



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Multihorizon discrete time survival models

**Citation for published version:**

Breeden, JL & Crook, J 2020, 'Multihorizon discrete time survival models', *Journal of the Operational Research Society*. <https://doi.org/10.1080/01605682.2020.1777907>

**Digital Object Identifier (DOI):**

[10.1080/01605682.2020.1777907](https://doi.org/10.1080/01605682.2020.1777907)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Journal of the Operational Research Society

**Publisher Rights Statement:**

This is an Accepted Manuscript of an article published by Taylor & Francis in Journal of the Operational Research Society on 7 August 2020, available online:  
<https://www.tandfonline.com/doi/full/10.1080/01605682.2020.1777907>

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Multihorizon Discrete Time Survival Models

Joseph L Breeden  
Prescient Models LLC  
breeden@prescientmodels.com

Jonathan Crook  
University of Edinburgh  
j.crook@ed.ac.uk

January 8, 2020

## Abstract

The new accounting standards of CECL for the US and IFRS 9 elsewhere require predictions of lifetime losses for loans. The use of roll rates, state transition and 'vintage' models has been proposed and indeed are used by practitioners. The first two methods are relatively more accurate for predictions of up to one year, because they include lagged delinquency as a predictor, whereas 'vintage' models are more accurate for predictions for longer periods, but not short periods because they omit delinquency as a predictor variable. In this paper we propose the use of survival models that include lagged delinquency as a covariate and show, using a large sample of 30 year mortgages, that the proposed method is more accurate than any of the other three methods for both short-term and long-term predictions of the probability of delinquency. We experiment extensively to find the appropriate lagging structure for the delinquency term. The results provide a new method to make lifetime loss predictions, as required by CECL and IFRS 9 Stage 2.

Keywords: Cox Proportional Hazards Models; Discrete-time Survival Models; Age-Period-Cohort Models

## 1 Introduction

IFRS9 is an international accounting standard that requires a lender to predict the expected cash flow and hence any loss from a loan if the risk of its default changes. All large banks were required to implement this regulation from January 2018. CECL is a similar regulation applicable to lenders of all sizes in the US. Between them they apply to all large banks globally. Whilst there is a large literature on the the accounting principles underlying these regulations there are few papers that address the appropriate modelling techniques to use and lenders have found the modelling of such losses challenging. The aim of this paper is to propose a new method to model expected losses over the lifetime of an account which will yield more accurate predictions of losses than established methods and that will enable the analyst to make predictions both in the short-term, that we define as over a 12-month horizon, and over a longer-term. The

method also gives point in time estimates, as required by IFRS 9 and useful in CECL.

Finalised in 2014, IFRS 9 makes requirements according to whether or not a loan experiences a significant increase in the risk of default since initial recognition. For a non credit impaired loan, forecasts of expected losses over the ensuing 12 months are required. If there has been, or is expected to be a significant increase in credit risk after the first 12 months, then the bank must predict the lifetime expected credit losses for each account with interest computed on a gross basis. This is known as a Stage 2 credit loss forecast. If there is a credit impairment then expected lifetime credit losses must be computed and interest treated on a net basis. In each case expected losses are to be computed as the probability of default in a month multiplied by the cash shortfalls in the event of default. Expected losses in each month in the future are to be discounted. CECL was implemented in the US in 2016 and required a lifetime expected loss to be computed for all loan accounts without use of stages.

Accurate prediction of expected losses from a loan portfolio is a vitally important challenge faced by lenders. The expected losses determine the provisions that each lender needs to charge borrowers and to retain as a form of capital to protect depositors in the event of the lender becoming bankrupt. So accurate prediction impacts borrowers in terms of the interest rates they are charged. It is also important for depositors who receive some degree of protection, and it impacts shareholders whose return on equity may be reduced if greater provisions are retained as accrued profits. Clearly banks wish to compute the most appropriate amount of provisions given the riskiness of the loans but not an excessive amount that unnecessarily reduces the ability of the lender to make loans.

One can consider models that may be used to predict losses for a portfolio as a whole, by segments of a portfolio, or at the level of an individual account. In the first case the literature discusses two methods to compute expected losses of a loan portfolio in the event that the credit risk increases (see for example Brunel et al 2016 [10] or Siddiqi 2006 [27]). One method is to use roll rates. A roll rate may be computed in terms of balances in which case it is defined as the aggregate value of balances outstanding in a state of delinquency at time  $t$  divided by the balance in the immediately lower delinquency state in the previous time period. It is computed for each number of months delinquent. The lifetime loss for the portfolio is then defined as the sum of the balances that reach a default state, such as six months delinquent, where the latter is the balance at the end of the previous period of accounts that are 5 payments behind times the roll rate from the state of being 5 behind to 6 behind.

Various authors have noted (Brunel et al 2016 [10] McPhail and McPhail (2014) [23] that roll rates traditionally assume past rates remain valid into the future, or they may be forecast using an ARIMA model, but they do not include any explanatory factors that may allow their value in the future to be predicted in a systematic way if such predictors change. However, best practice at financial institutions often model the time series of the roll rates with transformations of macroeconomic factors. That approach was followed by Breeden 2018 [7] and is

followed here. Still roll rates do not capture fluctuations due to changes in the credit quality of the loans originated or the maturing of loans from origination.

A second portfolio method is the so called 'vintage' or 'Age-Period Cohort' model [31, 22]. In this approach aggregate default rates are decomposed into the influence of the time a group of loans was started (originated), the time since the loan has been opened (age or duration time) and factors that influence the probability of default and that vary over time like macroeconomic factors (see Glenn 2005 [15], Breeden 2010 [6], and McPhail 2014 [23]).

Turning to account level models one may consider a state transition model. Here the probability that an account transits from one state of delinquency to another is predicted. The expected lifetime loss can be computed as the probability of default weighted by possible loss over the expected or defined life of a loan. There are several methodologies for making these predictions. In one, the probability that an account will transit from being in one specific state,  $j$ , to another,  $k$ , between time  $t$  and  $t + 1$  is estimated using a cumulative posterior probability model in terms of the value of the state the account is in in previous periods, lagged macroeconomic variables and the lagged age of the loan. In this approach the regression coefficients are specific to the state of the account at  $t$  and assumed to be the same no matter which destination state the account transits to (see for example Malik and Thomas 2012 [21]). Published studies (Malik and Thomas [21]) using this approach use a behavioural score as the state indicator, which has been estimated from some other model. This may be problematic if what the analyst is after is a state in terms of days past due because the behavioural scoring model may not be accurate and because a behavioural score does not translate directly into a delinquency state. The model of Malik and Thomas also omits covariates presumably because in some way they are included in the behavioural model, but it means one cannot predict transition probabilities for new cases without having the behavioural model score in previous periods. There may also be estimation issues. The model implicitly relates the transition probabilities to a set of covariates in a behavioural model, but the final model is estimated in two stage models with errors at both stages, yet the error distribution of the implicit model has not been specified.

A second method is to estimate a multistate intensity model where the probability of transiting between state  $j$  and  $k$  is specified as a hazard model where the set of regression parameters is specific to the transition rather than to the initial state. This approach is followed by Lando and Skodaberg 2002 [18] for corporates and by Leow and Crook (2014) [20] and Djeundje and Crook (2018) [13]. The latter included a frailty parameterisation and macroeconomic variables as well application and behavioural variables. These models enable probabilities of cure as well as advanced delinquency. Published applications of this methodology to credit risk have defined states in terms of days past due and estimated the coefficients in one step.

