Pseudo R-squares are similar with the preferred specification being the transformed multinomial logit with unobserved heterogeneity.

Table 2: Cross-Sample Validation for the Four Models

Models	In Sample	Out of Sample
Delinquency		
Multinomial Logit	0.272	0.267
Transformed Multinomial Logit	0.282	0.275
Multinomial Logit With Unobserved Heterogeneity	0.293	0.285
Transformed Multinomial Logit With Unobserved Heterogeneity	0.295	0.289
Prepayment		
Multinomial Logit	0.042	0.054
Transformed Multinomial Logit	0.055	0.068
Multinomial Logit With Unobserved Heterogeneity	0.104	0.116
Transformed Multinomial Logit With Unobserved Heterogeneity	0.120	0.133

Table 3: McFadden Pseudo R-square Results

Models	In Sample	Out of Sample
Multinomial Logit	0.0601	0.0579
Transformed Multinomial Logit	0.0688	0.0684
Multinomial Logit With Unobserved Heterogeneity	0.0700	0.0694
Transformed Multinomial Logit With Unobserved Heterogeneity	0.0710	0.0706

5 Conclusions

Ding, Tian, Yu, and Guo (2012) introduce transformed binary logit models that they apply to a univariate duration analysis of time to bankruptcy for banks. We

extend the Ding-Tian-Yu-Guo transformation to cover multinomial logit duration models which we apply to time to prepayment or delinquency for Fannie Mae mortgages. We show that a Wald test on the transformation parameter provides a test of the multinomial logit specification, like tests previously developed by McFadden, Train and Tye (1977), Hausman and McFadden (1984), and McFadden (1987).

An additional contribution of our work is exploring the relationship between the Ding-Tian-Yu-Guo transformation and over-dispersion in the data due to unobserved heterogeneity. We show that except for one special case, there is a value of the transformation parameter for which the outcomes are over-dispersed relative to the multinomial logit model. Follmann and Lambert (1989) address over-dispersion by generalizing binary logistic regression using non-parametric mixing for a balanced panel. In the present case, we use non-parametric mixing in a transformed multinomial logit duration model, which results in an unbalanced panel.

We use Fannie Mae public use data for single-family, 30-year fixed-rate mortgages to analyze 90-day delinquency and prepayment risks. Our empirical results generally support the hypothesis that prepayment and delinquency can be treated as financial options.

Our empirical results clearly reject the standard multinomial logit specifications. Including corrections for unobserved heterogeneity, as in Deng, Quigley and van Order (2000), Pennington-Cross (2003), and Clapp, Deng, and An (2006), improve both transformed and untransformed specifications. Similarly, transforming the model improves the specifications with and without controls for unobserved heterogeneity. We conjecture that the model that is both mixed and transformed works best because the transformation controls for residual mixing that cannot be precisely measured.

Appendix A: Explanatory Variables

A description of the explanatory variables follows.

Original Loan-to-Value Ratio (LTV): Many mortgages requires a minimum down payment. The loan-to-value ratio is the loan amount divided by property value at origination. A higher LTV will increase the risk of 90-day delinquency since it means a lower down payment, all else equal.

Original Debt-to-Income Ratio (DTI): Banks may require a debt-to-income ratio below a certain level. We expect that a higher DTI will cause the risk of 90-day delinquency to increase and the risk of prepayment to decrease.

Original Fair Isaac Corporation Score (FICO): This is the borrower credit score at origination. It is a number between 300 and 850, and it used to evaluate the quality of borrower credit. We expect that a higher FICO score will have a negative effect on the risk of 90-day delinquency.

Call Option Value: This is the value of the option to prepay. It is computed as the ratio of the difference between the market value of the mortgage and the book value of the mortgage to the market value of the mortgage.

$$Call\,Option_t = \frac{Market\,Value_{tj} - Book\,Value_{tj}}{Market\,Value_{tj}} \quad .$$

Market value and book value are calculated as follows:

$$Market\,Value_t = \sum_{t=1}^{360-t} \frac{Monthly\,Payment}{(1+market\,rate/12)^i} \quad ,$$
$$Book\,Value_t = \sum_{t=1}^{360-t} \frac{Monthly\,Payment}{(1+contract\,rate/12)^i} \quad .$$

In calculating book value, the contract rate is the interest rate at origination, provided by the Fannie Mae dataset. The market interest rate in a given calendar year is calculated as the average note rate of mortgages originated in that year. If the call option variable is positive, the market value is greater than the book value and the probability of prepayment increases.

