Incorporating lifecycle and environment in loan-level forecasts and stress tests

Joseph L Breeden
Prescient Models LLC
300 Catron St., Suite B
Santa Fe, NM 87501, USA
breeden@prescientmodels.com
+1 505-490-6042

June 3, 2016

Abstract

The new FASB current expected credit loss (CECL) proposal, IASB's IFRS 9, and regulatory stress testing all require that the industry move toward forecasting probabilities of future events, rather than simply rank-ordering loans. Even more importantly, effective loan pricing requires this same forward-looking, loan-level forecasting.

We created a loan-level version of Age-Period-Cohort (APC) models suitable for forecasting individual loan performance at a point-in-time or for the loan's lifetime. The APC literature explains that any model of loan performance must make either an explicit or implicit assumption around the embedded model specification error between age of the loan, vintage origination date, and performance date. We have made this assumption explicit and implemented a technique using augmented macroeconomic history to stabilize the analysis.

The preceding steps provide robust estimates of lifecycle and environmental impacts. We then use a Generalized Linear Model (GLM) with a population odds offset for each age / time combination derived from the lifecycle and environment functions in order to estimate origination and behavior scores. Analyzing a small US auto loan portfolio, we demonstrate that this model is robust out-of-sample and out-of-time for predicting both rank-ordering and probabilities by inserting the odds offset appropriate for the environment being modeled.

In addition to producing loan-level forecasts and stress tests, the scores produced have higher rank-order performance out-of-sample and out-of-time than standard scores. The scores prove to be robust years into the future with no measurable degradation in performance because of the stabilizing effect of the offset factor during model construction.

Keywords: Forecasting; Risk, Banking; Time series; Age-Period-Cohort Models

1 Introduction

Credit scores were originally developed to aid loan origination. Applicants would be screened by estimating a score from available application data. The score served as an impartial criterion for assessing risk. The "cut-off" score was the threshold below which riskier applicants were denied loans.

The transition to risk-based pricing meant that a wider range of applicants could be accepted by modifying the loan terms to match the riskiness of the applicant. Risk-based pricing came in response to competitive pressure among lenders. As margins contracted, lenders needed to better target their pricing.

As lenders became more reliant on scores, greater effort was put into improving score accuracy. Improved estimation techniques, the transition to logit and probit models, and the use of bureau attributes all enhanced the ability of the scores to assess risk.

The scorecard uses these factors from a dataset, $\boldsymbol{x}$, to predict the probability of "good", $p(G|\boldsymbol{x})$. The odds of a loan being good are [21]

$$o(G|\boldsymbol{x}) = \frac{p(G|\boldsymbol{x})}{p(B|\boldsymbol{x})} = \frac{p_G}{p_B}\frac{p(\boldsymbol{x}|G)}{p(\boldsymbol{x}|B)} \equiv o_{Pop} \times I(\boldsymbol{x}), \boldsymbol{x} \in X \tag{1}$$

or

$$\log(o(G|\boldsymbol{x})) \equiv \log(o_{Pop}) + \log(I(\boldsymbol{x})) \tag{2}$$

where p_G and p_B are the unconditional odds of good or bad, $\frac{p_G}{p_B}$ is the population odds o_{Pop}, and $I(\boldsymbol{x}) = \frac{p(\boldsymbol{x}|G)}{p(\boldsymbol{x}|B)}$ is the information odds. We could also say that o_{Pop} captures systematic effects for the portfolio and $I(\boldsymbol{x})$ captures the idiosyncratic effects for an individual loan.

When we create a credit score, we expect the information odds to be reasonably robust out-of-sample. The population odds, however, are dependent upon the macroeconomic conditions prevailing during the in-sample period. In future time periods, the population odds should change due to factors not captured in the credit score. For this reason, credit scores are used as risk ranking tools out-of-time, not as predictors of $o(G|\boldsymbol{x}_{oot})$ where $\boldsymbol{x}_{oot}$ are the loan attributes out-of-time. (Out-of-time refers to data from time periods not included in the training sample.)

Many practitioners expand the attributes $\boldsymbol{x}$ to include macroeconomic factors and the age of the loan in an attempt to predict the population odds as well, but with mixed results. As explained in the Age-Period-Cohort (APC) literature [18, 14] and applied to the context of credit risk modeling [8], a model specification error is embedded in the dynamics of retail lending. Traditional credit scoring attributes are measured in the loan origination month, also known as the vintage date, v. Macroeconomic data is measured with calendar date t. Lifecycle functions as in survival models [10, 16, 20, 12] are measured versus the age of the loan a. However, $a = t - v$, leading to a linear specification error if factors measured along all three dimensions are included in the model simultaneously and without constraint. In cases where some of these dimensions are excluded, as with traditional credit scores that rely solely on information from the origination (vintage) date, a unique solution is obtained, but at the cost of being unable to predict probabilities in future time periods.

The APC literature proves that no general solution exists for this specification error, with the implication that we can never be certain of the linear trends in lifecycle, macroeconomic, or credit risk functions. Instead, domain-specific solutions are recommended by incorporating constraints suitable to the specific situation being modeled. An inability to be certain of linear trends in time would me that we could not reliably predict the population odds in future time periods, brining the effectiveness of any stress test model into question.

Breeden and Thomas [8] described a constraint that would appear to be reasonable for retail lending. Namely, that after fitting macroeconomic factors to the environment function, the slope of that macroeconomic impact should be zero when extrapolated backward across multiple macroeconomic cycles. For any model that includes macroeconomic factors, this is equivalent to creating an environmental index from a subset of the model as $\hat{E} = f(E_i(t), c_i)$ where the E_i are the individual macroeconomic factors and c_i are the estimated coefficients for their inclusion in the credit risk model. Then create the constraint that $\mathrm{slope}(\hat{E}(t)) = 0$ when $\hat{E}(t)$ is extrapolated backward over multiple decades of macroeconomic history.

Over any short time frame of less than one economic cycle, the environment will certainly not show zero trend, but the trend is also unlikely to be identically zero even over longer periods. Nonetheless, the assumption of zero environmental trend over many cycles is consistent with the assumption that a through-the-cycle average can be defined for macroeconomic impacts, i.e. that a through-the-cycle probability of default (PD) exists in the sense that the population odds can be a constant when measured for a reference portfolio across multiple economic cycles.. Therefore, although the zero environmental trend assumption is a good starting point, it must be tested and potentially corrected if it is not precisely true.

Using the technique of Breeden and Thomas to control the model specification error, we demonstrate a method of creating credit scores that estimates both the population odds and information odds in-sample, and provides for extrapolating the population odds out-of-sample so that loan-level probabilities are forecasted. This paper represents the first time that APC algorithms have been used for account-level forecasting. Incorporating the retrending process of Breeden and Thomas means that the score created will be stable through economic cycles without risk of cross-correlation between scoring factors and macroeconomic factors.

The newly adopted accounting standards of IFRS9 [2] and CECL [13] require just such features. On the one hand, they are both based upon lifetime loss estimation from any point in the age of the loan. That necessitates inclusion of a lifecycle such as is found in survival models or APC models. However, other requirements within the guidelines for IFRS9 and CECL suggest that account-level models are to be preferred. The Cox proportional hazards algorithm [9] in principle applies to such problems, but the paper by Breeden and Thomas demonstrates that a risk of linear trend ambiguity exists in all credit risk models, and subsequent work by Breeden, et. al. [6] demonstrates this problem explicitly for Cox PH. Therefore, the authors believe that a new approach is needed that has the scoring and hazard function attributes of Cox PH, but with explicit control of the linear trend so that no estimation confusion occurs between the scoring and macroeconomic factors.

