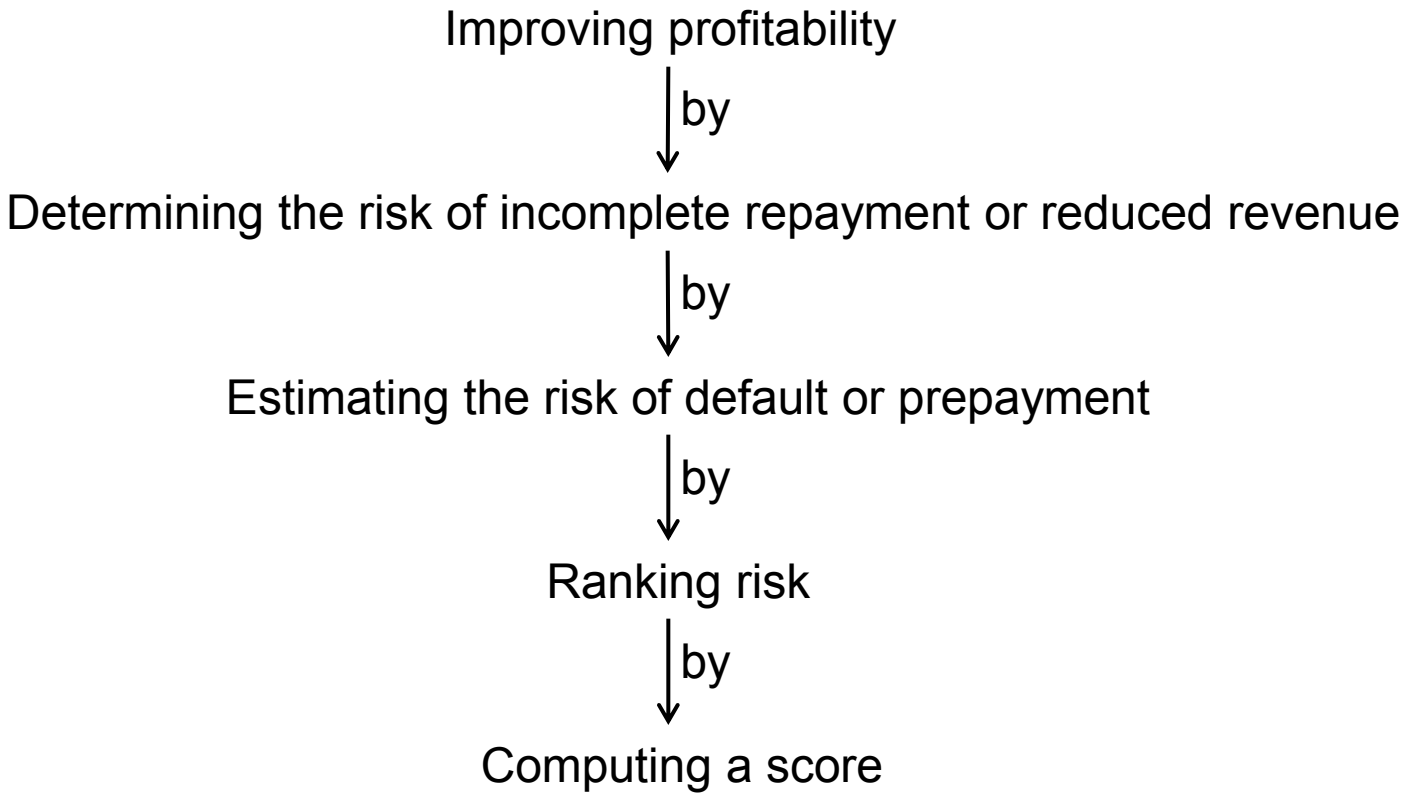
INCORPORATING LIFECYCLE AND ENVIRONMENT IN LOAN-LEVEL FORECASTS AND STRESS TESTS

Joseph L Breeden, PhD

Prescient Models

What are Scores for?



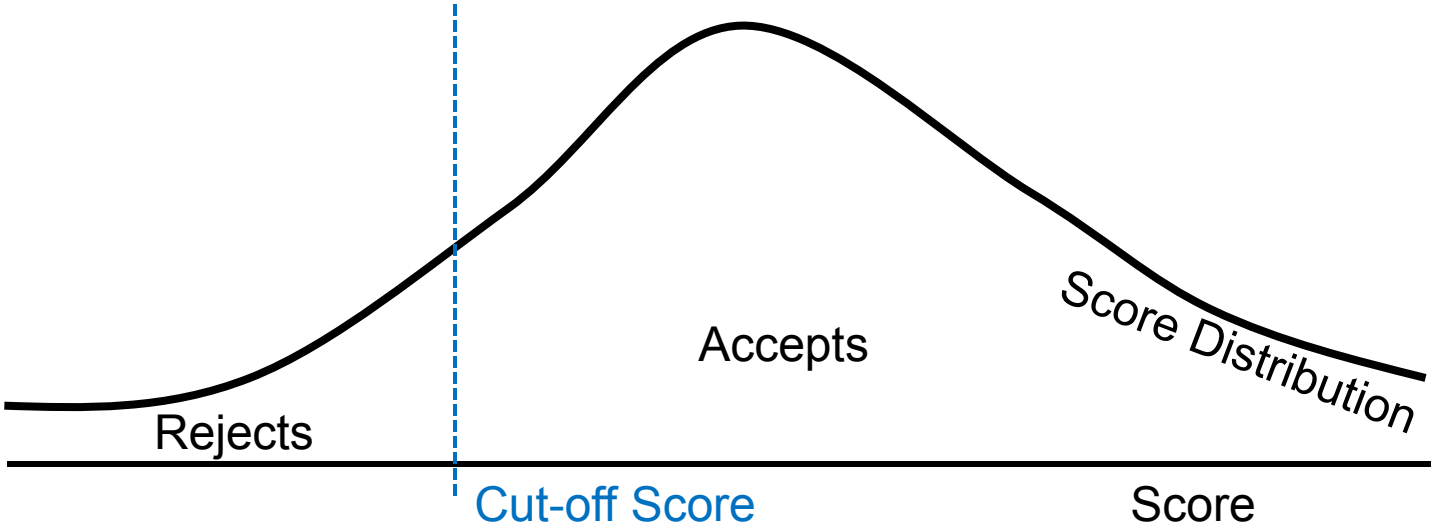
- This is the assumed hierarchy under which scores are created.