Negative Equity: This is an important determinant of default risk. The functions used to calculate this are:

$$X_i = \text{ current property value}_i \text{ - remaining balance }_i \quad ,$$

$$\text{negative equity }_i = \begin{cases} \text{absolute value of } X_i & \text{if } X_i < 0 \\ 0 & \text{if } X_i \geq 0 \end{cases} .$$

The current property value is calculated as following:

$$\text{current property value}_i = \left(\frac{\text{Case Schiller index}_i}{\text{Case Schiller index}_0}\right) \times \text{ original property value} \quad ,$$

where Case-Schiller index$_0$ is the Case-Shiller index at the origination date of the mortgage and Case-Schiller index$_i$ is the Case-Shiller index at year i. The remaining balance at month i is provided in the data. If the current property value is less than the remaining balance, it is more likely that the borrower will default. So we expect a positive effect of negative equity on 90-day delinquency.

Unemployment Rate: This is the annual unemployment rate for Miami. The higher the unemployment rate in Miami, the higher should be the risk of 90-day delinquency.

Original Loan Size: This is the original loan amount. We expect that the larger the loan amount, the less likely is prepayment, and the more likely it is delinquency.

Table 4 presents the definitions of our explanatory variables, and Table 5 presents the means and standard deviations.

Table 4: Definition of Explanatory Variables

Variables	Definition
Call Option	(Market Value-Book Value)/Book Value.
Negative Equity	The absolute value of current property value minus current unpaid principal if the current unpaid principal is greater than the current property value, and zero otherwise. The variable is divided by a hundred thousand in the estimation.
Recession Indicator	The recession indicator equals 1, when the calendar year is 2008, 2009, or 2010.
Original FICO Score	The score has a minimum value of 300 and a maximum value of 850. The variable is divided by a thousand in the estimation.
Log Loan Size	Log of the original loan amount. The variable is divided by 10 in the estimation.
Original Debt-to-Income	Monthly Payment/Stable Monthly Income.
Original Loan-to-Value	The original loan amount over the value of the mortgaged property.
Unemployment Rate.	The annual unemployment rate of Miami. The variable is divided by a hundred in the estimation
Dummy Units > 1	The number of units comprising the related mortgaged property. If the number of units is greater than 1, the dummy equals 1, and zero otherwise.
Loan Age Dummy 1 to 16	The loan age dummy i equals 1, when loan age in years equals i. The variables are multiplied by 10 in the estimation.

Table 5: Descriptive Statistics on Mortgage Loans

Variables	Mean	Standard Deviation
Call Option	0.0377	0.0883
Negative Equity	7210.28	22021.01
Recession Indicator	0.1767	0.3813
Original FICO	724.40	56.36
Loan Size	172007.34	92202.62
Original Debt-to-Income	35.81	12.33
Original Loan-to-Value	72.76	16.16
Unemployment Rate	5.773	2.362
Dummy Units>1	0.0195	0.1381
Loan Age 1	0.2220	0.4158
Loan Age 2	0.1921	0.3939
Loan Age 3	0.1423	0.3493
Loan Age 4	0.1030	0.3039
Loan Age 5	0.0785	0.2690
Loan Age 6	0.0622	0.2416
Loan Age 7	0.0487	0.2153
Loan Age 8	0.0377	0.1906
Loan Age 9	0.0300	0.1706
Loan Age 10	0.0236	0.1517
Loan Age 11	0.0169	0.1289
Loan Age 12	0.0130	0.1132
Loan Age 13	0.0105	0.1018
Loan Age 14	0.0084	0.0912
Loan Age 15	0.0064	0.0799
Loan Age 16	0.0045	0.0671

Appendix B

Theorem 1. *Suppose we have a transformed multinomial logit model with J outcomes or choices where $\beta_J = 0$. Let Y_j, $j = 1, ..., J-1$ be the indicator variable for outcome or choice j. Define $\rho_j = \exp(z_j\beta)$ and $S = \sum_{k=1}^{J-1} \exp(z_k\beta)$. Then, for each j, there exists a transformation parameter value c such that the value of $var(Y_j)$ increases above its value in the corresponding untransformed model (the case where $c = 1$), except when $\frac{\rho_j}{1+S} = \frac{1}{2}$.*