2 Modeling approach

Although a single-stage approach is in principle possible using a constrained optimization, we followed a sequential analysis using simpler algorithms. The following steps were performed in the analysis.

1. Decompose loan-level performance data with an Age-Vintage-Time (AVT)
2. Fit the time function to macroeconomic data
3. Retrend the age and vintage functions
4. Fit a credit score with age and time offsets

2.1 Age-Vintage-Time decomposition

The first step is to estimate the lifecycle as a function of age of the loan, the vintage quality as a function of vintage, and the environment as a function of time (calendar date). This analysis should be performed on the longest history available, preferably longer than the two to three years of history that is typical of credit scores.

For the Age-Vintage-Time (AVT) decomposition, we can use standard Age-Period-Cohort (APC) implementations to analyze vintage-aggregate time series with a logit transformation, or use an equivalent loan-level implementation of APC. Each vintage will be measured each month to create an appropriate rate. For example, to predict the default rate, default accounts and active accounts would be reported each month. The APC algorithm would estimate

$$r(a, v, t) = \frac{\text{Defaults}(t)}{\text{Active.Accounts}(t-1)} = \frac{1}{1 + e^{-(F(a)+G(v)+H(t))}} \tag{3}$$

where $r(a, v, t)$ is the default rate, $F(a)$ is the lifecycle with age, $G(v)$ is the credit quality by vintage, and $H(t)$ is the environment function over time. With standard implementations of the APC algorithm, all three functions are estimated via splines with the analyst specifying the number of spline nodes. Although $F(a)$ is usually given fewer nodes on the assumption of a relatively smooth lifecycle function, $G(v)$ and $H(t)$ should have as many nodes as the data will support in order to capture sudden changes in the portfolio composition or macroeconomic environment respectively. Alternatively, standard APC implementations usually support nonparameteric estimation of any of the three functions.

For loan-level data, we created an implementation equivalent to the APC algorithm. Although we can use a range of distributions and link functions, a logit transform is again desirable to model default rate. The loan-level data will contain periodic observations of each account until default or voluntary attrition. Each observation will report a binary value for default. After default or voluntary attrition, the loan is no longer reported. In a loan-level analysis, reporting only active loans is equivalent to using active accounts as the denominator of the aggregate rate modeling in Equation 3.

$$logit(p_i(a, v, t)) = F(a) + G(v) + H(t) \tag{4}$$

With the loan-level AVT algorithm, the same spline approximations were available, but nonparametric estimation of any of the functions is also available assuming sufficient data exists to estimate all the coefficients.

Note that the loan-level estimation of Equation 4 still results in population-wide functions of $F(a)$, $G(v)$, and $H(t)$. In fact, aside from estimation errors, both approaches will estimate the same functions on a given data set. These functions essentially capture the population odds in-sample. To predict the population odds out-of-sample, we need to extrapolate $H(t)$ for future environments and move all the loans along the lifecycle function $F(a)$ as they age. Since these functions are designed to capture all of the systematic effects in the portfolio, the remaining structure should be loan-level idiosyncratic effects, as found in the information odds.

The form of Equation 4 expresses independence between the age, vintage, and time functions, except that we know a linear specification ambiguity is present. Therefore, a more precise representation given the assumption here of no linear trend in time would be

$$logit(p_i(a, v, t)) = \alpha_0 + \alpha_1 a + F'(a) + \beta_1 v + G'(v) + H'(t) \tag{5}$$

where α_0, α_1, and β_1 are the constant and linear term coefficients such that $F'(a)$, $G'(v)$, and $H'(t)$ are just the nonlinear pieces.

Previous work [15] has demonstrated that all components of Equation 5 are uniquely estimable, but this does not prove that no cross-terms between age, vintage, or time are expressed in the data. In fact, a test for missing cross-terms was previously developed [5] and applied to the residuals of APC models of consumer loan performance data. The test is essentially a spatial correlation test on the model residuals in the dimensions of age and time. Use to date has shown that in cases where significant cross-terms are present, this can be resolved through proper segmentation by product type, geography, or risk band.

2.2 Macroeconomic fit

The environmental function $H(t)$ is initially estimated by the APC algorithm with the assumption of no net trend with time over the observed portfolio data. Some explicit assumption on the allocation of the linear trend must be made in order for the algorithm to converge. Assuming no linear trend is expected to be appoximately correct when modeling at least a full economic cycle, but we have no proof that the trend is identically zero.

As the data is fit to macroeconomic data, the no-trend constraint is tested via the inclusion of a constant trend in time in order to find the best fit to available macroeconomic factors. To avoid overfitting, we only consider factors that are close to the consumer balance sheet: employment / unemployment / under-employment; house prices; real wages; interest rates, etc.

In each of these cases, careful consideration must be given to the transformations used. Since a logit transformation is used in Equation 4, the $H(t)$ function will be roughly normally distributed. Any explanatory macroeconomic factors should be transformed to be roughly normally distributed as well. For example, the house price index (HPI) is usually reported as year-over-year percentage change. Although intuitively useful, percentage change is asymmetric. A 10% period-over-period increase followed by a 10% decrease does not return the index to its original value. Instead, we borrow from the investment analytics world to select transformations that are symmetric and approximately normally distributed so that linear regression may be employed. A table of preferred transformations maybe be found in Breeden (2010) [4]. For changes in HPI, we should use a log-ratio transformation, $log - ratio(HPI) = log(HPI(t)/HPI(t - w)$ where w is the window over which the change is computed.

Once a transformation has been chosen, we consider lags and moving averages of the transformed values. This is equivalent to a simplified Distributed Lag Model [17]. The lags and moving averages become part of the transformation of the macroeconomic variable prior to creating the final regression model. If several variables are found that contain predictive power, we will attempt to create a multiple regression model,

$$\hat{H}(t) = c_0 t + \sum_i c_i E_i(t) + \epsilon_t, \tag{6}$$

where c_0 is the linear trend coefficient, c_i are the coefficients against transformed macroeconomic factors $E_i(t)$, $\epsilon_t \in \mathcal{N}(0, \sigma)$ and t spans the date range of the observed portfolio performance data, $t \in [0, T_o]$. The process of fitting the environmental function to macroeconomic data has been previously demonstrated for APC-class models [7, 8].

2.3 Retrending the functions

Using the technique of Breeden and Thomas [8], the fitted environmental function $\hat{H}(t)$ is extrapolated backward through previous economic cycles $[-T_h, 0)$ not represented in the portfolio performance data. A straight line is fit through the extrapolation of $\hat{H}(t)$ over the range $t \in [-T_h, T_o]$ as $\hat{H}(t) = \alpha + \beta t$.

The original lifecycle, vintage quality, and time functions are retrended as

$$\begin{aligned} F'(a) &= F(a) + \beta a \\ G'(v) &= G(v) + \beta v \\ H'(t) &= H(t) - \beta t \end{aligned} \tag{7}$$

Although the above equations are mathematically correct, they do not express the data weighting present when the original functions were estimated. This could potentially produce a biased result.

Therefore, the safest way to assure that the retrending will retain unbiased functions, it is best to rerun the AVT estimation with the retrended $H'(t)$ as a fixed input. The estimated $F'(a)$ and $G'(v)$ will preserve the original data density weighting and is generally the easiest approach to guarantee the functions are optimized to the data.

Retrending these functions by leveraging a greater macroeconomic history provides a reasonable solution to the extrapolation problem. We have seen in practice for many forecast or stress test models, particularly in the context of CCAR [1], results that trend strongly upward or downward into the distant future because of the uncontrolled or unobserved trend acquired during initial modeling. By detrending over multiple economic cycles, we create a model that is stationary and consistent with the philosophy that a long-run PD exists and can be modeled. Simply subtracting off the linear trend is reasonable, since the assigned linear trend was arbitrary in the original estimation.