A multistate intensity model would enable an analyst to predict losses arising from each transition into each state, for example when a payment is missed but the account does not default. But from a practitioner point of view such models are relatively complex and a simpler method that still enables the prediction

of losses in each month in the life of a loan may suffice. Hazard models of the probability that an account will move into the default state in the next month, given that it has not done so before is an example of such a method. There is a rapidly growing literature on the application of survival models to the probability of credit default. Again there are several approaches. One is to use application, and lagged behavioural and macroeconomic variables as predictors. Recent literature for retail credit default includes Banasik et al (1999) [1] and Stepanova and Thomas (2002) [28] (both omit macroeconomic variables) and Bellotti and Crook (2009, 2012, 2013) [5] [4] [3] and Djeundje and Crook (2018) [13] that include macroeconomic variables.

Empirical evidence and practitioner experience of using the roll rate and transition probabilities approaches is that they are relatively accurate when predicting over 12 months but less accurate when predicting over longer time horizons (see for example FDIC 2007 and McPhail and McPhail 2014). Vintage models such as age-period-cohort models are accurate for aggregate forecasting over long horizons, but can be quite inaccurate in the beginning as they do not consider account-level details such as delinquency. When predicting losses for IFRS9 and CECL, some accounts will have only a short remaining duration while others require a long forecast horizon. Therefore, we need model accuracy over both short and long forecast horizons.

In this paper we propose a variant of the hazard model approach that is more accurate than either roll-rates or transition models over both the short term and as good as vintage models for long term forecasts. Such predictions are necessary across all loan types. We parameterise the model using a combined Fannie Mae and Freddie Mac conforming mortgage portfolio and demonstrate its relative accuracy in comparison to roll rate, vintage, and state transition models. Thirty-year term, fixed interest rate mortgages were selected for the test, because they are an important asset class and a worst-case for loss reserves under CECL. The data provides loan-level performance detail with commonly considered origination and behavioural factors.

This paper makes the following contributions to the literature. First we present a new variant of a hazard model to predict the probability of default in each month of the life of an account. Our innovation is to include a thorough investigation of the role of past delinquency of various degrees as a predictor. Second our proposed model is tested on a large sample of 30 year conforming mortgage accounts. Third our tests show that our proposed model makes more accurate predictions than either roll rate methods or a state transition method using scoring and macroeconomic factors. Fourth we show that this is because of the enhancement of predictive power of lagged delinquency terms and we show that the marginal effects of lagged delinquency varies systematically with the lag, in the same way, for each type of delinquency.

Our paper is structured as follows. Following this Introduction we explain the model. Section three describes the application date and the fourth section presents the results of this application. In sections five and six we compare the predictive accuracy of the model to roll rate, vintage and state transition models. the final section concludes.

## 2 Model Description

The model development begins with concepts from Age-Period-Cohort (APC) models [25, 22, 15], where loan performance can be described as a combination of three functions. This is a panel model with time varying variables. One function is loan repayment performance as a function of age of the loan,  $F(a_i)$  where  $a$  denotes age of a loan. This represents a lifecycle pattern of performance over the life of a loan. A second function represents performance as a function of origination date,  $G(v_i)$  where  $v_i$  is the origination date for account  $i$ . Loans originating around a similar date, would be expected to have been of a similar maximum level of risk and a pool of such loans to be of similar risk composition. This relationship is often referred to as a 'vintage' effect. The third function represents performance as a function of calendar date,  $H(t_i)$  where  $t_i$  denotes calendar time for account  $i$ . This relationship relates performance to exogenous factors such as macroeconomic conditions. The functions can be parameterised in many ways, but the most general form is to use a set of dummy variables, one for each value of the relevant time variable. Each dummy variable is represented as  $\delta(u)$  where

$$\delta(u) = \begin{cases} 1, & \text{if } u_i = \text{age } a_i \text{ or vintage } v_i \text{ or } t_i \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

and  $i$  denotes case  $i$ .

$$F(a) = \sum_a \alpha_a \delta(a) \quad (2)$$

$$G(v) = \sum_v \beta_v \delta(v) \quad (3)$$

$$H(t) = \sum_t \gamma_t \delta(t) \quad (4)$$

The complete model can then be written as

$$Def(i, a, v, t) \sim F(a_i) + H(t_i) + G(v_i) \quad (5)$$

where  $Def(i, a, v, t) = 1$  if account  $i$  defaults and 0 if it does not.

The coefficients  $\alpha_a$ ,  $\beta_v$ , and  $\gamma_t$  need to be estimated. Because only one overall intercept term may be estimated, the estimates are constrained so that

$$\sum_v \beta_v \delta(v) = \sum_t \gamma_t \delta(t) = 0. \quad (6)$$

Because of the relationship  $t = v + a$ , an assumption must be made about how to remove the linear specification error. Consistent with earlier work (see Appendix in Breeden and Canals-Cerdá [8]), this is accomplished using an orthogonal projection onto the space of functions that are orthogonal to all linear functions. The coefficients obtained are then transformed back to the original specification.

Traditional APC or Bayesian APC estimators [26] process vintage aggregate data. But in the current work we are interested in a loan-level estimator.

Default is modelled conditioned on the account not having previously closed. Any account closure that is not a default is called attrition or prepayment. To capture this effect, a parallel equation for modeling attrition is employed.

$$Att(i, a, v, t) \sim F^{att}(a_i) + H^{att}(t_i) + G^{att}(v_i) \quad (7)$$

Although the structure is the same as for the default model, the functions  $F^{att}(a)$ ,  $G^{att}(v)$ , and  $H^{att}(a)$  will be different, because the drivers of attrition performance will be different from those for default.

By using dummy variables for vintage in the APC formulation, the model can measure credit risk but not explain it. Therefore, to enhance the usefulness of the model and to better predict default, information collected at the time of application is added to the formulation.

$$Def(i, a, v, t) \sim F(a) + H(t) + \sum_j c_j s_{ij} + \sum_v \beta_v' \delta(v) \quad (8)$$

where  $s_{ij}$  are the available attributes at origination for account  $i$ . Such attributes include origination FICO score and origination loan-to-value (LTV), for example. The  $c_j$  are the coefficients to be estimated. Vintage dummies are still included, but with new coefficients  $\beta_v'$  such that for a given  $v$ ,

$$G(v_i) = \sum_v \beta_{v_i} \delta(v_i) = \sum_{j, i \in v} c_j s_{ij} + \sum_v \beta_v' \delta(v_i) \quad (9)$$

This approach was used to study root causes of the 2009 US Mortgage Crisis [8]

We know from the work by Holford [16] that once the constant and linear terms are appropriately incorporated, the nonlinear structure for  $F$ ,  $G$ , and  $H$  is uniquely estimable. The treatment of the constant and linear terms described above satisfy this condition, meaning that the vintage aggregate structure  $G(v)$  must be equivalent to the aggregate credit risk for the accounts  $i$  in that vintage, as shown in Eq. 9. Said differently, the estimates of  $F(a)$  and  $H(t)$  do not change with the loan-level attribution of credit risk in Eq. 8. Of course, Eq. 8 will be more accurate at the account level than Eq. 5 that only has a vintage level credit risk measure.

Note that this loan-level version of an APC model is identical to a discrete time survival model where a dummy variable is used for each value of age and time. Therefore, the same parameter estimator is used here as for a discrete time survival model. Notice also that this discrete time survival model is just a panel model with age and time factors added beside the usual explanatory factors.