Proof. With or without the Ding-Tian-Yu-Guo transformation in a multinomial logit model, the probability function for the indicator variable, Y_j, $j = 1, ..., J-1$, is the same as for the binomial indicator variable:

$$P(Y_j = y_j) = \pi_j^{y_j}(1-\pi_j)^{1-y_j} \quad .$$

The variance of Y_j is $\pi_j(1-\pi_j)$. We now show that for each j, there exists a value of c such that the variance of Y_j in a transformed multinomial logit has a larger variance than the corresponding Y_j in an untranformed model.

Let σ_{jc}^2 be for the variance in a transformed model, and σ_j^2 be the variance in an untransformed model. Then $f_j = \sigma_{jc}^2 - \sigma_j^2$ is equivalent to

$$f_j = [(1 - \frac{1}{(1+cS)^{\frac{1}{c}}})\frac{\rho_j}{S}][1 - (1 - \frac{1}{(1+cS)^{\frac{1}{c}}})\frac{\rho_j}{S}] - (\frac{\rho_j}{1+S})(\frac{1+S-\rho_j}{1+S}) \quad , \quad (1)$$

where $\rho_j = \exp(z_j\beta)$ and $S = \sum_{k=1}^{J-1} \exp(z_k\beta)$. Define $l_j = (1 - \frac{1}{(1+cS)^{\frac{1}{c}}})\frac{\rho_j}{S}$. Equation (1) can now be rewritten as:

$$f_j = l_j(1-l_j) - (\frac{\rho_j}{1+S})(\frac{1+S-\rho_j}{1+S}) \quad . \quad (2)$$

Taking the first derivative with respect to c, we get

$$\frac{\partial f_j}{\partial c} = \frac{\partial l_j}{\partial c} - 2l_j\frac{\partial l_j}{\partial c} = (1-2l_j)\frac{\partial l_j}{\partial c} \quad . \quad (3)$$

Define $g = \frac{1}{(1+cS)^{\frac{1}{c}}}$. Then

$$\frac{\partial g}{\partial c} = \frac{\partial ln(g)}{\partial c} g = \frac{\partial(-\frac{1}{c} ln(1+cS))}{\partial c} \frac{1}{(1+cS)^{\frac{1}{c}}} = (\frac{1}{c^2} ln(1+cS) - \frac{1}{c} \frac{S}{1+cS}) \frac{1}{(1+cS)^{\frac{1}{c}}} \quad , \tag{4}$$

and

$$\frac{\partial l_j}{\partial c} = -\frac{\rho_j}{S} \frac{\partial g}{\partial c} = -\frac{\rho_j}{S} (\frac{1}{c^2} ln(1+cS) - \frac{1}{c} \frac{S}{1+cS}) \frac{1}{(1+cS)^{\frac{1}{c}}} \quad . \tag{5}$$

Plugging equation (5) into equation (3) yields:

$$\frac{\partial f_j}{\partial c} = -(1-2l_j) \frac{\rho_j}{S} \frac{1}{c^2} (ln(1+cS) - \frac{cS}{1+cS}) \frac{1}{(1+cS)^{\frac{1}{c}}} \quad . \tag{6}$$

It is straightforward to show that $ln(1+cS) - \frac{cS}{1+cS} > 0$. Therefore, we have an internal solution only if $l_j = \frac{1}{2}$, which in turn requires that $\frac{\rho_j}{S} > \frac{1}{2}$. Moreover,

$$\frac{\partial^2 f_j}{\partial c^2} = \frac{\partial^2 l_j}{\partial c^2} - 2(\frac{\partial l_j}{\partial c})^2 - 2l_j \frac{\partial^2 l_j}{\partial c^2} = -2(\frac{\partial l_j}{\partial c})^2 < 0 \tag{7}$$

at $l_j = \frac{1}{2}$.