2.4 Fitting the Score

The final modeling step is to use the retrended lifecycle and environment functions to compute the monthly populations odds as a function of the age of the loan and macroeconomic environment. This is referred to as the $\text{offset}(a,t) = F'(a) + H'(t)$ where $H'(t)$ is the detrended fit to macroeconomic data.

With this, a scoring function is created to predict defaults as a function of typical scoring attributes $\mathbf{X}$,

$$\text{logit}(p_i(a,t)) = \text{offset}(a,t) + \mathbf{BX}, \tag{8}$$

where $\mathbf{B}$ are the score coefficients. The offset is equivalent to an attribute with a fixed coefficient of 1. The coefficients are estimated via a generalized linear model (GLM) on a loan-level data matrix of monthly performance observations.

3 Numerical Example

This process was tested on a small US auto loan portfolio. Historic loan-level performance data was available from 2007 through 2012 for vintages originated from January 2000 through December 2012. Figure 1 shows the number of defaults for this portfolio as a function of time. From the graph, a peak in losses is apparent in 2009 and 2010. Although we could directly correlate the loss rate to economic data, we know that multiple problems coincided at this time. In order to create models with future predictive value, we need to separate economic effects from changes in the volume and quality of loan originations through time.

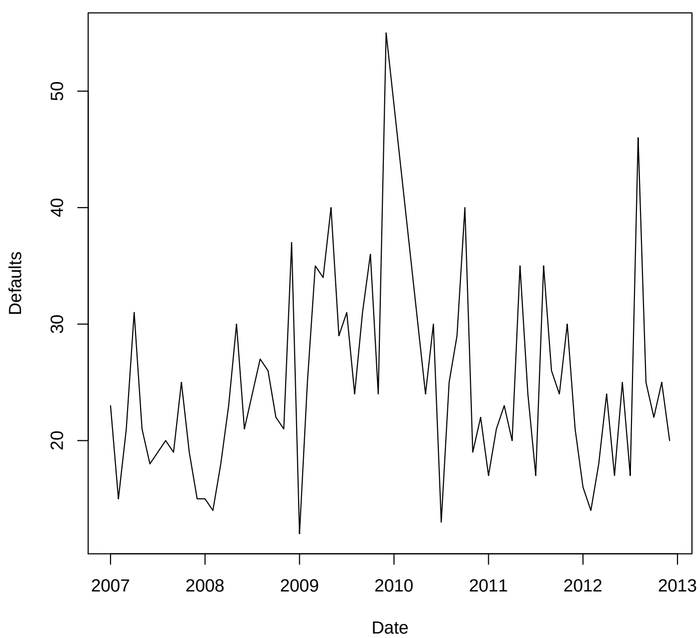

Figure 1: Number of defaults as a function of time.

To develop an intuitive understanding of portfolio performance, managers have used vintage plots for decades. The monthly default rate was definted as

$$DR(a, v, t) = \frac{\text{Default Accounts}(t)}{\text{Active Accounts}(t-1)} \tag{9}$$

Figure 2 shows annual vintage performance for default rate graphed versus calendar date and age (months-on-book).

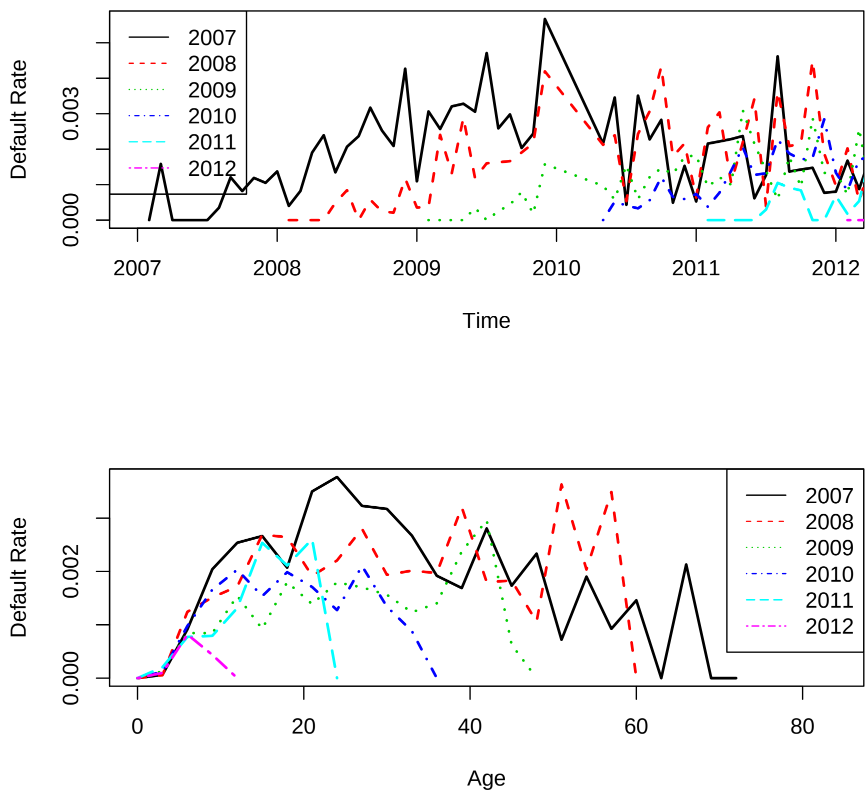

Figure 2: Number of defaults as a function of time.

The vintage graph versus calendar time shows that recent vintages are peaking lower than the vintages which experienced the 2009 recession. The vintage graph versus age provides a visual estimate of the default rate lifecycle (loss timing function).

To quantify what is suggested visually in the above graphs, the steps described in Section 2 were followed.

3.1 APC Analysis

Using a loan-level implementation of Age-Period-Cohort models with an initial assumption that the environment function had no trend over the in-sample data provided the following results. To extract the full structure of the component functions, a Bayesian APC algorithm [19] was employed. The Bayesian approach uses a quick spline estimation as a starting point, but then runs millions of Monte Carlo simulations to measure the distribution of possible values at each point on the functions. The mean of the distribution, point-by-point, provides the functional values and the confidence intervals are measured as the 5% and 95% points on those distributions.

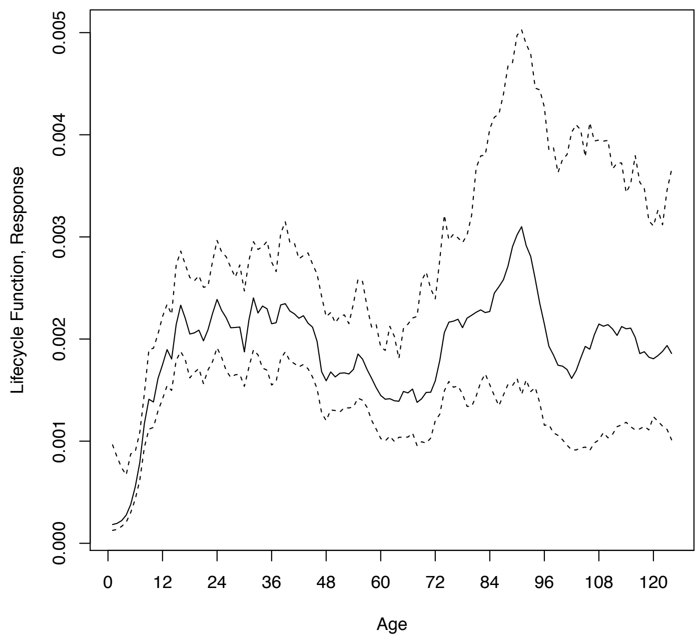

Figure 3: Lifecycle function with account age measured for a small US auto loan portfolio

The lifecycle function extracted from the data in Figure 2 is shown in Figure 3. The y-axis scaling is monthly default rate and the x-axis is months since origination. Although visually noisy from the small data set, the graph captures the expected features. Overall, auto loans generally exhibit a peak in default rate around 24 months and again at end of term. The peak around 24 months is due to gradual financial degradation of the consumer relative to what was known when the loan was approved during underwriting. The peak at end of term appears to have several causes, but is in part a selection effect, since the less risky loans generally pay off before the end of term.