Profit Model

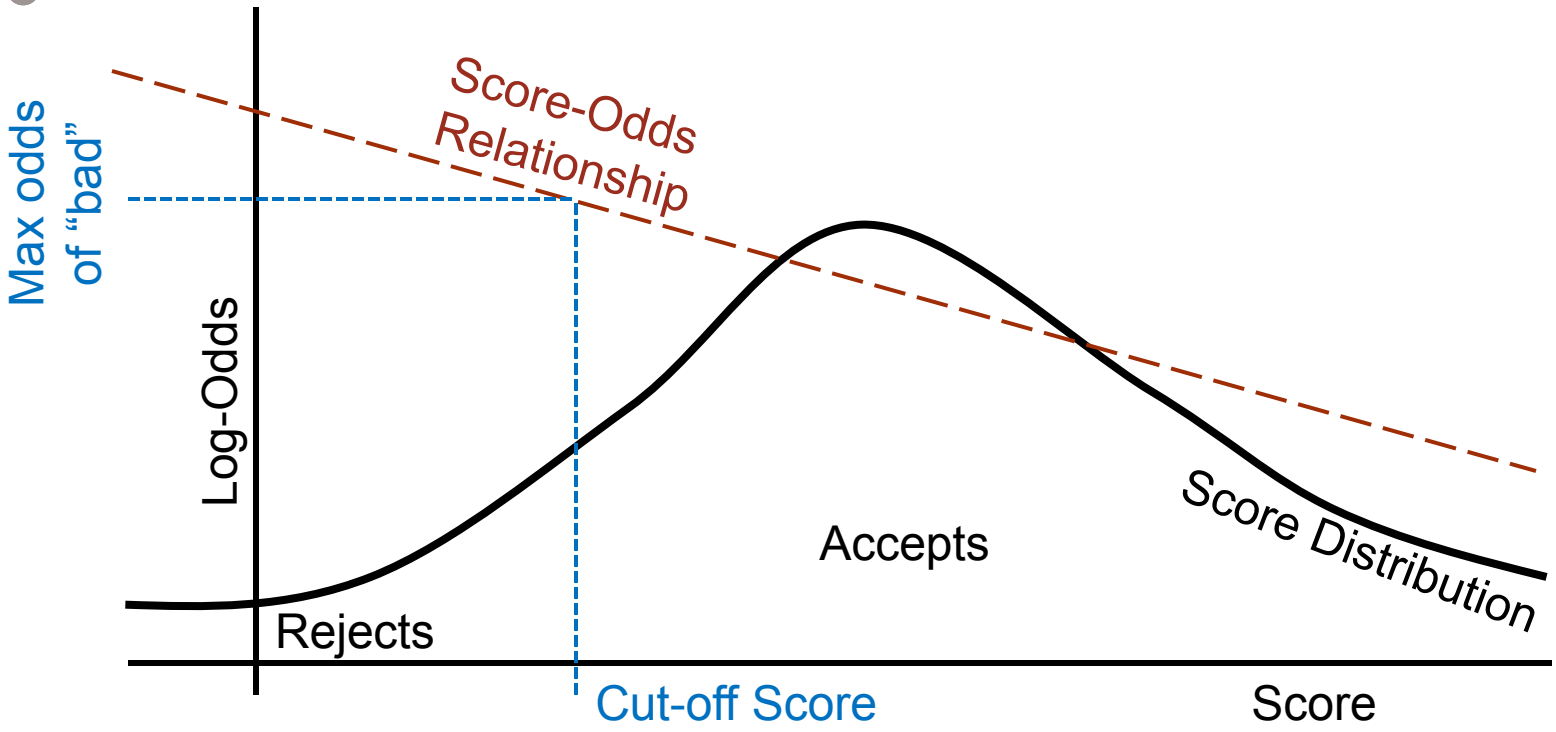
Cut-off Score	620	Max Loss	
Max PD	%	Exp. Revenue	
Loan Amt		Expenses	
EAD		Min Margin	
LGD			

- The connection between cut-off score and PD is tenuous.
- We can be much more accurate in pricing if we start with a real PD forecast.

Using Scores: Setting Pricing



Using Scores: Setting Pricing



- The score distribution shifts with the economy.
- The score-odds relationship shifts with the economic cycle and credit cycle.
- Cut-off score is usually adjusted intuitively.

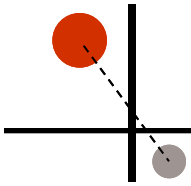
Standard Bank Practice

- Credit Risk has a group that builds scores that rank-order.
- Model Audit validates the scores rank-ordering ability.
- Credit Policy sets the cut-off score to control risk (odds of default).
- Finance creates a profit model that assumes a certain odds of default.

This structure fails by design.

What we needed from the start was a probability model.

Any score provides that in-sample, but how do we produce it out of sample?



MODELING PROBABILITIES

Prescient Models

Drivers of Performance

- *Origination:* Consumer & loan information.
- *Vintages:* Biases for specific cohorts.
- *Lifecycle:* Changes with age of the loan.
- *Behavior:* As the loan matures, observed loan performance can refine the predictions.
- *Account Management:* Changes made to individual loans.
- *Environment:* Changes in the economy or portfolio management.
- *Loan Idiosyncratic Effects*: Every loan is unique.

Age-Vintage-Time (A-V-T) Dependency

Information Type	…is a function of
Originations	Recent history or loan (i) up to origination date (v)
Vintage	Vintage origination date (v)
Lifecycle	Age of the loan (a)
Behavior	All time since origination (v, v+1, … , t=v+a)
Account Management	Time since event (t_e)
Environment	Recent time (t-n, … t-1, t)
Loan Idiosyntricities	Loan (i)

- Because $a = t - v$, measurements from one information source will be contaminated by other sources:
  - Behavior is dependent on Lifecycle and Environment.
  - The linear components of any vintage, age, or time dependencies cannot be uniquely measure.

Modeling Implications

- From the Age-Period-Cohort literature, we know that *linear* effects in age, vintage, and time cannot be uniquely determined regardless of the model used.
- For short data sets (such as a typical 2-year scoring dataset), the macroeconomic environment by time and credit risk by vintage are mostly linear trends.
- Coefficients estimated against the linear components of macroeconomic variables and scoring factors will be unstable, i.e. adding macroeconomic factors to scoring models rarely works.

GLM-AVT Model

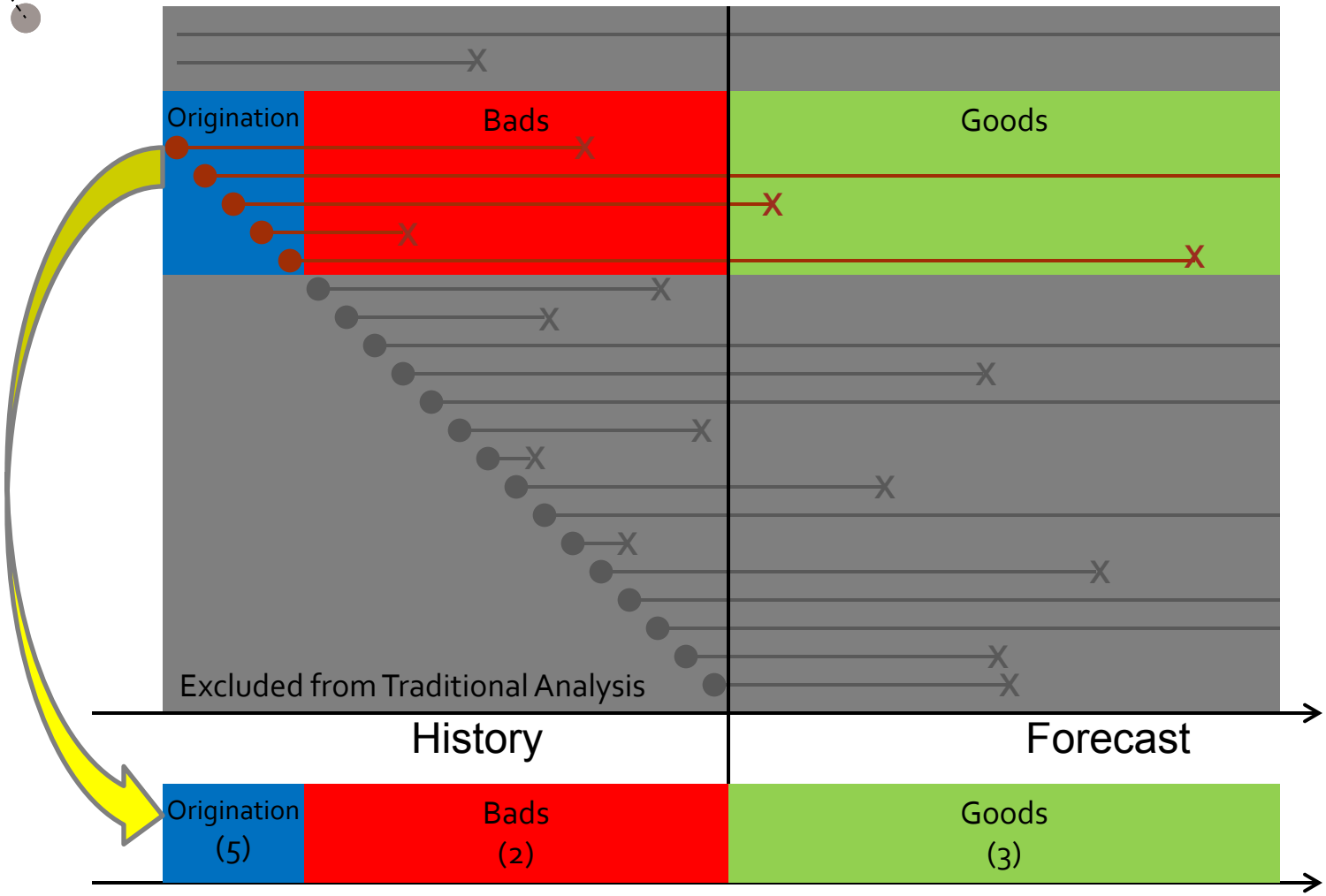
1. Use the Age-Vintage-Time structure found in APC, but in a loan-level model.