To achieve the short term accuracy of roll rate and state transition models, we need to incorporate delinquency in the model. This creates complications in any APC or survival model structure, because delinquency is also a function of

the age of the account and the economic and portfolio management environment and therefore correlated to  $F$  and  $H$  that appear in Eqs. 5 & 8. Eq. 10 provides the general structure including delinquency.

$$Def(i, a, v, t) \sim F(a_i) + H(t_i) + \sum_j c_j s_{ij} + \sum_k d_k D_i^k(t - L) + \sum_v \beta_v'' \delta(v) \quad (10)$$

where  $D_i^k(t - L)$  are again indicator variables, one for each possible delinquency state leading up to default.

If all parameters in Eq. 10 are estimated simultaneously,  $F$  and  $H$  will likely change from the values in Equation 8 because of the multicollinearity with the  $d_k D_i^k(t - L)$ . An alternative that will be explored here is to constrain the estimation of the  $d_k$  such that  $F$  and  $H$  are retained from the estimate in Eq. 5. The result would be that

$$G(v_i) = \sum_v \beta_v \delta(v_i) = \sum_{j, i \in v} c_j s_{ij} + \sum_k d_k D_i^k(t - L) + \sum_v \beta_v'' \delta(v_i) \quad (11)$$

Whenever time-varying predictive factors are used, referred to in the lending industry as behavioral factors, one faces the question of how to extrapolate those factors. The question becomes acute in lifetime loss forecasting where long range extrapolations would be required. However, forecasting the delinquency factors above is comparable in difficulty to forecasting default. Since the primary drivers of future delinquency would again be some measure of age and time effects, little new information is gained beyond that already available in the default equation.

Therefore, the current approach is to create a set of models using the structure of Eq. 10, each with different lags on the behavioral factors. If a model uses only behavioral factors with lags as small as  $L$ , then that model can be used to forecast  $L$  steps into the future without the need to forecast the input factors. Thus, if we have a set of  $N$  models, one each for  $L \leq l$  for  $l \in [1, N]$ , then a full set of forecasts out to  $N$  periods into the future can be created without forecasting the input factors. The hope is that as  $N$  becomes large, the coefficients  $c_j$  and  $d_k$  approach limiting values such that forecasts can continue to be generated for forecast horizons greater than  $N$ , again without forecasting the input factors.

If we define the model by the minimum allowed lag for the behavioral factors,

$$Def^L(i, a, v, t) = F(a_i) + H(t_i) + \sum_j c_j s_{ij} + \sum_v \beta_v' \delta(v_i) + \sum_k d_k D_i^k(t - L), \quad (12)$$

then forecasts are generated from age  $a_0$  and time  $t_0$  as

$$Def(i, a_0 + L, v, t_0 + L) = M^L(i, a_0, v, t_0). \quad (13)$$

$Def^L$  is a kind of discrete time survival model [11, 3, 30], one such model for each forecast horizon (value of  $L$ ).



A parallel approach is used for modeling account attrition / pay-off. To get to charge-off balance, additional models of exposure at default (EAD) and loss given default (LGD) would be required. Those are not considered here, since they are independent questions potentially requiring different modeling techniques.

## 3 Data

### 3.1 Mortgage Data

Data was obtained from Fannie Mae and Freddie Mac for 30-year, fixed-rate, conforming mortgages. The data contained de-identified, account-level information on balance, delinquency status, payments, pay-off, origination (vintage) date, origination score, postal code, loan to value, debt to income, number of borrowers, property type, and loan purpose. Risk grade segmentation was defined such that Subprime is less than 660 FICO, Prime is 660 to less than 780, and Superprime is 780 and above.

The data in this study represents more than \$2 trillion of conforming mortgages. The data made available by Fannie Mae and Freddie Mac is a large share of their respective portfolios, but not the entirety. Figure 1 shows the historic default rate aggregated by annual vintage.

### 3.2 Macroeconomic Data

As part of the government's implementation of the Dodd-Frank Stress Test Act (DFAST), the Federal Reserve Board regularly releases Base, Adverse, and Severe scenarios for a set of macroeconomic factors. Since these factors and scenarios have become industry standards, this study has focused on the use of these factors for incorporating macroeconomic sensitivity. For mortgage modeling, the most interesting are real gross domestic product (GDP), real disposable income growth, unemployment rate, CPI, mortgage interest rate, house price index, and the Dow Jones stock market index (DJIA).

## 4 Model Estimation

Model estimation occurs in three stages. The first step is to apply an APC decomposition to estimate the lifecycle, environment, and vintage quality functions. During that stage, methods such as described in Breeden & Thomas [9] may be applied to achieve linear trend stability. The second step is to include those lifecycle and environment functions into Equation 12, and estimate Equation 12 as a logistic regression, one estimation for each forecast horizon,  $L$ . The third step is to estimate a time series model of the environment function using macroeconomic factors.

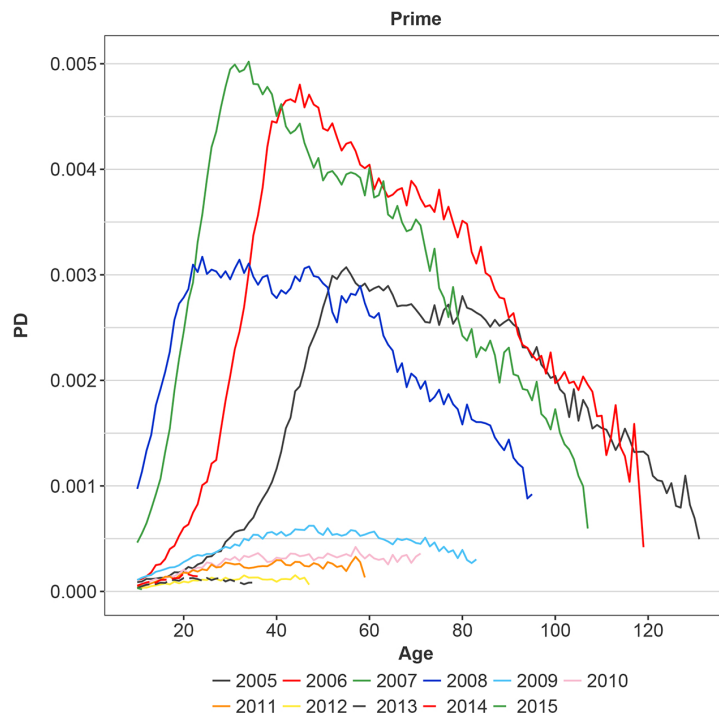


Figure 1: A plot of probability of default (PD) for the training data by annual vintage.

## 4.1 APC Decomposition

A Bayesian APC algorithm was used to estimate the functions in equation 5. The estimated PD lifecycle versus age of the account,  $F(a)$ , is shown in Figure 2. The estimated relationship between PD and vintage is shown in Figure 3. The estimated relationship between PD and calendar date,  $H(t)$  is shown in Figure 4. In all three figures the functions are segmented by credit quality: subprime, prime or superprime.

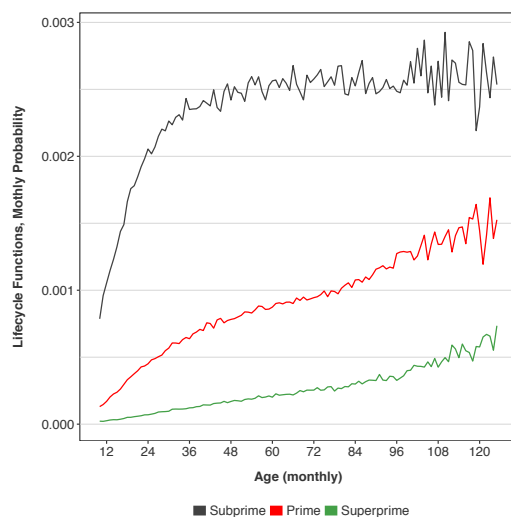


Figure 2: The estimated lifecycle PD function versus age of the account estimated for Equation 5 segmented by origination score.