The graph also shows a blend of terms. Peaks around 12, 24, 36, 60, 72, and 84 are because of blending different terms. With a larger portfolio, one option to improve forecast accuracy would be to segment the analysis by term and estimate separate lifecycles for each.

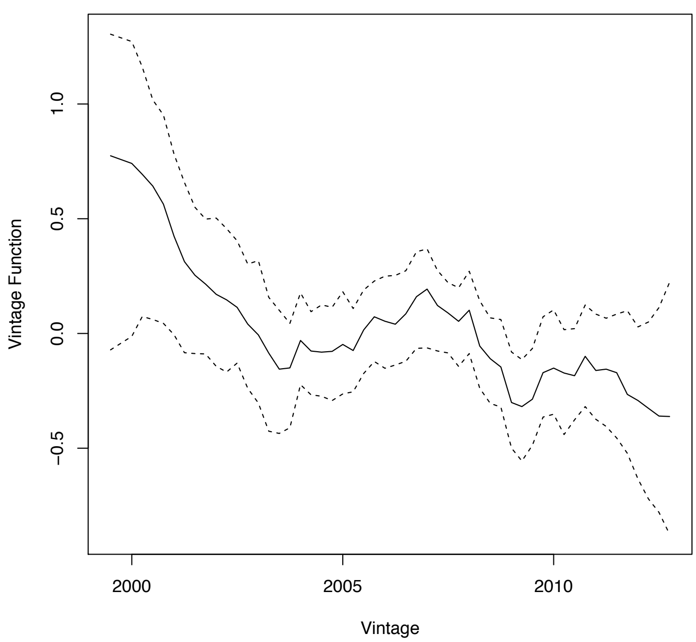

Figure 4: Credit risk function with vintage measured for a small US auto loan portfolio

Relative to the lifecycle function, vintages will have different risk scalings. Figure 4 shows the estimated risk scaling for each vintage. The x-axis is the origination month, and the y-axis is the change in log-odds of default relative to the lifecycle function. The lifecycle function was originally estimated as log-odds of default but was plotted as default rate to aid interpretation.

The pattern shown here matches business intuition. Research on the credit risk of US mortgage originations [3] also showed heightened credit risk for vintages originated between 2005 and 2008 with dramatic improve following the onset of the recession. That work explained the causes as a combination of changes in underwriting and macroeconomic drivers of consumer risk appetite, both of which are probably at present here.

The environment function in Figure 5 is measured as a change in log-odds

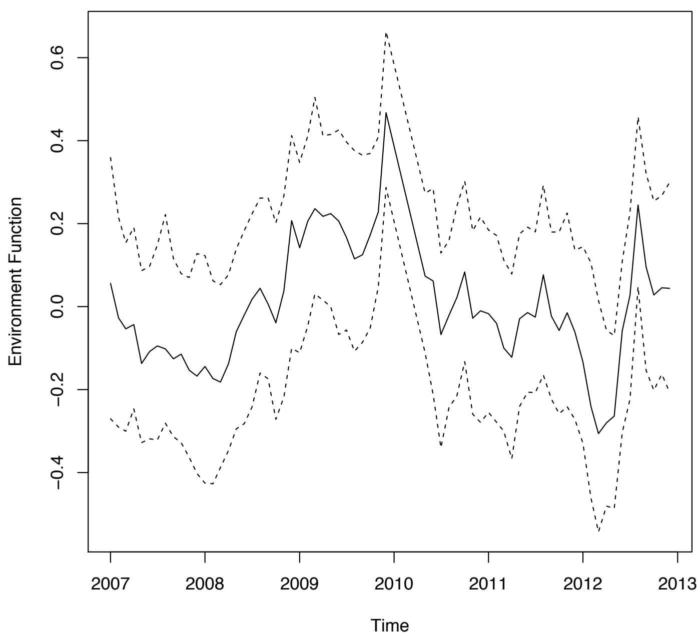

Figure 5: Environmental function with time measured for a small US auto loan portfolio

for all active loans on a given calendar date. The 2009 and its aftermath are clearly visible. In addition, a regional deterioration is apparent in late 2012.

3.2 Macroeconomic Fit

The environment function was compared to a range to macroeconomic indicators available in the geographic region for this portfolio: unemployment rate, housing starts, house price index, current economic indicator, leading economic indicator, CredAbility risk measure, income per capita, and wages per capita. Each economic factor was first transformed using either moving averages for smoothing or log-ratios to compute change over time. The width of the moving average or log-ratio comparison is given by the window parameter. Each transformed macroeconomic variable could also be lagged. A negative lag refers to future values, which is possible, implying that defaults appeared before changes in the corresponding macroeconomic factor.

A grid search was conducted across possible values of lag and the window for the transformation to optimize the fit of each candidate macroeconomic variable independently. The variables with significant correlation were unemployment rate, housing starts, and income per capita. GDP was not available in a timely manner and with enough history for this region, but housing starts serves as a proxy for GDP.

Interest rates were not considered, because the auto loans are fixed rate. In addition, interest rates fell during the recession, giving the seemingly counter-intuitive result that declining rates correlates to rising default. This is true, but only because the Federal Reserve manages interest rates lower in response to economic stress, so the impact to consumers is very indirect and captured by other factors.

The coefficients in Table 1 produce a model with multiple R-squared: 0.60, adjusted R-squared: 0.58, F-statistic: 31.46 on 3 and 64 DF, and p-value: 1.287e-12.

Table 1: Coefficients for the model predicting the environmental function from macroeconomic indicators.

Dependent Variable	Lag	Window	Estimate	Std. Error	z value	$Pr(>\|z\|)$
(Intercept)			9.42e-02	3.66e-02	2.58	0.0123
t			-3.27e-05	5.51e-04	-0.059	0.953
Unemp.lwLogRatio	1	17	0.453	0.116	3.90	0.000235
IncPerCap.lwLogRatio	-1	15	-2.69	0.634	-4.241	7.3e-05

Because the linear trend is not uniquely estimable, we include time as a factor in the model. This tests whether a different trend would produce a better macroeconomic fit. For the model in Table 1, the secular trend was insignificant.

Figure 6 shows the fit of the model in Table 1 to the environment function and the backward extrapolation using older macroeconomic data. A linear fit to to the model is shown.

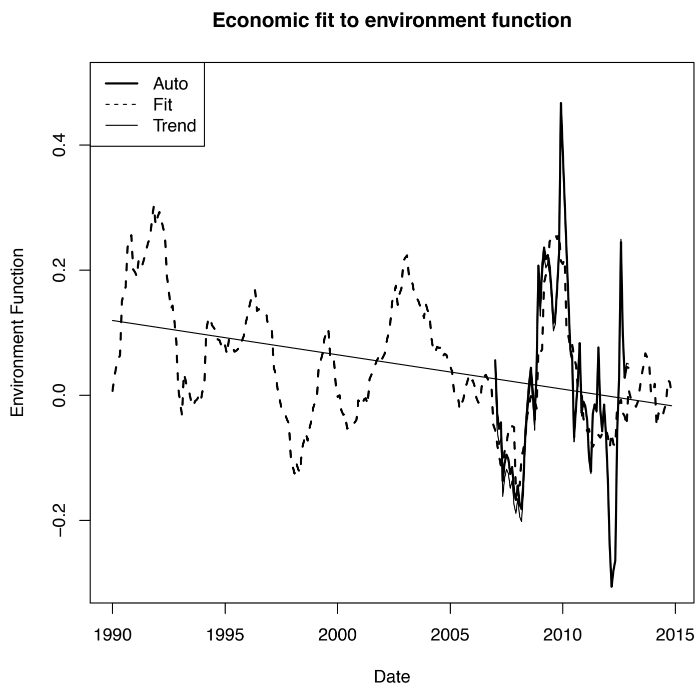

Figure 6: A comparison of the environment function and the macroeconomic fit to that function. The linear trend overlay was fit to the macroeconomic model extrapolation.