$$\log\left[\frac{p_i(a,v,t)}{1-p_i(a,v,t)}\right] = f(a)+g(v)+h(t), \quad a = t - v$$

Make an explicit choice about trend allocation between $f(a)$, $g(v)$, and $h(t)$ to obtain a unique solution.

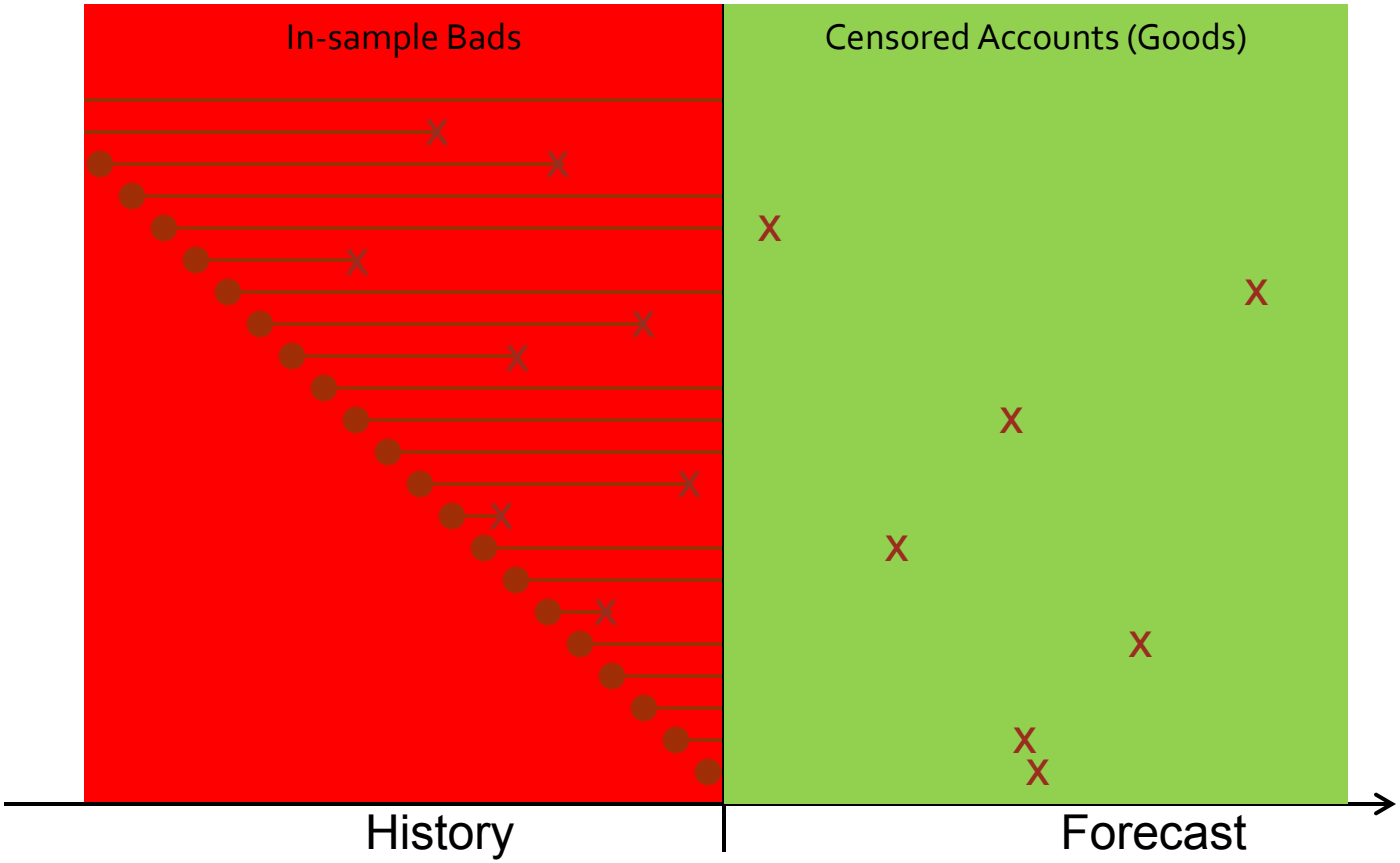
1. (Optional) Fit $h(t)$ to macroeconomic data. Retrend lifecycle and credit risk trends to compensate.
2. Introduce either origination or behavioral scoring factors to estimate score.

Data for Loan-level Modeling



- A standard approach to defining Good and Bad accounts for logistic-style regression without inclusion of temporal information.

Data for V-A-T Decomposition Methods



- All in-sample loans are included in analysis with emphasis on time of Event occurrence.
- Goods are simply accounts that were not observed to go bad. They are included as censored accounts in the estimation process.

1. Create Loan-level Age-Vintage-Time Model

Put an APC structure into a loan-level GLMM framework.

$$\log\left[\frac{p_i(a,v,t)}{1-p_i(a,v,t)}\right] = \alpha_0 + \alpha_1 a + \sum_{j=2}^{N_a-1} \alpha_j O_j(a) + \beta_1 v + \sum_{j=2}^{N_v-1} \beta_j O_j(v) + \sum_{j=2}^{N_t-1} \gamma_j O_j(t)$$

- 1 constant term, 2 linear terms, and nonlinear terms in a, v, and t. No explicit cross-terms.
- O_i are orthogonal basis functions.
- N_a, N_v, and N_t are the number of observations in a, v, and t.
- Assumes environmental trend = 0.