The lifecycles include the overall constant term for the analysis and are transformed to probability of default per month so that they may be understood more intuitively than on a log-odds scale. They show that the PD of the subprime segment rises faster relative to account age and to a higher level compared to the PD of the superprime segment that rises gradually and always at a much lower level.

The PD vintage functions in Figure 3 are mean-zero and on a log-odds scale. They show that the credit cycle was most severe for the superprime segment. Proportionally, the 2006 - 2008 vintages were worst for superprime even though loss amounts were smaller than for subprime. After 2010 the estimates for each vintage are much more volatile, because those vintages have far fewer cases.

The PD calendar time functions in Figure 4, also on a log-odds scale, show that the macroeconomic impacts were nearly equivalent across all risk bands on a proportional basis. Since these risk bands are all for conforming mortgages, they do not include the most extreme non-conforming loans. Differences may arise when taken to those extremes.

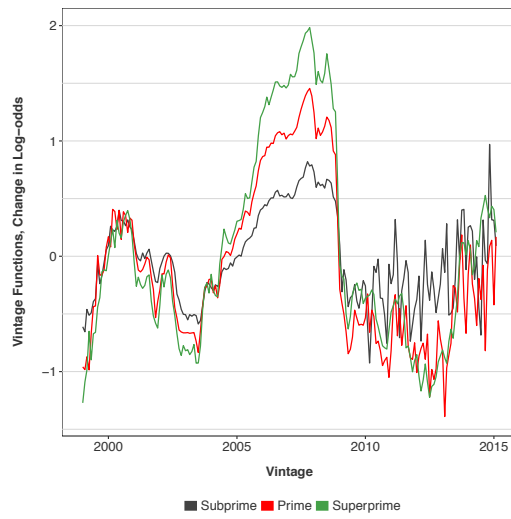


Figure 3: The credit quality function versus vintage estimated for Equation 5 segmented by origination score.

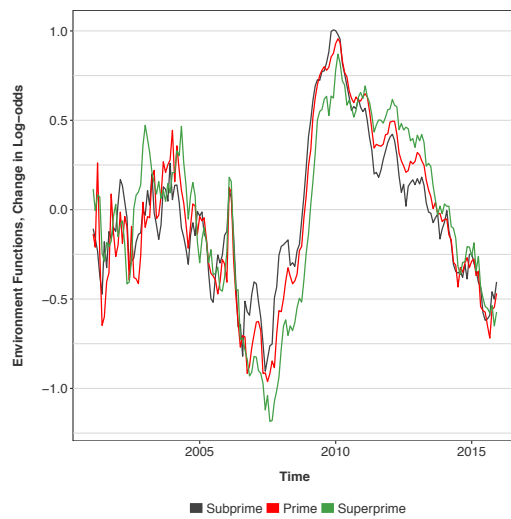


Figure 4: The environment function versus calendar date estimated for Equation 5 segmented by origination score.

## 4.2 Loan-level Coefficients

The discrete time survival model represented by equation 12 was estimated separately for each credit risk segment and each forecast horizon. Through testing, the model coefficients were found to stabilize by horizon 12, so no further coefficients were estimated. The coefficients for horizon 12 were applied to all horizons greater than 12.

A separate model was created to predict at the time of origination that did not include any behavioral factors such as delinquency since no values of behavioural factors exist in the first few months after account opening. The origination model was applied to all accounts less than 6 months on book at the start of the forecast.

The estimation of equation 12 takes the estimated parameters for the life-cycle function,  $F(a)$ , and the environment function,  $H(t)$ , as fixed, so the coefficients  $c_j$ ,  $d_k$ , and  $\beta_v''$  are estimated to replace the vintage effects in the PD model. The result is a set of estimations, one for each value of  $L$  and risk segment. Vintage dummies,  $\beta_v''$ , are included as possible explanatory factors in order to capture adverse selection or consumer risk appetite as was observed by Breeden & Canals-Cerda [8].

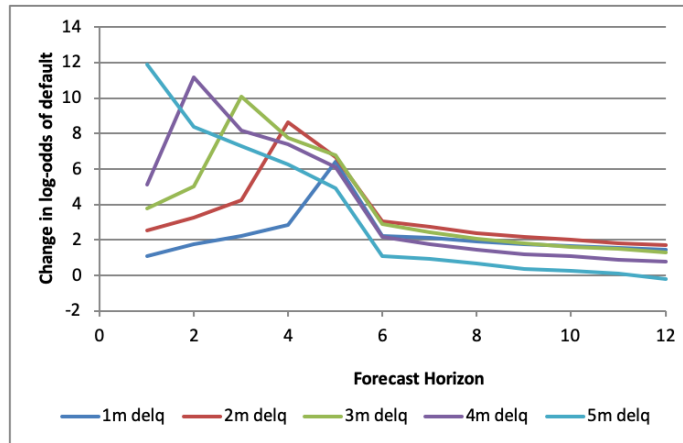
Figures 5 and 6 graph the coefficients  $d_k$  and  $c_j$  respectively, one for each lag and credit quality. Although estimated independently, Figures 5 by risk grade show nearly identical structure. Each line is the probability of default with forecast horizon for accounts in a certain delinquency state at the start of the forecast. Default is either contractual default at 6 months delinquency or caused by non-contractual reasons such as bankruptcy, fraud, or deceased.

Because the coefficients are functions of the forecast horizon, the importance to the forecast depends upon how far into the future one is trying predict. Figure 5 is a good example of this.

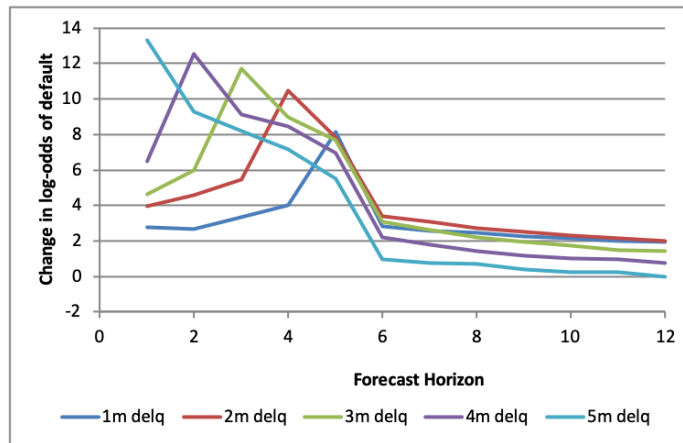
When trying to predict one month forward, severe delinquency is the strongest predictor of default. An account that is 2 months delinquent at the start of the forecast has the greatest risk of default at horizon 4. For all delinquency states, the predicted value declines dramatically beyond 6 months into the future, because delinquent accounts will most likely have either cured or defaulted by then. In fact, the most severely delinquent accounts (5 months delinquent in this analysis) are less likely to default at horizon 6 than a less delinquent account. Presumably, this is because any 5-month delinquent accounts that are still active by 6 months into the future must have cured and therefore are not such a severe default risk.

All of the delinquency coefficients are measured relative to non-delinquent accounts, which are thus assigned a coefficient of 0. This means that any delinquent account is more risky than a non-delinquent account, but the relationship for delinquent accounts is highly nonlinear. If only contractual default were considered, then the coefficients for all delinquency states would be 0 until enough months had elapsed for the account to move from the current state to default, defined as 6 months delinquent.

