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
July 21, 2014

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, $x$, to predict the probability of “good”, $p(G|x)$. The odds of a loan being good are

$$o(G|x) = \frac{p(G|x)}{p(B|x)} = \frac{p_G}{p_B} \frac{p(x|G)}{p(x|B)} = o_{pop} \times I(x), x \in X$$  \hspace{1cm} (1)

or

$$\log(o(G|x)) = \log(o_{pop}) + \log(I(x))$$ \hspace{1cm} (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 $I(x) = \frac{p(x|G)}{p(x|B)}$ is the information odds. We could also say that $o_{pop}$ captures systematic effects for the portfolio and $I(x)$ 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 $o(G|x_{out})$ where $x_{out}$ 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 $x$ 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 [10, 7] and applied to the context of credit risk modeling [3], 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 [4, 8, 11, 6] 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 mean that we could not reliably predict the population odds in future time periods, bringing the effectiveness of any stress test model into question.

Breeden and Thomas [3] 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 \( \text{slope}(\hat{E}(t)) = 0 \) 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 assumption of zero 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.

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.

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

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

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 nonparametric 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.

\[ \text{logit}(p_i(a, v, t)) = F(a) + G(v) + H(t) \]  

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.

2.2 Macroeconomic fit

The environmental function \( H(t) \) is initially estimated with the assumption of no net trend with time over the observed portfolio data. As the data is fit to macroeconomic data, the no-trend constraint is relaxed 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) [1]. For changes in HPI, we should use a log-ratio transformation, 

\[
\log - \text{ratio}(\text{HPI}) = \log(\text{HPI}(t)/\text{HPI}(t-w))
\]

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 [9]. 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,

\[
\hat{H}(t) = c_0 t + \sum_i c_i E_i(t) + \epsilon_t,
\]

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, \( t \in [0, T_0] \). The process of fitting the environmental function to macroeconomic data has been previously demonstrated for APC-class models [2, 3].

2.3 Retrending the functions

Using the technique of Breeden and Thomas [3], 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

\[
F'(a) = F(a) + \beta a \\
G'(v) = G(v) + \beta v \\
H'(t) = H(t) - \beta t
\]  

(6)

Retrending can also be performed by rerunning 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. Many forecast or stress test models 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 offset \( \text{offset}(a, t) = F'(a) + H'(t) \) 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 \( X \),

\[
\logit(p_i(a, t)) = \text{offset}(a, t) + B X,
\]

(7)

where \( B \) 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 2004 through 2012 for all vintages during that period. The monthly default rate was modeled

\[
DR(a, v, t) = \frac{\text{Default Accounts}(t)}{\text{Active Accounts}(t - 1)}
\]

(8)

The steps described in Section 2 were followed.
3.1 AVT 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.

![Lifecycle function with account age measured for a small US auto loan portfolio](image)

Figure 1: Lifecycle function with account age measured for a small US auto loan portfolio

3.2 Macroeconomic Fit

The environment function was compared to a range to macroeconomic indicators. 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.

The values in Table 1 produce a model with multiple R-squared: 0.3918, adjusted R-squared: 0.3848, F-statistic: 56.52 on 4 and 351 DF, and p-value: \( \times 2.2e-16 \).

3.3 Retrending

The fit to macroeconomic data is shown in Figure 4 as the black line between 2005 and 2013. The older part of the black line is the backward extrapolation over previous recessions. This extrapolation is implausible and fully explainable
Figure 2: Credit risk function with vintage measured for a small US auto loan portfolio

Figure 3: Environmental function with time measured for a small US auto loan portfolio
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)        | 5.91| 1.2    | 4.727    | 3.31e-06   |         |          |
| t                  | -0.00955 | 0.0025 | -3.765   | 0.000195   |         |          |
| Unemployment Moving Avg | -6 | 3 | 2.05 | 0.54 | 3.828 | 0.000153 |
| House Price Index Log Ratio | 5 | 24 | -2.26 | 1.1 | -2.177 | 0.030153 |
| Housing Starts Log Ratio | 8 | 15 | -0.307 | 0.048 | -6.440 | 3.94e-10 |

from the trend ambiguity described earlier. The red line in Figure 4 shows the detrended result after fitting a straight line to the original fit (black line) and removing. The lifecycle and credit risk functions were adjusted to compensate for the detrending of the environmental function.

Figure 4: A comparison of the fit to the environment function using macroeconomic data (black line) and the detrended version (red line).

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 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, and term of the loan. The coefficients of this AVT origination model fit via a generalized linear model (GLM) with a binomial family are given in Table 2.

Table 2: 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 |
| Open Score | -0.0129 | 0.0018 | -7.043 | 1.88e-12 |
| log(LTV) | 1.20       | 0.55    | 2.201 | 0.0277 |
| SubcategoryNew | 0.134 | 0.31 | -0.429 | 0.6680 |
| SubcategoryUsed | -0.919 | 0.19 | -4.888 | 1.02e-06 |
| Log Deposit Balance | -0.0173 | 0.0090 | -1.937 | 0.0528 |

The factors in the model are the usual ones with reasonable coefficients and sensitivities. The New / Used distinction shows the correct relationship, but 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 [5] 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

The AVT 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 5 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 6.

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 7 actually shows that
Figure 5: The AVT origination model was trained on half the data randomly sampled by loan and tested on the excluded half.

Figure 6: The AVT origination model was trained on data from 2004 through 2010 and tested on 2011 through 2012.
the AVT origination model does output-perform the simpler model by a small amount.

Figure 7: In-sample comparison of the AVT origination model to a similar model that excludes the offset (population odds).

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 8, to within the granularity of the data.

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 its 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.

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