Create Age-Vintage-Time Functions

Recast as functions $F(a), G(v), H(t)$ as

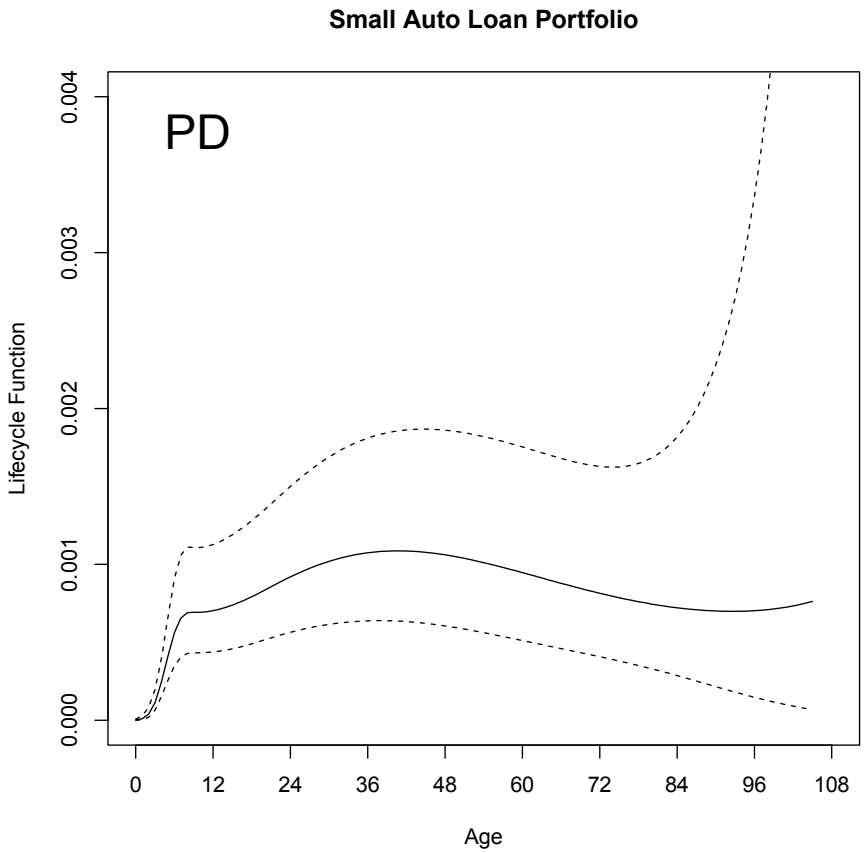
$$F(a) = \alpha_0 + \alpha_1 a + \sum_{j=2}^{N_a - 1} \alpha_j O_j(a)$$

$$G(v) = \beta_1 v + \sum_{j=2}^{N_v - 1} \beta_j O_j(v)$$

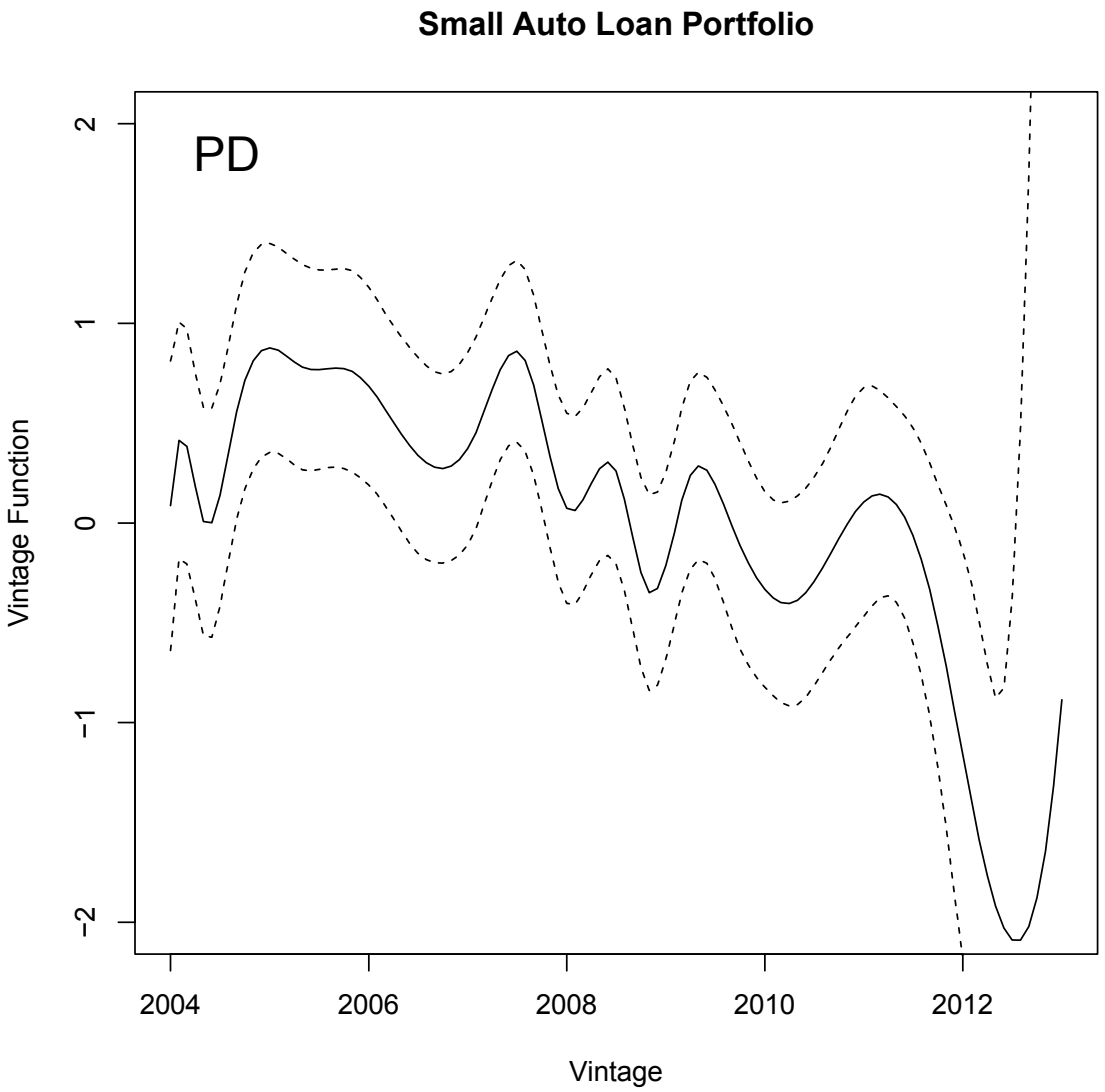
$$H(t) = \sum_{j=2}^{N_t - 1} \gamma_j O_j(t)$$