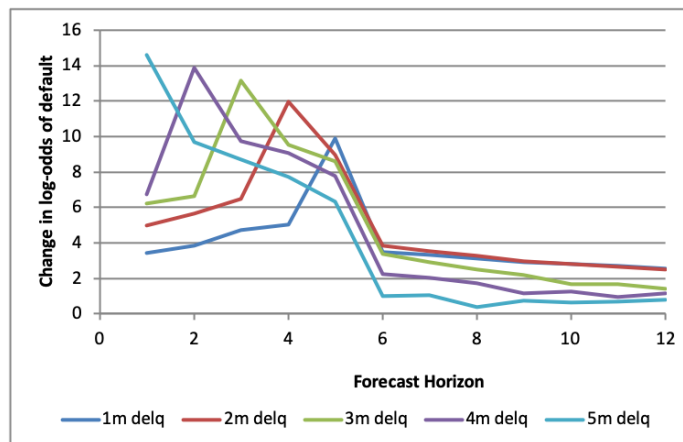
Figure 6 shows the coefficients for some of the other variables in the scores.



(a) Subprime



(b) Prime



(c) Superprime

Figure 5: Coefficients predicting default probability by forecast horizon for each delinquency state.

Note that in the first few months of the forecast, the other candidate variables make almost no contribution to the forecast – delinquency is everything. As the predictive value of delinquency diminishes around horizon 6, the other factors take over. By horizon 12, the coefficients have almost converged to the values that appear in the origination score, meaning that for long-range forecasting, origination information still dominates behavioral information. This is probably not true for credit cards, for example, where the transactor / revolver distinction is critical for the entire forecast.

The model is already segmented by FICO, so it does not appear directly in the explanatory factors. Although tested, it was no longer significant beyond the initial segmentation.

The importance of current delinquency decays rapidly with time. However, as delinquency loses value, other measures such as loan to value ratio (LTV), debt to income ratio (DTI), and number of borrowers take over. These coefficients are shown in Figure 6. Unlike delinquency, the information content of these three origination variables is stable for long horizons. The coefficients do not return to the origination model values, because behavioral variables like delinquency provide some information. The ability to adapt to the information decay rate of each variable is what provides the combination of short-term and long-term forecast accuracy.

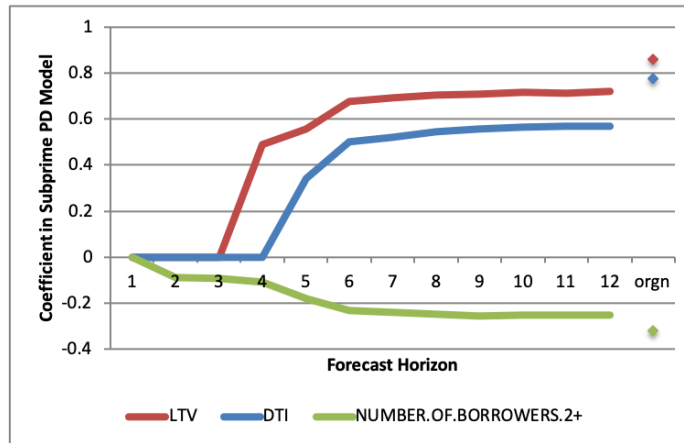
In addition to the factors shown here, the PD models for subprime and superprime also included origination balance. The model for prime included origination balance, loan purpose, and occupancy status.

Figure 7 shows the coefficients for annual vintage dummy variables. Note that the Superprime line is the most volatile, because the fewest number of defaults were available for the modeling. Overall, these agree with the results of a previous study on consumer risk appetite [8], showing that the 2005-2008 vintages had significant residual credit risk that was not captured by the available scoring factors.

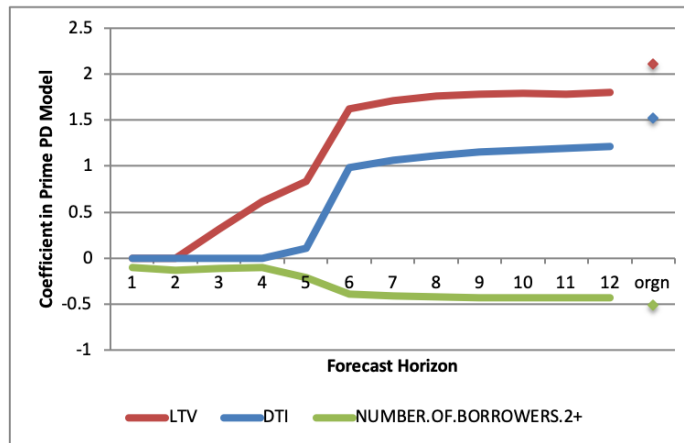
#### 4.2.1 Model fit tests

Because the multihorizon model uses 12 separate regressions for the behavioral model plus one for the origination model, the model fits and discrimination tests of each model were measured separately. Figure 8 shows the coefficients for pseudo- $R^2$ , defined as  $1 - \text{Residual Deviance} / \text{Null Deviance}$ . For short forecast horizons, the pseudo- $R^2$  is very high. It falls dramatically as delinquency information decays and approaches the pseudo- $R^2$  for the origination model as the forecast horizon becomes large.

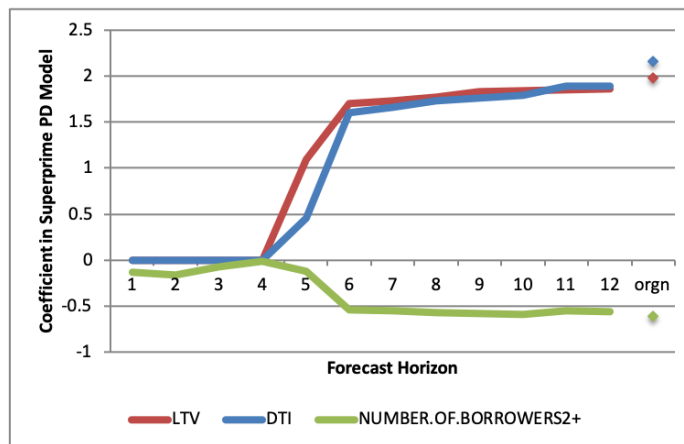
Discrimination ability was measured by the Gini coefficient, which was again computed separately for each forecast horizon and the origination model, Figure 9. This shows that discrimination ability is strong through the first 6 months of the forecast and then again decreases toward the origination values.



(a) Subprime



(b) Prime



(c) Superprime

Figure 6: Coefficients predicting default probability by forecast horizon or origination score for some key predictive factors.



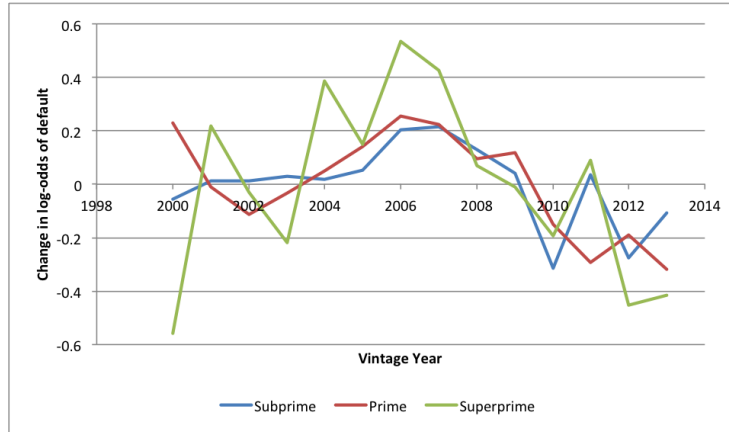


Figure 7: Coefficients for the vintage dummies in the discrete time survival model. The same coefficients were used for all forecast horizons.