3.3 Retrending

The long-term linear trend shown in Figure 6 suggests that a long-range forecast would be biased downward. The measured environment function does not appear to have a trend, but relative to macroeconomic factors it could have a hidden trend, as revealed here. Following Breeden & Thomas [8], the linear trend shown in Figure 6 was subtracted from the environment function. The change was slight, but the macroeconomic model was then refit and extrapolated again.

The coefficients in Table 2 produce a model with multiple R-squared: 0.60, adjusted R-squared: 0.58, F-statistic: 31.81 on 3 and 64 DF, and p-value: 1.045e-12. The changes in coefficients are slight. Mostly just the linear trend has changed.

Table 2: Coefficients for the model predicting the environmental function from macroeconomic indicators.

Dependent Variable	Lag	Window	Estimate	Std. Error	z value	$Pr(>\|z\|)$
(Intercept)			0.0675	0.0366	1.845	0.0696
t			0.000425	0.000551	0.770	0.444
Unemp.lwLogRatio	1	17	0.453	0.116	3.898	0.000235
IncPerCap.lwLogRatio	-1	15	-2.687	0.634	-4.24	7.3e-05

Figure 7 shows the detrended environment function and revised macroeconomic fit. These two models (before and after detrending) are statistically equivalent, but we prefect the second one because it agrees better with the notion of a through-the-cycle PD.

Using the detrended environment function, the lifecycle and vintage functions were re-estimated. Visually the change was slight, but the PD model will not be stable through long-range forecasts.

3.4 Scoring

Using the retrended lifecycle and environment functions to compute the monthly populations odds as a function of the age of the loan and macroeconomic environment, we then estimated an origination score using typical attributes. The scoring function was created to predict defaults as a function of variables are observed at loan origination: FICO score at origination, log of the initial loan-to-value (LTV), subcategory (New or Used), log of the total balance on deposit by the borrower with the lender at origination, and term of the loan. LTV is the ratio of amount borrowed to the value of the vehicle at time of loan origination, and may include roll-over balances from a previous auto-loan such that LTV may be up to 130%.

The coefficients of this AVT origination model fit via a generalized linear model (GLM) with a binomial family are given in Table 3.

The factors in the model are the usual ones with reasonable coefficients and sensitivities. New or Used is a discrete variable, so the coefficient for New is

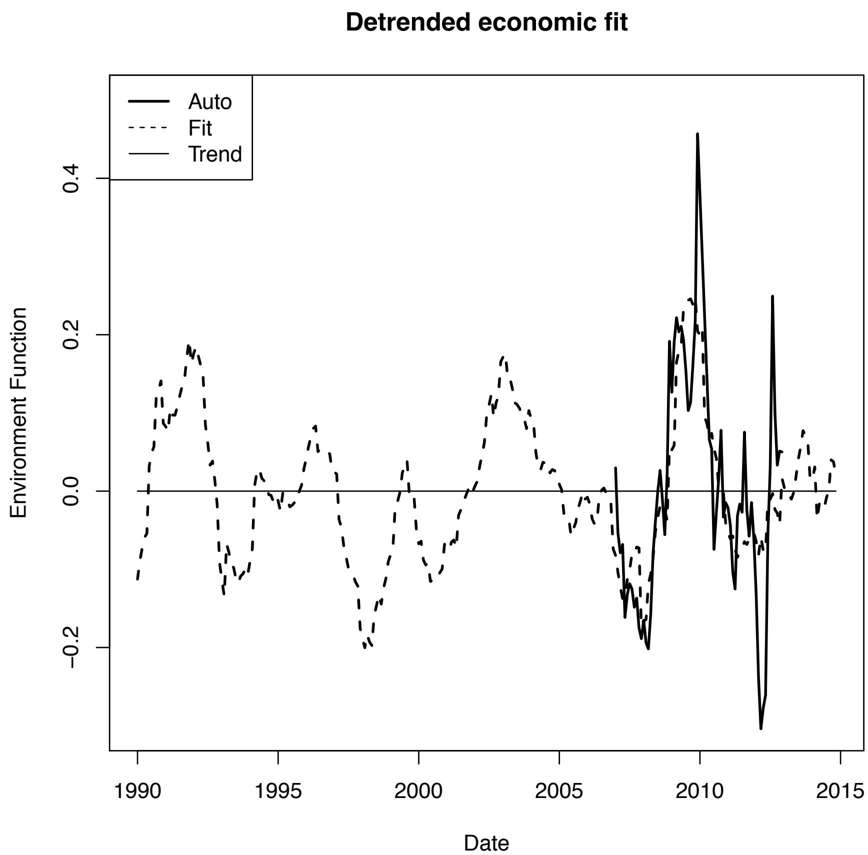

Figure 7: A comparison of the detrended environment function and the macroeconomic fit to that function. The linear trend overlay was fit to the macroeconomic model extrapolation.

Table 3: Coefficients for the origination score fit using lifecycle and macroeconomic impacts as fixed offsets (population odds).

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
(Intercept)	10.3	1.4	7.538	4.76e-14
FICO Score	-0.0129	0.0018	-7.043	1.88e-12
log(LTV)	1.20	0.55	2.201	0.0277
SubcategoryNew	0			
SubcategoryUsed	0.134	0.31	0.429	0.6680
Log Deposit Balance	-0.919	0.19	-4.888	1.02e-06
Term	-0.0173	0.0090	-1.937	0.0528

assigned to be zero as the reference level. The New / Used distinction shows the correct relationship whereby used car loans are expected to be higher risk than comparable new car loans. Nevertheless, the estimated coefficient appears to be insignificant. This is assumed to be due to the small dataset, since we know from experience that this distinction is important.

This model does not yet incorporate a random effects term for the borrower. Recent research [11] suggests that including a random effects term can be useful in cases where loans are observed on successive months as was the case here.

3.5 Validation

Since the model contains elements of scoring and time series modeling, it makes sense to validate using both scoring and time series validation methods.

3.5.1 Score Validation

The origination model was validated by computing ROC curves, K-S, and Gini coefficients. For each of the tests below, the model was completely retrained on the provided subset (lifecycle, environment, macroeconomic fit, and scoring coefficients were all re-estimated) and tested on the hold-out data.

For an in-sample / out-of-sample test, the data was split in half randomly by loan. Figure 8 gives the in-sample / out-of-sample comparison of the model as defined above. The model shows no measurable degradation when tested out-of-sample but over the same range of vintage and time.

For an in-time / out-of-time test, the last two years of the data were held out. Over the out-of-time period, the population odds (offset) were computed using the estimated lifecycle function and the macroeconomic model applied to actual macroeconomic data for that period. No statistically significant degradation occurs for the out-of-time test, Figure 9.

Lastly, we compare the AVT origination model to a model created with exactly the same data and structure, but excluding the offset for the population odds. Generally speaking, we do not anticipate a significant improvement in-sample from including the population odds, since the overall population odds for the period will appear in the model's intercept. Figure 10 actually shows that the AVT origination model does output-perform the simpler model by a small amount.