Example

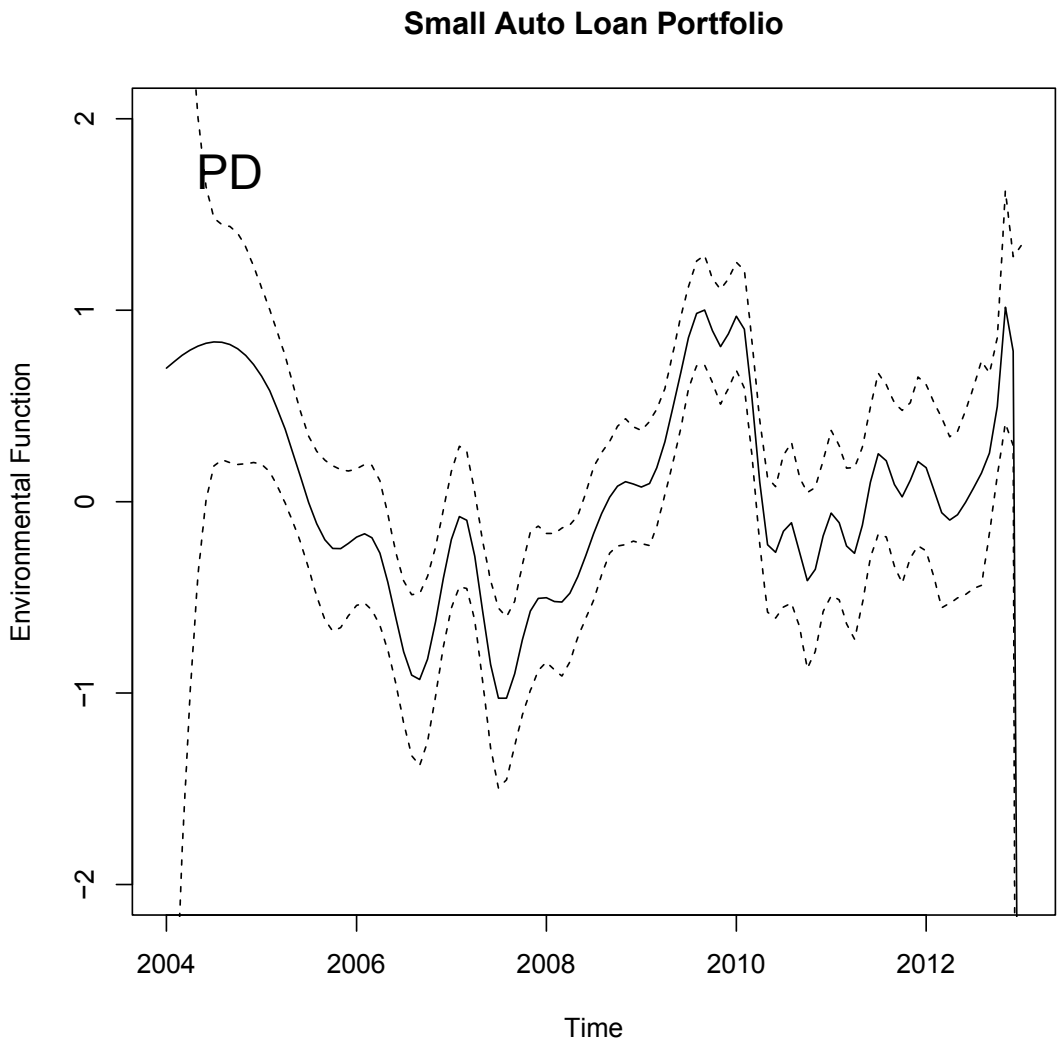
- Regional US auto loan portfolio, ~$400 mln receivables.
- We modeled PD, EAD, LGD, and attrition to predict losses.



Credit Risk Function Example



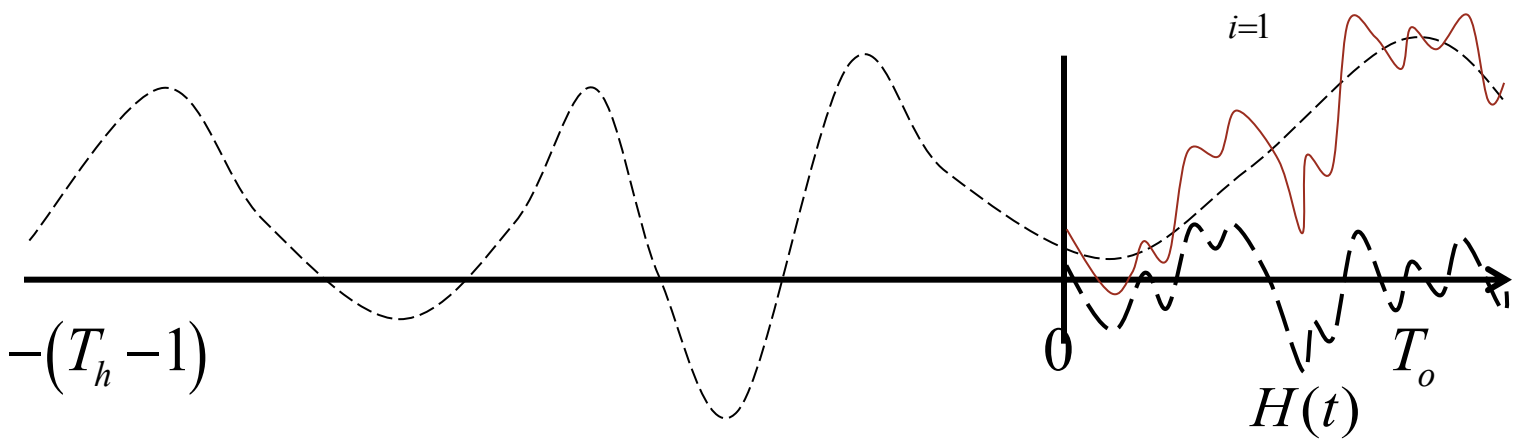
Environment Function Example



2. Fit to Macroeconomic Data

Create OLS fit of $H(t)$ to macroeconomic factors $E_i(t)$ with an explicit trend term. Possibly use a constrained fit so the long-run trend = 0.

$$H(t) = c_0 t + \sum_{i=1}^{N} c_i E_i(t) + \varepsilon$$



- We have T_0 observations of function $H(t)$.
- We have T_h measurements of n economic factors $E_i(t)$ prior to the start of the performance data.
- Total length of economic data is $T = T_0 + T_h$.