Denote the maximum value of f_j as f_j^*. Then

$$f_j^* = \begin{cases} > 0, & \frac{\rho_j}{1+S} \neq \frac{1}{2} \\ = 0, & \frac{\rho_j}{1+S} = \frac{1}{2} \end{cases} \quad . \tag{8}$$

To see this, substitute $l_j = \frac{1}{2}$ into equation (2):

$$\begin{aligned} f_j &= \frac{1}{2}(1 - \frac{1}{2}) - (\frac{\rho_j}{1+S})(\frac{1+S-\rho_j}{1+S}) \\ &= (\frac{\rho_j}{1+S})^2 - (\frac{\rho_j}{1+S}) + \frac{1}{4} \end{aligned} \quad . \tag{9}$$

We want to show that when $\frac{\rho_j}{S} \leq \frac{1}{2}$, there always exits a value of c for which $f_j > 0$. For this, we need to demonstrate two facts: (1) $\frac{\partial f_j}{\partial c} < 0$ for $\frac{\rho_j}{S} \leq \frac{1}{2}$, and (2) $f_j > 0$ as c approaches 0 from the right. For the first fact, note that all of the factors in $\frac{\partial f_j}{\partial c}$ in equation (6) are positive, except $-(1-2l_j)$. But this factor is negative because

$$1 - 2l_j = 1 - 2(1 - \frac{1}{(1+cS)^{\frac{1}{c}}}) \frac{\rho_j}{S} > 1 - 2\frac{\rho_j}{S} > 0 \quad . \tag{10}$$

To see the second fact, note that $\frac{\partial f_j}{\partial c} < 0$ implies that $f_j(c) > 0$ for c in the interval $(0, 1)$, because the transformed and standard multinomial logit models are equivalent when $c = 1$.

□

Appendix C

Table 6: Multinomial Logit with and without Transformation

(Standard deviations are in parentheses.)

Estimate	Model 1	Model 2
Delinquency		
Baseline Intercept (Group 1)	3.301	6.596
	(0.242)	(0.694)
Baseline Intercept (Group 2)	0.302	4.126
	(0.263)	(0.635)
Baseline Intercept (Group 3)	0.967	-2.338
	(0.253)	(2.182)
Baseline Intercept (Group 4)	-7.859	0.832
	(0.145)	(0.643)
Recession Indicator	0.942	1.325
	(0.114)	(0.176)
Negative Equity	6.714	8.750
	(0.389)	(0.738)
Negative Equity Square	-2.749	-3.973
	(0.290)	(0.432)
Negative Equity * Recession	-0.727	-0.433
	(0.271)	(0.334)
FICO	-1.526	-2.132
	(0.067)	(0.186)
Log Loan Size	-0.039	2.192
	(0.298)	(1.132)
Dummy Units>1	0.004	0.095
	(0.142)	(0.151)
Debt-to-Income	1.402	2.109
	(0.127)	(0.500)
Unemployment Rate	0.582	0.823
	(0.169)	(0.222)
Original Loan-to-Value	3.248	4.719
	(0.233)	(0.541)
Prepayment		
Baseline Intercept (Group 1)	-1.594	-4.203
	(0.175)	(1.606)
Baseline Intercept (Group 2)	-5.219	-0.708
	(0.375)	(0.299)
Baseline Intercept (Group 3)	-2.579	-2.375
	(0.286)	(0.284)
Baseline Intercept (Group 4)	-2.839	-6.297
	(0.285)	(0.477)
Call Option	7.176	10.46
	(0.105)	(0.637)
FICO	-0.178	-0.312
	(0.018)	(0.048)
Log Loan Size	2.024	2.469
	(0.424)	(0.534)
Dummy Units>1	-0.595	-0.871
	(0.120)	(0.176)
Debt-to-Income	-0.602	-0.711
	(0.126)	(0.199)
Mass Point (Group 1)	0.572	-0.441
	(0.182)	(0.426)
Mass Point (Group 2)	-1.049	2.445
	(0.149)	(0.202)
Mass Point (Group 3)	0.057	1.690
	(0.364)	(0.255)
c		5.473
		(0.399)
Log Likelihood	-21944.80	-21909.46

References

[1] Agarwal, S., Deng, Y., and He, J. 2014. "Time preferences, mortgage choice and mortgage default," *Available at SSRN 2447684.*

[2] Ambrose, B., Capone, C., and Deng, Y. 2001. "Optimal put exercise: an empirical examination of conditions for mortgage foreclosure," *Journal of Real Estate Finance and Economics* 23:213-234.