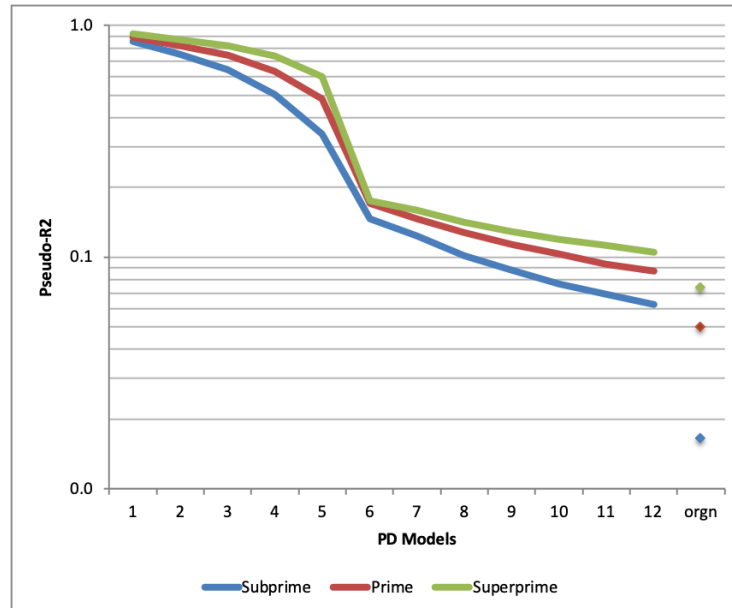


Figure 8: Pseudo- $R^2$  by forecast horizon.

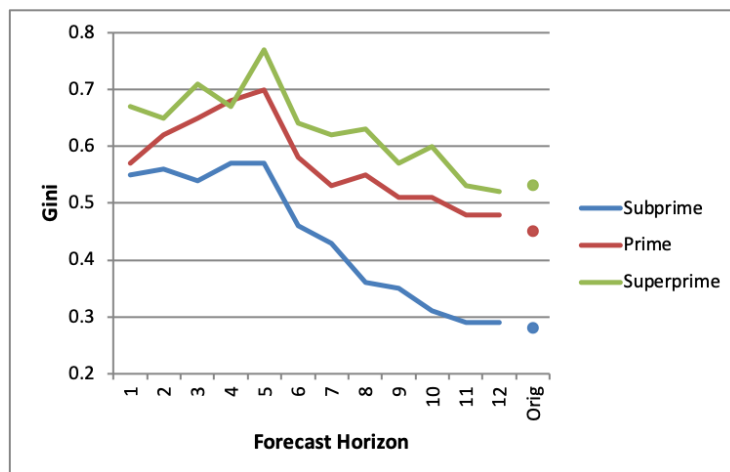


Figure 9: Gini discrimination tests for PD by forecast horizon and the origination model.

### 4.3 Macroeconomic modeling

The macroeconomic model was created as a time series model of the environment functions in Figure 4 using transforms and lags of macroeconomic factors [9]. The macroeconomic model is used only to translate macroeconomic scenarios into scenarios of the environment function. In the current context, the primary purpose is to provide a fair comparison to alternative methods where all models can be given the same input macroeconomic factors.

The transformations were optimized individually via an exhaustive search of the lags and transform widths. The optimized transforms were then combined in all possible combinations up to a maximum of five factors to find the one that had the best adjusted  $R^2$  while still having the intuitively correct sign for the factor and significant p-values using a Newey-West robust estimator [24]. Five was considered to be a stopping point, since no five-factor models survived the above constraints due to loss of significance.

The macroeconomic modeling was done concurrently with estimating seasonal effects. The seasonality coefficients are shown in Figure 10 along with confidence intervals. All of the other coefficients in the loan-level and macroeconomic analysis were tested to ensure that the p-values were significant. However for seasonality, the zero level was arbitrarily set to January, so the monthly coefficients should not be tested separately for significance. Instead, it must be viewed as a function to determine if the range of variation is significant relative to the confidence intervals, which it is.

The macroeconomic models are summarized in Table 1. The factors are listed along with the transformation used, lags applied, and the width of the transformation. For example, House.Price.Index.LogRatio.L12.W24 in the subprime model refers to the transformation in Equation 14

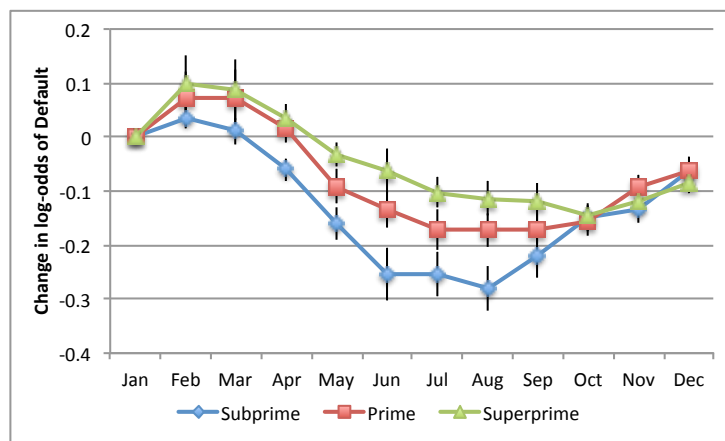


Figure 10: Seasonality coefficients measured for each segment including the standard errors.

$$House.Price.Index.LogRatio.L12.W24 \equiv \log \left( \frac{HPI(t-12)}{HPI(t-12-24)} \right) \quad (14)$$

Although all the consumer-related factors in the DFAST scenarios were considered, the ones chosen through this automated process were intuitively obvious: unemployment rate (UR), house price index (HPI), and disposable personal income (DPI). DPI and HPI were only considered with the log-ratio transformation demonstrated in Equation 14 in order to ensure stationarity. The proper transformation of UR is not clear, because past studies have shown that both levels and changes in UR can be predictive. In this case, both were tested but moving averages were found to be optimal.

## 5 Alternate Approaches

To provide a means of comparison, several other standard models were estimated on the same data set. For brevity, the full estimations for those models are not shown, but a summary description is provided for each and the results are included in the forecast comparisons. We wish to compare the predictive accuracy of the multihorizon model with alternative methodologies.

### 5.1 Vintage models

The multihorizon survival model developed here can be viewed as an extension of age-period-cohort (vintage) models. Therefore it is reasonable to ask what would happen if the analysis stopped with aggregate vintage modeling instead of going to loan level modeling.

Subprime Coefficients	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	-0.90	0.17	-5.2	1.15E-06
Real.DPI.LogRatio.L1.W24	-0.67	0.40	-1.7	9.43E-02
UR.MovingAvg.L0.W1	0.16	0.03	5.8	5.90E-08
HPI.LogRatio.L12.W24	-0.39	0.11	-3.6	5.71E-04
Adj-R2	0.91			
Prime Coefficients	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	-1.77	0.15	-11.8	2.2e-16
UR.MovingAvg.L0.W2	0.27	0.02	12.4	2.2e-16
Adj-R2	0.93			
Superprime Coefficients	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	-1.94	0.38	-5.2	1.00E-06
UR.MovingAvg.L3.W2	0.28	0.05	6.0	2.17E-08
Adj-R2	0.90			

Table 1: The macroeconomic models created for each segment.

The vintage models for probability of default and probability of attrition / pay-off were used to predict the expected number of defaults each month for each vintage. Balances were not modeled, since different models are often used to model exposure at default and loss given default. The test here is only of the probability of default and probability of attrition models.