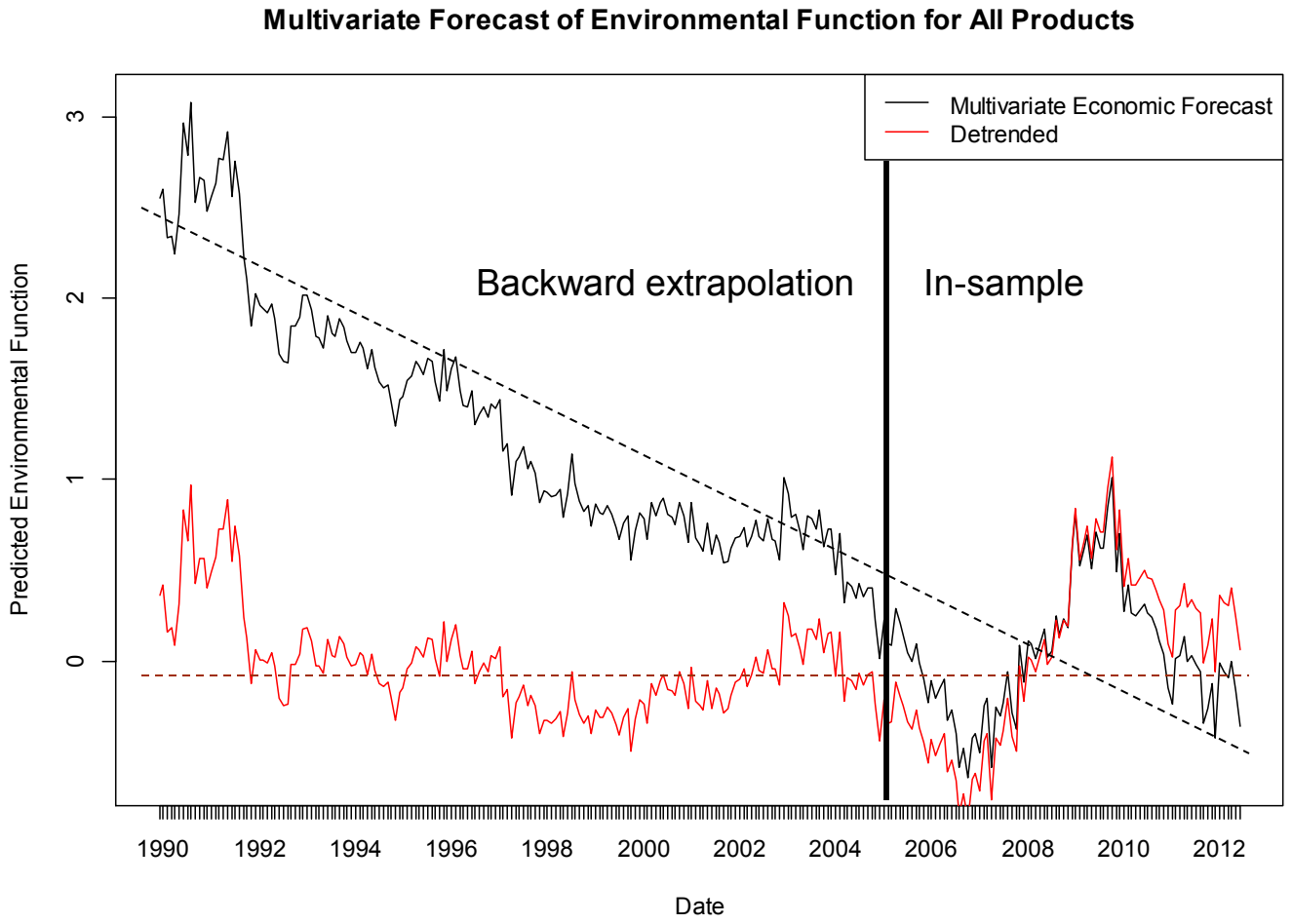
Create a macroeconomic model

- Model just the environmental function with macroeconomic factors. (Macroeconomic adverse selection impacts on vintage quality is a separate effect.)
- Use the correct transformations: log-ratio, not % change; log interest rates; etc.
- Allow for a linear component.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.908513	1.250051	4.727	3.31E-06
t	-0.009547	0.002535	-3.765	0.000195
Unemp.lwMovingAvg.Lm6.W3	2.050558	0.535645	3.828	0.000153
HPI.lwLogRatio.L5.W24	-2.262367	1.039257	-2.177	0.030153
Housing.Starts.lwLogRatio.L8.W15	-0.306789	0.047639	-6.44	3.94E-10

```
Residual standard error: 0.434 on 351 degrees of freedom
  (12 observations deleted due to missingness)
Multiple R-squared: 0.3918,Adjusted R-squared: 0.3848
F-statistic: 56.52 on 4 and 351 DF,  p-value: < 2.2e-16
```

Retrend the environmental function



Readjust Age-Vintage-Time Functions

Compensate for the trend in $H(t)$.

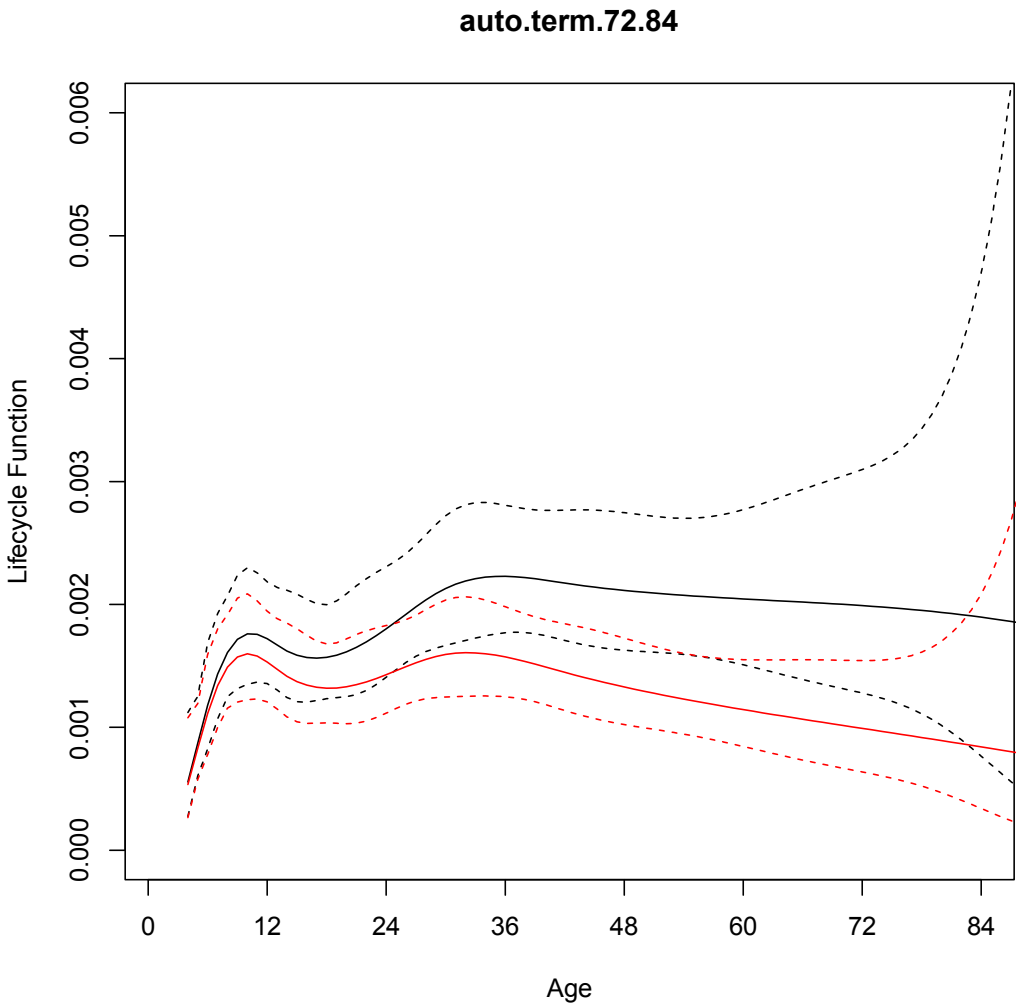
$$F'(a) = F(a) + c_0 t$$

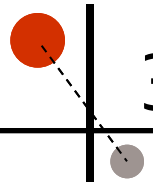
$$G'(v) = G(v) + c_0 t$$

$$H'(t) = H(t) - c_0 t = \sum_{i=1}^{N} c_i E_i(t) + \varepsilon$$

The linear specification problem is NOT unique to APC models. In fact, it appears to be the cause of scores constantly being recalibrated.

Counter-Adjustments to Lifecycle and Credit Risk





3. Fit the Score

- Using GLM and monthly performance data, use $F(a) + H(t)$ as a fixed offset.
- The score is estimated as
- `glm(Bad ~ 1 + offset(F+H) + factor1 + factor2 + …)`
- At this point, the factors can be either origination or behavior variables without destabilizing the probability model.