[3] Amemiya, Takeshi. 1978. "The estimation of a simultaneous equation generalized probit model." *Econometrica: Journal of the Econometric Society* 1193-1205.

[4] An, X., Deng, Y., and Gabriel, S.A. 2019. "Default option exercise over the financial crisis and beyond," *Available at SSRN 2764026.*

[5] Archer, W.R., Ling, D.C., and McGill, G.A. 2002. "Prepayment risk and lower-income mortgage borrowers," *Low Income Homeownership: Examining the Unexamined Goal* 279-321.

[6] Calhoun, C.A., and Deng, Y. 2002. "A dynamic analysis of fixed-and adjustable-rate mortgage terminations," *In New Directions in Real Estate Finance and Investment* 9-33.

[7] Chipman, J.S. 1960. "The foundations of utility," *Econometrica* 28:193-224.

[8] Clapp, J.M., Deng, Y., and An, X. 2006. "Unobserved heterogeneity in models of competing mortgage termination risks," *Real Estate Economics* 34:243-273.

[9] Cox, D.R. 1972. "Regression Model and Life-tables," *Journal of Roal Statistical Society*, Series B, 34:187-220.

[10] Debreu, Gerard. 1960. "Review of RD Luce, Individual choice behavior: A theoretical analysis." *American Economic Review* 50: 186-188.

[11] Deng, Y., Quigley, J.M., van Order, R. 2000. "Mortgage terminations, heterogeneity and the exercise of mortgage options," *Econometrica* 68:275-307.

[12] Ding, A., Tian, S., Yu, Y., and Guo, H. 2012. "A class of discrete transformation survival models with application to default probability prediction," *Journal of the American Statistical Association* 107:900-1002.

[13] Fagerland, M.W., Hosmer, D.W., and Bofin, A.M. 2008. "Multinomial goodness-of-fit tests for logistic regression models," *Statistics in Medicine* 27:4238-4253.

[14] Findley, M.C. and Cappozza, D.R. 1977. "The Variable Rate Mortgage: An Option Theory Perspective," *Journal of Money, Credit and Banking* 9:356-364.

[15] Follain, J.R., Ondrich, J., and Sinha, G.P. 1997. "Ruthless prepayment? Evidence from multifamily mortgages," *Journal of Urban Economics* 41:78-101.

[16] Follmann, Dean A., and Diane Lambert. 1989. "Generalizing logistic regression by nonparametric mixing." *Journal of the American Statistical Association* 84(405): 295-300.

[17] Hausman, J. and McFadden, D. 1984. "Specification tests for the multinomial logit model," *Econometrica* 52:1219-1240.

[18] Heckman, James, and Burton Singer. 1984. "A method for minimizing the impact of distributional assumptions in econometric models for duration data." *Econometrica: Journal of the Econometric Society* 271-320.

[19] Kau, J.B., Donald C.K., Walter, J.M., and James, F.E. 1992. "A generalized valuation model for fixed-rate residential mortgages," *Journal of Money, Credit and Banking* 24:279-299.

[20] LaCour-Little, M., and Malpezzi, S. 2003. "Appraisal quality and residential mortgage default: evidence from Alaska," *The Journal of Real Estate Finance and Economics* 27:211-233.

[21] Lancaster, Tony. 1979. "Econometric methods for the duration of unemployment." *Econometrica: Journal of the Econometric Society* 939-956.

[22] McFadden, D. 1978. "Modeling the choice of residential location," Cowles Foundation Discussion Paper No.477, Yale University.

[23] McFadden, D. 1987. "Regression-based specification tests for the multinomial logit model," *Journal of Econometrics* 34:63-82.

[24] McFadden, D., Train, K., and Tye, W.B. 1977. "An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model," Institute of Transportation Studies, University of California.

[25] Merton, R.C. 1973. "Theory of rational option pricing," *The Bell Journal of Economics and Management Science* 4:141-183.

[26] Pennington-Cross, A. 2003. "Credit history and the performance of prime and nonprime mortgages," *The Journal of Real Estate Finance and Economics* 27:279-301.

[27] Shumway, T. 2001. "Forecasting bankruptcy more accurately: A simple hazard model," *The Journal of Business* 74:101-124.

[28] Trussell, James, and Toni Richards. 1985. "Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure." *Sociological methodology* 15: 242-276.