Comparing these two models via an out-of-time test is the ultimate goal of this paper. By including the expected future population odds, the AVT origination model should out-perform the simpler model without the offset. This is in fact what we observe, Figure 11, to within the granularity of the data.

3.5.2 Time series Validaton

For time series validation, we performed a typical in-time and out-of-time comparisons of the forecast to actuals. In Figure 12 the model was trained on data

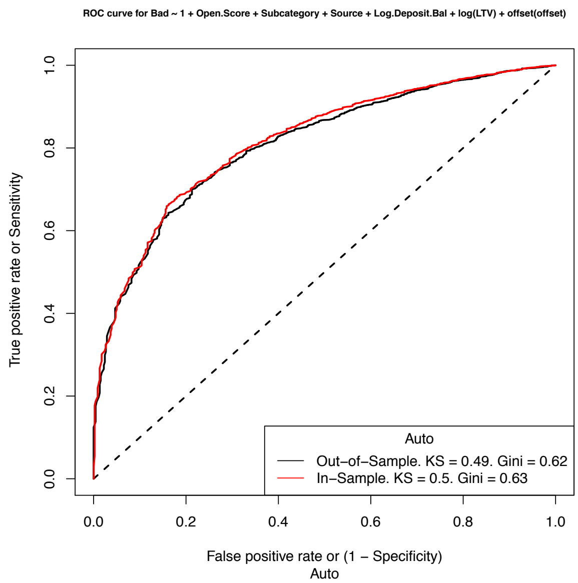

Figure 8: The AVT origination model was trained on half the data randomly sampled by loan and tested on the excluded half.

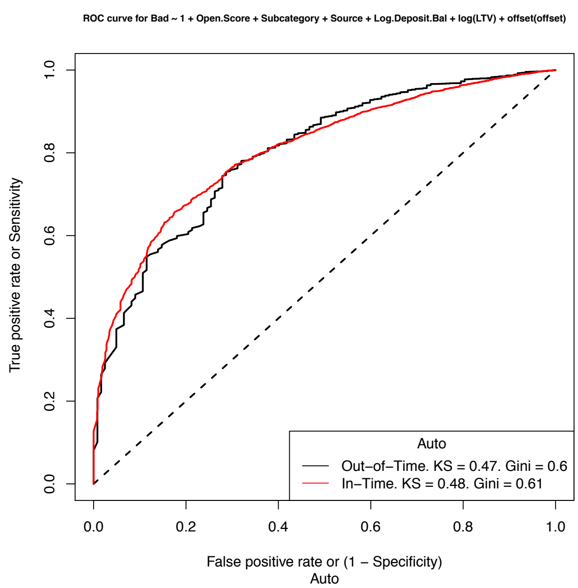

Figure 9: The AVT origination model was trained on data from 2004 through 2010 and tested on 2011 through 2012.

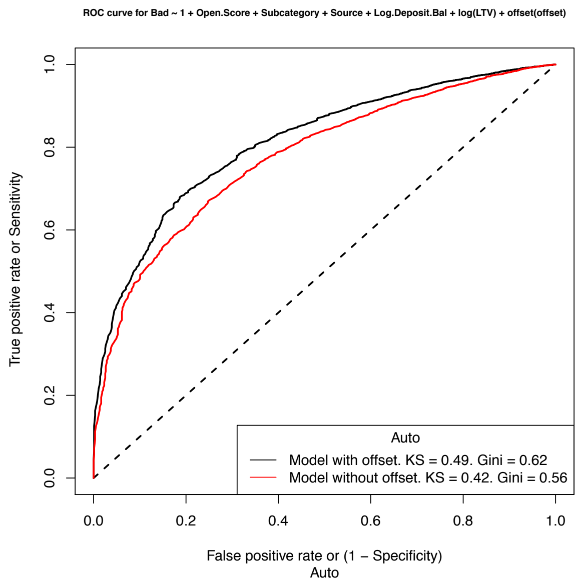

Figure 10: In-sample comparison of the AVT origination model to a similar model that excludes the offset (population odds).

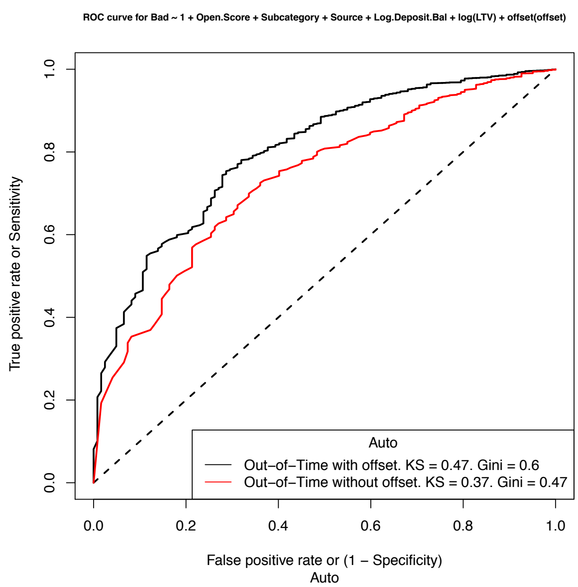

Figure 11: Out-of-time comparison of the AVT origination model and the simpler model excluding the offset (population odds).

through Dec-2012. Creating forecasts from the model produces a loan-level distribution of PD for each month. The mean of the monthly PD distribution is compared to the observed default rate. In addition, the forecasts were created for the period Jan-2013 through Dec-2013 using the actual values of the macroeconomic environmental variables observed during that period. Any new accounts originated during this period were also included in the forecast. This kind of test where the real macroeconomic values and new bookings are provided as inputs is referred to as Ideal Scenario Validation.

As shown in the graph, the model fit well both in-time and out-of-time. The in-time mean error was 0.24% and the RMS error divided by the mean PD was 0.0004%. Out-of-time the mean error was -3.9%, apparently due to an anomalous burst of defaults in May-2013. The normalized RMS error was 0.0005%.

The standard deviation of the forecast distribution is also shown in the plot to provide a sense of the amount of credit risk variation across the portfolio through time.

3.6 Stress Tests

One of the goals in creating this model was to produce a loan-level stress test model. Using only the active accounts at the end of the historic data, Dec-2013, forecasts are created for the following two years under different macroeconomic scenarios. The account attributes are held fixed, but the age of the loans is incremented each month and a new environment is considered.

To demonstrate this, a severe macroeconomic scenario was created whereby unemployment rises significantly and income per capita falls.

A baseline scenario was created where no changes occur in either economic factor from Dec-2014. The adverse scenario was created midway between the severe and baseline scenarios.

By applying the model in Table 2 to the three economic scenarios in Figure 13, three scenarios for the environment functions are obtained, as shown in Figure 14.

The scenarios for the environment are then combined with the lifecycle and credit scoring effects to create forecasts for the distribution of accounts at each month, Figure 15.

Since no new loans were included in the forecast, the average age of the portfolio increases rapidly. Consequently, the lifecycle component of the forecast $F(a)$ causes a dramatic increase in the PD for the portfolio overall. This replicates the well know effect that the blended loss rate rises when no new loans are originated.

On top of the lifecycle-driven rise in PD, we also see a spread of outcomes caused by the divergence in the three macroeconomic scenarios. Although the graph only shows the mean of the distribution, each account is being separately forecasted. Since they are at different ages prior to the start of the forecast, this is not just a parallel shift of the distribution. Forecasts for individual loans will cross, but the mean of the distribution follows the expected patterns.