Score Example

- "Offset" captures the population odds. The rest of the model is the idiosyncratic variation – a score.

```
glm(formula = Bad ~ 1 + offset(offset) + Bureau.Score + log(LTV) +
    Subcategory + Log.Dep.Bal + Term, family = binomial, …)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.2726  -0.0276  -0.0133  -0.0042   4.7835

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     10.306961   1.367256   7.538 4.76e-14 ***
Bureau.Score    -0.012886   0.001830  -7.043 1.88e-12 ***
log(LTV)         1.203598   0.546832   2.201   0.0277 *
SubcategoryNew          0
SubcategoryUsed  0.133842   0.312062   0.429   0.6680
Log.Dep.Bal     -0.918796   0.187975  -4.888 1.02e-06 ***
Term            -0.017266   0.008915  -1.937   0.0528 .
---
```

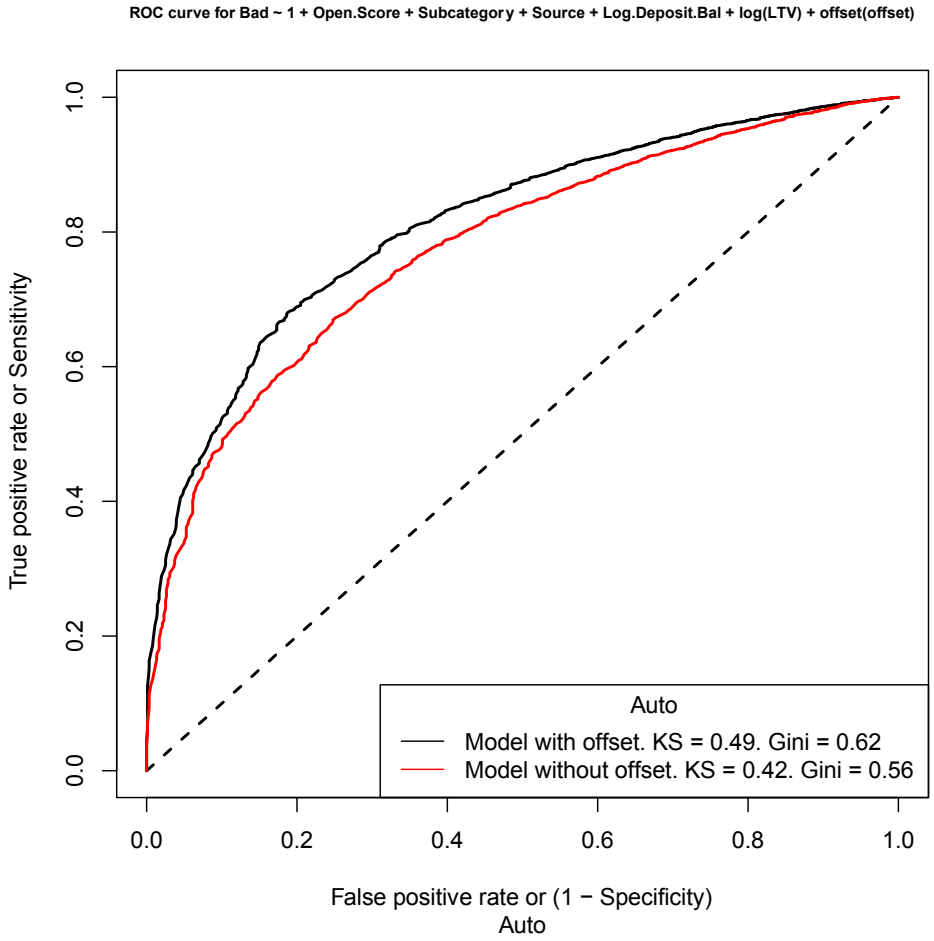
Outputs – The Score

- The glm-avt result is still a "score" if you exclude the offset term.
  - It can be scaled to match scores currently in use.
  - It will be more robust to changes in the economy – recalibrate much less often.
  - It can incorporate measured adverse selection for existing loans.
  - It can be built across both older and more recent data history to capture more consumer dynamics.

Outputs – The Probability

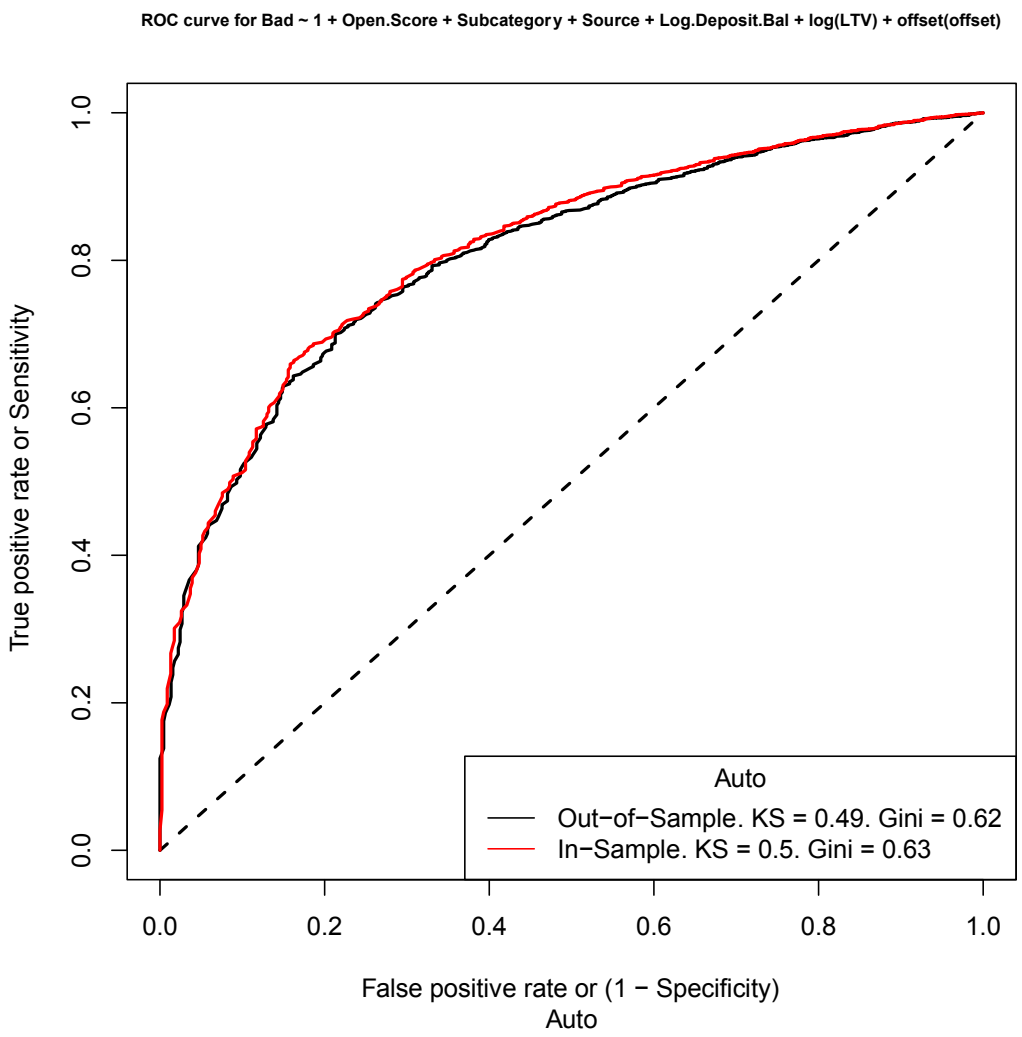
- p_i is the complete probability including the offset term to provide population odds over time.
  - It is conditional on age of the loan, "score", and scenario for the environment.
  - It is available for each month, arbitrarily far into the future.
  - It can be aggregated over any desired time window to predict total probability.