The vintage model results shown below use the lifecycle and environment from Figures 2-4, but re-estimate the vintage coefficients at the beginning of each test period. This is only a partial out-of-sample test, because the training data would be too short to create a stable estimate of the lifecycle and macroeconomic model without at least one full economic cycle.

## 5.2 Roll Rate Models

Roll rate models are another aggregate modeling approach [14], and one that has been in use in lending since at least the 1960s. Balance roll rates are the most commonly used, because they are the most simple. However, to create a fair comparison to the default account forecasts for the other models, account roll rates are used here, defined as

$$R_k(t) = \frac{Accounts_k(t)}{Accounts_{k-1}(t-1)} \quad (15)$$

where  $k \in [1, 6]$  is the months delinquent,  $Accounts_k(t)$  is the number of accounts in delinquency  $k$  at calendar month  $t$ , and  $R_k(t)$  is the net roll rate.  $R_6 \equiv R_D$ , which is the roll to default.

To capture account pay-offs, an additional payoff rate,  $R_p$ , is required.

$$R_p(t) = \frac{\text{Payoff Accounts}(t)}{\text{Accounts}_0(t-1)} \quad (16)$$

These roll rates are modeled historically with macroeconomic factors using the same kind of process described in Section 4.3, wherein the roll rate time series are modeled with transformations of the same macroeconomic factors listed previously. Again, a grid search over possible lags and windows for the transformations was used, and all models must satisfy criteria related to significance and the sign of the relationship.

To create forecasts, the following equations were used.

$$\text{Accounts}_0(t) = \text{Accounts}_0(t-1) * (1 - R_p(t)) \quad (17)$$

$$\text{Accounts}_k(t) = \text{Accounts}_{k-1}(t-1) * R_k(t) \quad (18)$$

The final lifetime loss is calculated by summing the default accounts until all  $\text{Accounts}_k$  reach 0. Note that in these formulas, no new accounts are included, so  $\text{Accounts}_0$  should decrease with time.

### 5.3 State Transition models

State transition models [17, 2, 29, 19, 12] are the loan-level equivalent of roll rate models. Rather than modeling aggregate movements between delinquency states, the probability of transition is computed for each account. The states considered are current (not delinquent), delinquent up to a maximum of five months delinquent, default, and pay-off. Account transition probabilities are modeled. For an account  $i$  in state  $j$  at time  $t$ , the transition probability to state  $k \in [0...5, D, P]$  is given as

$$p_{j \rightarrow k}(i, t) \sim c_0 + \sum_{l=1}^n c_l x_l(i, t) \quad (19)$$

Each transition with sufficient data was modeled with a separate logistic regression to estimate coefficients  $c_l$  for predictive factors  $x_l(i, t)$ . This includes all state transitions  $p_{j \rightarrow k}$  for which  $k - j \in [2, 1, 0, -1]$ , meaning two forward transitions and one backward transition. Transitions to other states that were too rare for modeling ( $k - j \notin [2, 1, 0, -1]$ ) were included as constant probabilities. Also modeled were all transitions  $p_{i \rightarrow 0}$ ,  $s_{j \rightarrow D}$ , and  $s_{j \rightarrow P}$ , where  $D$  and  $P$  refer to default and prepayment. As noted before, default can occur from any delinquency state because of bankruptcy, fraud, deceased, or abandonment.

The transition matrix is close to diagonal because monthly transitions are being modeled. Transitions over longer periods, like quarters or years, would cause a greater spread in the transition matrix.

The regression factors  $x_l(i, t)$  in Equation 19 include macroeconomic factors such as for the roll rate models, measures of age of the account such as ( $age, age^2, sqrt(age), log(age)$ ), and scoring factors such as FICO, LTV, etc.

Factors with insignificant coefficients were removed following the same process as in Section 4.2.

Although tested, measures of previous delinquency states were not predictive for a given transition. Therefore, these models satisfied the Markov criterion and forecasting was done via efficient matrix multiplies. When summed over all accounts, expectation values in each state were produced. The forecasts are run iteratively until all accounts either payoff or default.

## 6 Model Accuracy

The goal of the multihorizon survival model is to provide both near-term and long-term accuracy. Therefore, instead of reporting a single cumulative error over a forecast horizon, the accuracy was measured monthly for each forecast horizon. Starting periods for the tests were scattered non-uniformly across the data range to avoid synchronizing with seasonal effects. The test periods were 36 months each; 36 months was chosen as a maximum in order to maximize the number of test periods.

Even through the performance data was available back to 2005, this time period represents only one economic cycle. To properly test a model using macroeconomic factors, an out-of-sample recession would be needed. Instead, we acknowledged that we only have enough data to estimate the macroeconomic models and lifecycles in-sample. The out-of-sample testing kept those macroeconomic correlations and lifecycles from the full time history, but re-estimated all scoring coefficients using data only up to the start of each test period.

Figure 11 shows the median average absolute error across tests by forecast horizon. Figure 12 shows the cumulative forecast error with forecast horizon.

For the first three months of the forecast, the roll rate, state transition, and multihorizon DTS models outperform the vintage model. This is consistent with industry observations. Since roll rate and state transition models focus on delinquency and the results above show the importance of delinquency, this is to be expected. The vintage model does not consider any current performance information, so it is prone to near-term discontinuities.

However, in the long-term, the vintage model largely maintains its level of accuracy due to its emphasis on the lifecycle effect. The roll rate and state transition models deteriorate substantially over the long-term, because delinquency loses its predictive value. Since both models are one step ahead predictors, their coefficients are optimized for the near-term use of delinquency and largely lose sensitivity to other possible effects that would be useful in long-term forecasting.

The multihorizon survival model outperforms all models in the early months and attains a long-run accuracy just above that of the vintage model. When the cumulative errors are considered, the early advantages of including delinquency give the multihorizon survival model an advantage over the vintage model.

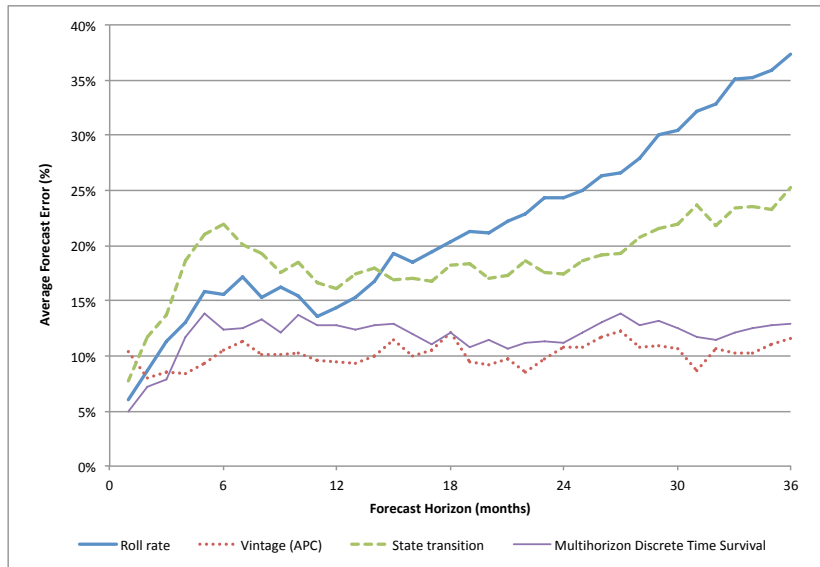


Figure 11: The average absolute forecast error is shown versus forecast horizon for each model, averaged over all test periods.