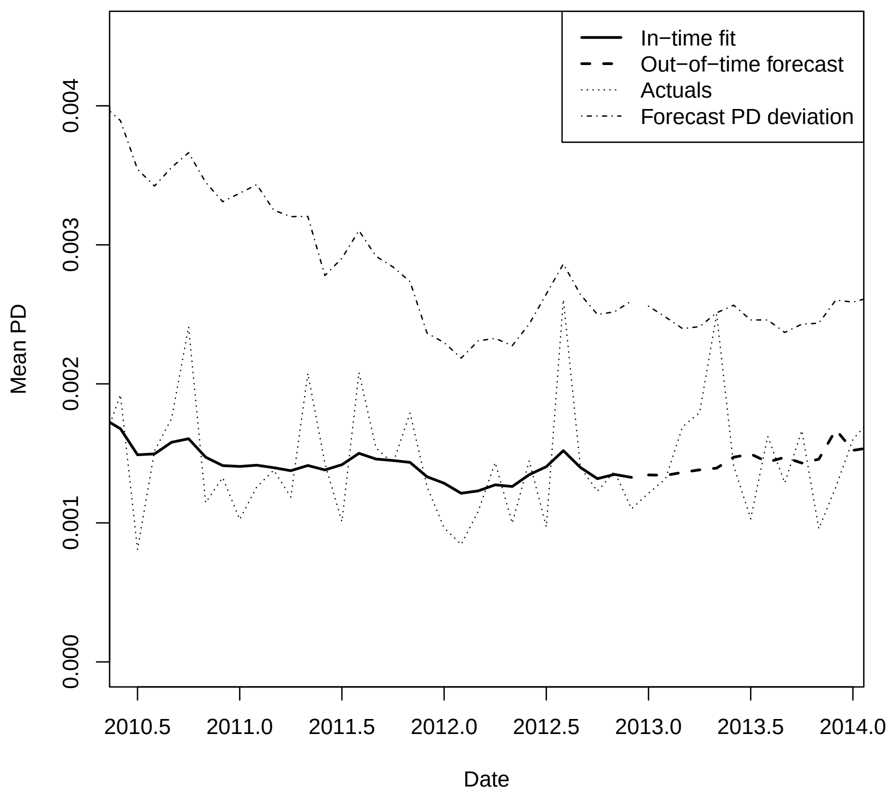

Figure 12: A comparison of the mean of the distribution of the monthly PD forecasts to the average monthly default rate. The standard deviation of the monthly forecast distribution is also shown for comparison.

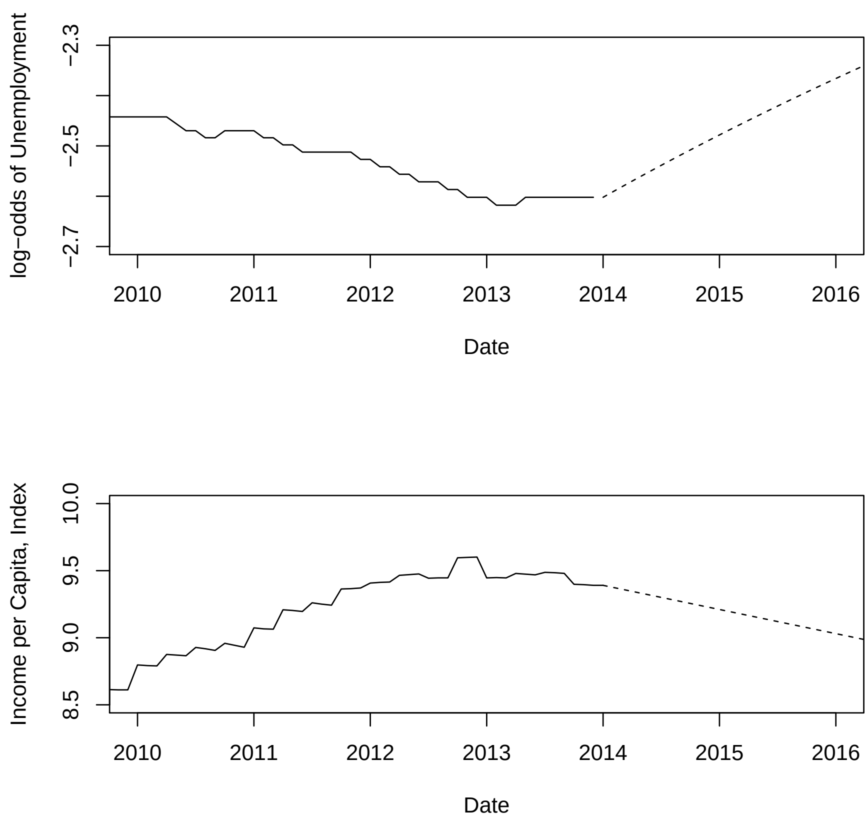

Figure 13: The history (solid line) and scenario (dashed line) for the log-odds of unemployment rate and income per capita.

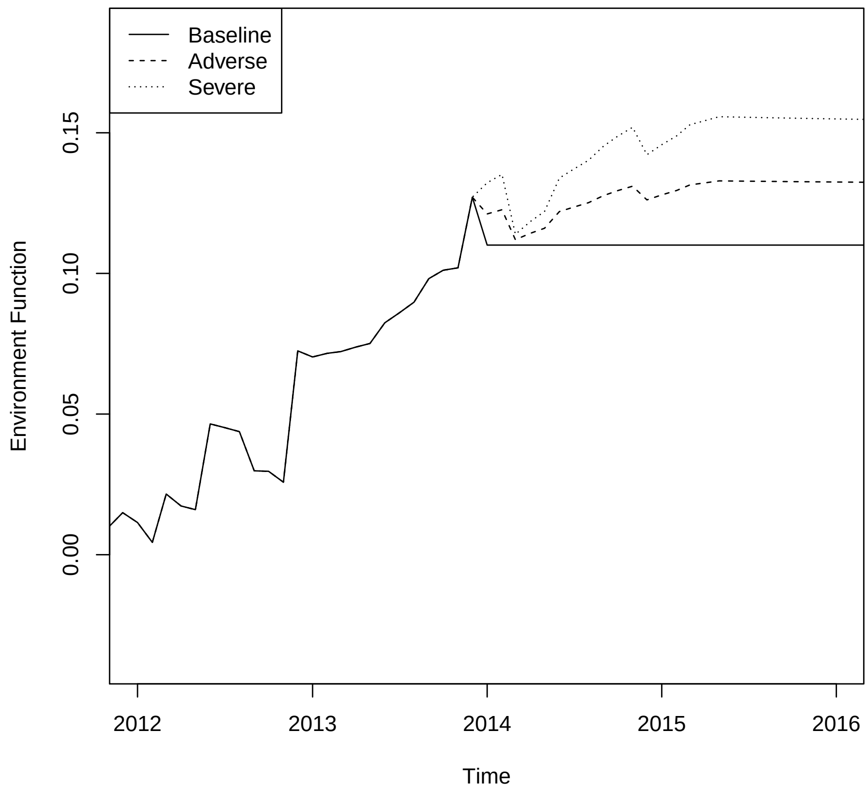

Figure 14: The scenarios for the enviornment function $H(t)$ resulting from applying the macroeconomic model from Table 2 to the three economic scenarios.