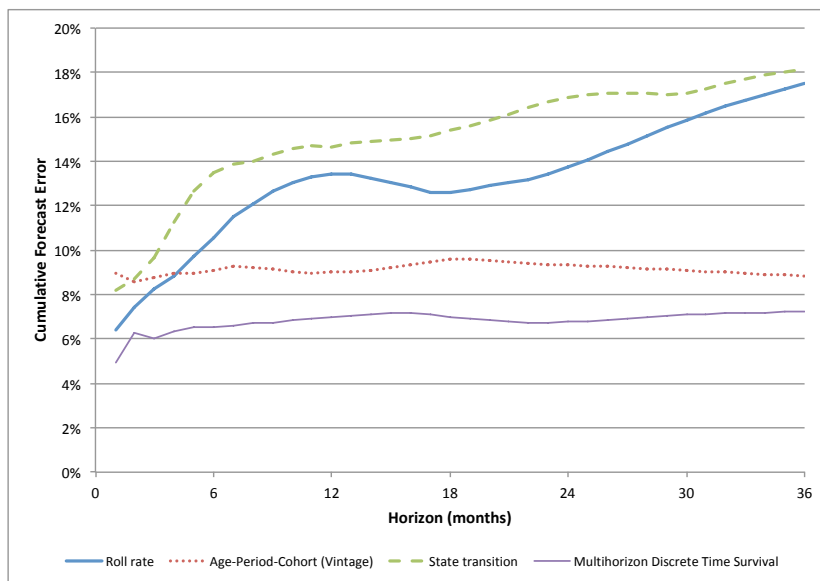


Figure 12: The absolute cumulative forecast error is shown versus forecast horizon for each model, averaged over all test periods.

## 7 Conclusions

The multihorizon survival model achieves its intended goal of being more accurate than other models for short term forecasting and comparable to vintage models for long term forecasting. Many organizations currently use two separate models for short-term and long-term forecasting, so this provides a best-of-both-worlds approach in a single modeling framework. This is particularly beneficial in the context of IFRS 9 and CECL so that accurate loss reserves are created that capture the best current information about the loans and the lifetime expectations.

In the process of estimating the multihorizon survival model, reviewing the coefficients versus forecast horizon appears to explain this performance advantage. For short horizons, the coefficients capture the extreme nonlinearity of delinquency. For longer horizons, these coefficients capture the rate of decay of information content across explanatory variables, generally replacing short term predictors (delinquency) with long term predictors (LTV, DTI, etc.)]

Although originally designed to solve the lifetime loss forecasting problem for IFRS 9 and CECL, this modeling technique provides advantages in many other contexts. As a behavior score for account management or collections, the forecasts can be aggregated over any desired horizon. Given an economic scenario, the "score" is immediately calibrated to a probability. Moving from rank-order scores to probability estimates has immediate advantages across a range of applications, such as loan pricing and cash flow estimates for existing loans.

## References

- [1] J Banasik, J Crook, and L Thomas. Not if but when borrowers default. *Journal of the Operational Research Society*, 50:1185–1190, 1999.
- [2] Anil Bangia, Francis X Diebold, André Kronimus, Christian Schagen, and Til Schuermann. Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking & Finance*, 26(2):445 – 474, 2002.
- [3] T. Bellotti and J. Crook. Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting*, 29(4):563 – 574, 2013.
- [4] T. Bellotti and J.N. Crook. Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society*, 60:1699–1707, 2009.
- [5] Tony Bellotti and Jonathan Crook. Loss given default models incorporating macroeconomic variables for credit cards. *International Journal of Forecasting*, 28(1):171 – 182, 2012. Special Section 1: The Predictability of



Financial Markets Special Section 2: Credit Risk Modelling and Forecasting.

- [6] Joseph L. Breeden. Testing retail lending models for missing cross-terms. *Journal of Risk Model Validation*, 4(4), 2010.
- [7] Joseph L. Breeden. *Living with CECL: Mortgage Modeling Alternatives*. Prescient Models LLC, 2018.
- [8] Joseph L. Breeden and Jose J. Canals-Cerdá. Consumer risk appetite, the credit cycle, and the housing bubble. *Journal of Credit Risk*, 14(2):1–30, 2018.
- [9] Joseph L. Breeden and Lyn C. Thomas. The relationship between default and economic cycle for retail portfolios across countries: identifying the drivers of economic downturn. *Journal of Risk Model Validation*, 2(3):11 – 44, 2008.
- [10] Vivien Brunel. Lifetime pd analytics for credit portfolios: A survey. 2016.
- [11] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.
- [12] V. Djeundje and J. Crook. Incorporating heterogeneity and macroeconomic variables into multistate delinquency models for credit cards. *European Journal of Operational Research*, 271(2):697–709, 2018.
- [13] Viani Biatat Djeundje and Jonathan Crook. Incorporating heterogeneity and macroeconomic variables into multi-state delinquency models for credit cards. *European Journal of Operational Research*, 2018.
- [14] FDIC. Credit card activities manual. [https://www.fdic.gov/regulations/examinations/credit\\_card/ch12.html](https://www.fdic.gov/regulations/examinations/credit_card/ch12.html), 2007.
- [15] Norval D. Glenn. *Cohort Analysis, 2nd Edition*. Sage, London, 2005.
- [16] T R Holford. The estimation of age, period and cohort effects for vital rates. *Biometrics*, 39(2):311–324, 1983.
- [17] R. Israel, J. Rosenthal, and J.: Wei. Finding generators for markov chains via empirical transition matrices, with application to credit ratings. *Mathematical Finance*, 11:245–265, 2000.
- [18] D. Lando and T. M. Skdeberg. The estimation of transition matrices for sovereign credit ratings. *Journal of Banking & Finance*, 26(2):423–444, 2002.
- [19] M. Leow and J. Crook. Intensity models and transition probabilities for credit cards loan delinquencies. *European Journal of Operational Research*, 236(2):685–694, 2014.

- [20] Mindy Leow and Jonathan Crook. Intensity models and transition probabilities for credit card loan delinquencies. *European Journal of Operational Research*, 236(2):685–694, 2014.
- [21] Madhur Malik and Lyn C Thomas. Transition matrix models of consumer credit ratings. *International Journal of Forecasting*, 28(1):261–272, 2012.
- [22] W.M. Mason and S. Fienberg. *Cohort Analysis in Social Research: Beyond the Identification Problem*. Springer, 1985.
- [23] Joseph McPhail and Lihong McPhail. Forecasting lifetime credit losses: Modelling considerations for complying with the new fasb and iasb current expected credit loss models. *Journal of Risk Management in Financial Institutions*, 7(4):375–388, 2014.
- [24] Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.
- [25] Norman B. Ryder. The cohort as a concept in the study of social change. *American Sociological Review*, 30(6):843–861, 1965.
- [26] Volker Schmid and Leonhard Held. Bayesian age-period-cohort modeling and prediction - bamp. *Journal of Statistical Software, Articles*, 21(8):1–15, 2007.
- [27] Naeem Siddiqi. *Credit risk scorecards: developing and implementing intelligent credit scoring*, volume 3. John Wiley & Sons, 2012.
- [28] Maria Stepanova and Lyn Thomas. Survival analysis methods for personal loan data. *Operations Research*, 50(2):277–289, 2002.
- [29] L. C. Thomas, J. Ho, and W. T. Scherer. Time will tell: behavioural scoring and the dynamics of consumer credit assessment. *IMA Journal of Management Mathematics*, 12(1):89–103, 2001.
- [30] Gerhard Tutz and Matthias Schmid. *Modeling Discrete Time-to-Event Data*. Springer, 2016.
- [31] Y. Yang and K.C. Land. *Age-Period-Cohort Analysis*. Taylor and Francis, Boca Raton, 2014.