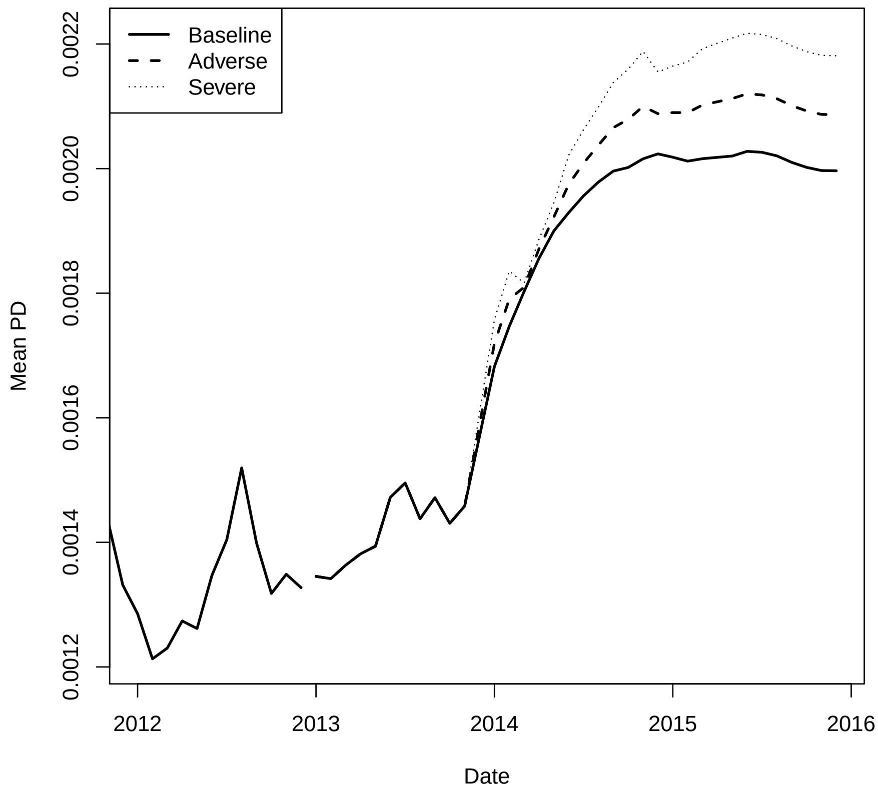

Figure 15: The mean of the forecast distribution each month given the environment scenarios in Figure 14.

In this way, we obtain loan-level, monthly forecasts and stress tests which can be combined in a competing risks framework, loss forecasting system, or lifetime expected profit estimation.

4 Conclusions

The analysis shown here demonstrates that we can build loan-level models that are at least as good as conventional scores for rank-ordering, but can also be applied to forecasting and stress testing. By including the retrending adjustment, we guarantee that the model produced will be stationary when extrapolated forward or backward, meaning that is can also be used reliably to predict through-the-cycle PD estimates simply by providing an average environmental function or macroeconomic scenario.

The AVT origination model created here also has the advantage of being able to incorporate performance data from the most recent vintages. A standard approach of taking loans originated three years ago and monitoring their performance over the next two years to create a score leaves a model that is at least three years out-of-date as soon as it is applied out-of-time. For the AVT origination model, any loan at least old enough to experience defaults is incorporated in the training data.

Many practitioners try to overcome the above lag by using shorter observation periods and continuously rebuilding the model. The out-of-time tests of the AVT origination model show that it's performance did not degrade measurably two years into the future. We assume that a retraining will eventually be required, but this remarkable stability comes from the separation of macroeconomic conditions from scoring attributes. By avoiding confusing the two, the origination score is much more robust.

Logistic regression panel data models are probably the most commonly employed alternative to the approach described here. A regression model could be created with scoring and macroeconomic attributes, dummy variables or basis functions for the lifecycle, and even additional dummy variables for vintage adverse selection. (Use of dummy variables in addition to scoring attributes in a model framework similar to that developed in this paper is described Breeden and Canals-Cerda 2016 [3].) However, the unconstrained single-step approach does not solve the linear specification error. Since attributes are used instead of APC-style basis functions, a solution is obtained, but multicolinearity exists between the various attributes of scoring, economics, and lifecycle. Therefore, the out-of-time extrapolation properties are uncontrolled and prone to extreme answers.

Although the two-step process developed here may seem less elegant than a single-step model, it has the effect of imposing the necessary constraint on the regression so that the out-of-time performance is robust. This is the greatest advancement of the current approach relative to standard logistic regression models.

In addition, if behavior score variables such as delinquency or utilization are included among the scoring attributes, i.e. attributes whose values change with time, the multicolinearity problem becomes more severe. Most practitioners would recognize that for proper use out-of-time, the behavior scoring attributes would need to be predicted in additional to running the overall model with macroeconomic scenarios. Because the current approach concentrates the macroeconomic sensitivity first in the macroeconomic factors, this constraint serves to solve the multicolinearity problem. The behavior scoring attributes act as adjustment factors centered around the known macroeconomic drivers. Therefore, when used out-of-time, a behavioral PD model created via the current approach does not need to forecast the behavioral attributes. This provides a a dramatic improvement over standard approaches.

Lastly, this model has the benefit that it is a PD model from the start. With traditional scores, most practitioners attempt to fit a score-odds calibration after the model has been created. That process risks confounding the macroeconomic environment with the scoring attributes and has shown limited effectiveness. By explicitly capturing the population odds throughout the creation of the AVT origination model, the PD forecast is as robust as the rank ordering.

References

[1] Capital planning at large bank holding companies: Supervisory expectations and range of current practice. Technical report, Board of Governors of the Federal Reserve System, August 2013.

[2] IFRS 9 financial instruments. Technical report, International Financial Reporting Standards Foundation, July 2014.

[3] Joseph Breeden and Joseè J. Canals-Cerdaá. Consumer risk appetite, the credit cycle, and the housing bubble. Technical report, Federal Reserve Bank of Philadelphia, Research Department, February 2016.

[4] Joseph L. Breeden. *Reinventing Retail Lending Analytics: Forecasting, Stress Testing, Capital and Scoring for a World of Crises.* Risk Books, London, 2010.

[5] Joseph L. Breeden. Testing retail lending models for missing cross-terms. *Journal of Risk Model Validation*, 4(4), 2010.

[6] Joseph L. Breeden, Anthony Bellotti, Aleh Yablonski, and Eugenia Leonova. Instabilities using cox ph with age-vintage-time dimensions. In *Festschrift Conference in Honour of Lyn Thomas*, Southampton, April 22 2016.

[7] 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.

[8] Joseph L. Breeden and Lyn C. Thomas. Solutions to specification errors in stress testing models. *Journal of the Operational Research Society*, 67:830–840, December 2016.

[9] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187.

[10] D. R. Cox and D. O. Oakes. *Analysis of Survival Data*. Chapman and Hall, London, 1984.

[11] J. Crook and T. Bellotti. Asset correlations for credit card defaults. *Applied Financial Economics*, 22:87–95, 2012.

[12] Bradley Efron. The two-way proportional hazards model. *Journal of the Royal Statistical Society B*, 64:899 – 909, 2002.

[13] Financial Accounting Standards Board. Financial InstrumentsCredit Losses (Subtopic 825-15). *Financial Accounting Series*, December 20 2012.

[14] Norval D. Glenn. *Cohort Analysis, 2nd Edition*. Sage, London, 2005.

[15] T R Holford. The estimation of age, period and cohort effects for vital rates. *Biometrics*, 39:311–324, 1983.

[16] David W. Hosmer, Jr. and Stanley Lemeshow. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. Wiley Series in Probability and Statistics, New York, 1999.

[17] George Judge, William E. Griffiths, R. Carter Hill, Helmut Ltkepohl, and Tsoung-Chao Lee. *The Theory and Practice of Econometrics*. Wiley Publ., 1985.

[18] W.M. Mason and S. Fienberg. *Cohort Analysis in Social Research: Beyond the Identification Problem*. Springer, 1985.

[19] Volker J. Schmid and Leonhard Held. Bayesian age-period-cohort modeling and prediction BAMP. *Journal of Statistical Software*, 21.

[20] Terry M. Therneau and Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York, 2000.

[21] Lyn C. Thomas. *Consumer Credit Models: Pricing, Profit and Portfolios*. Oxford University Press, 2009.

References