Scoring with and without Offset

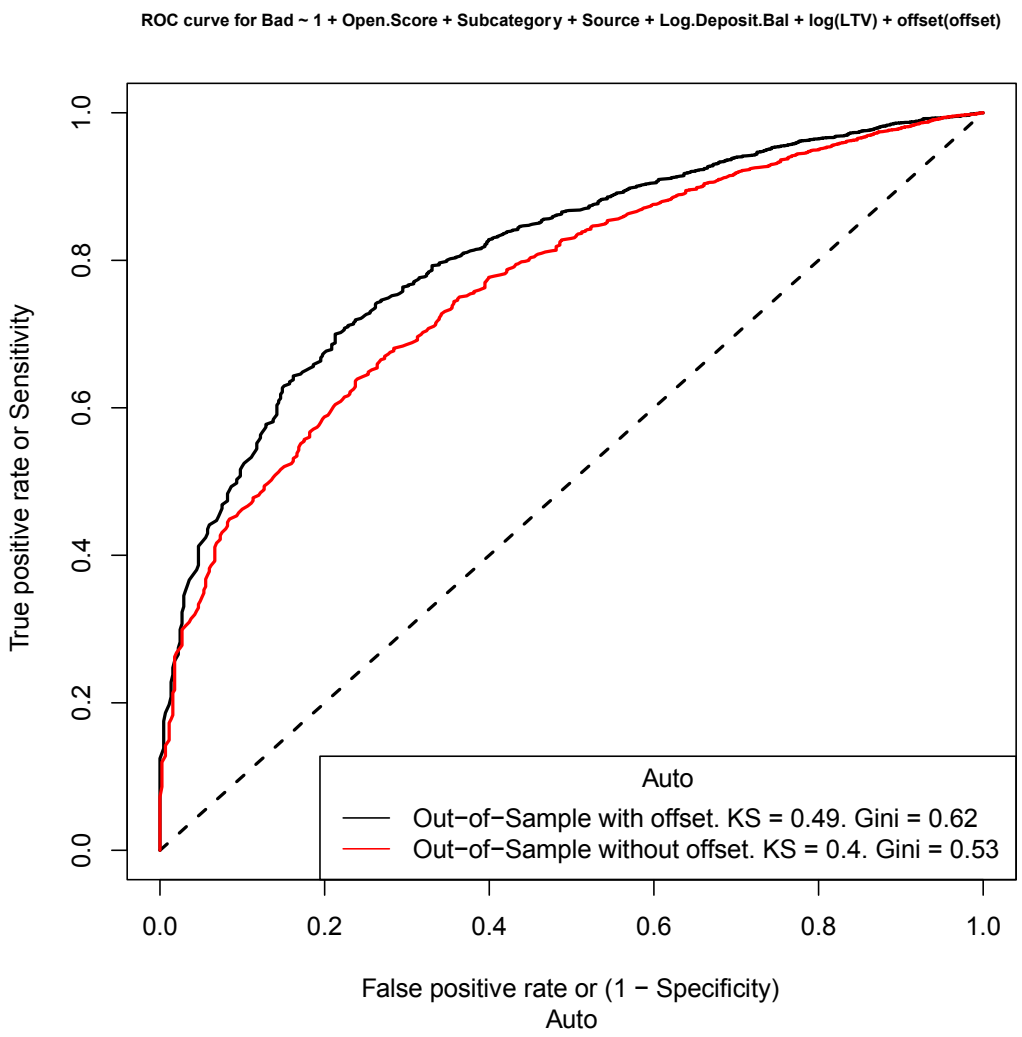


The scoring offset = F(a) + H(t)

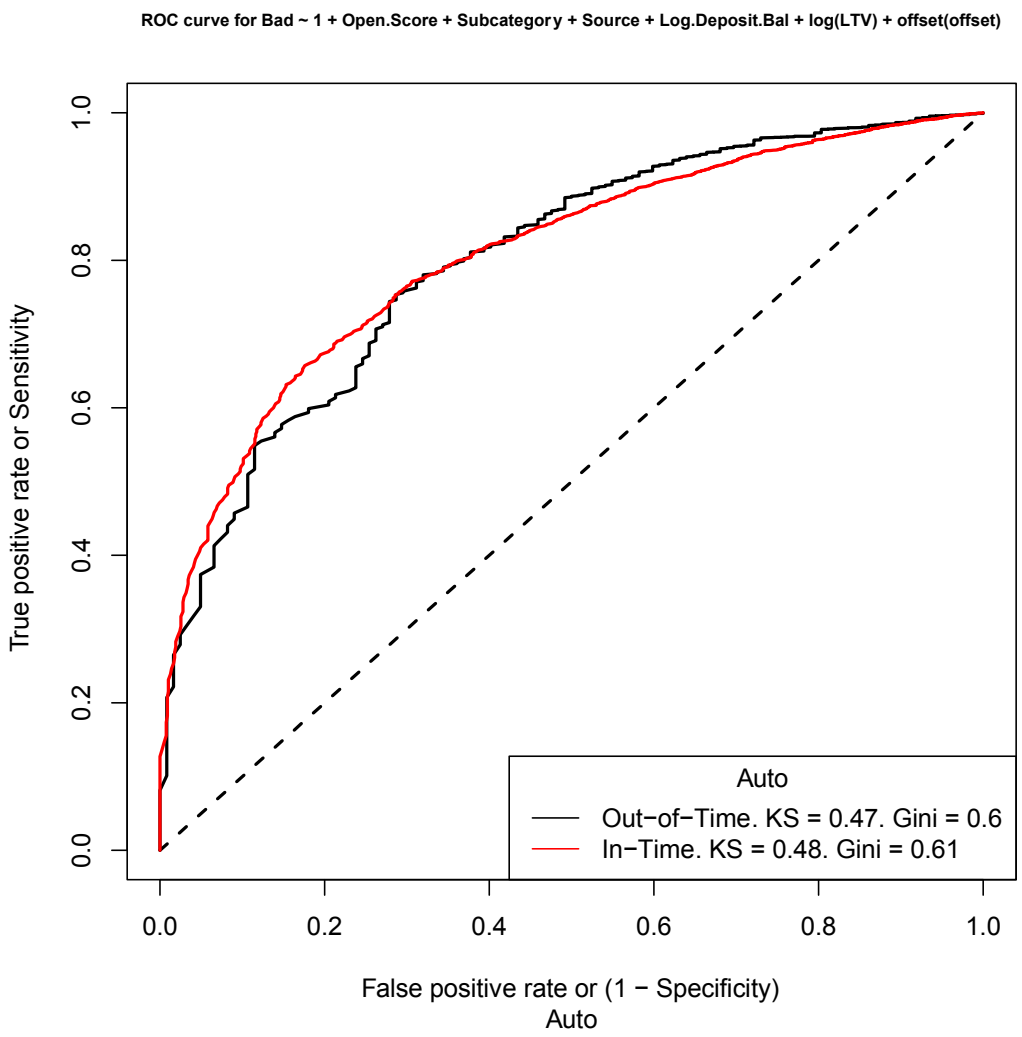
Out-of-Sample Validation



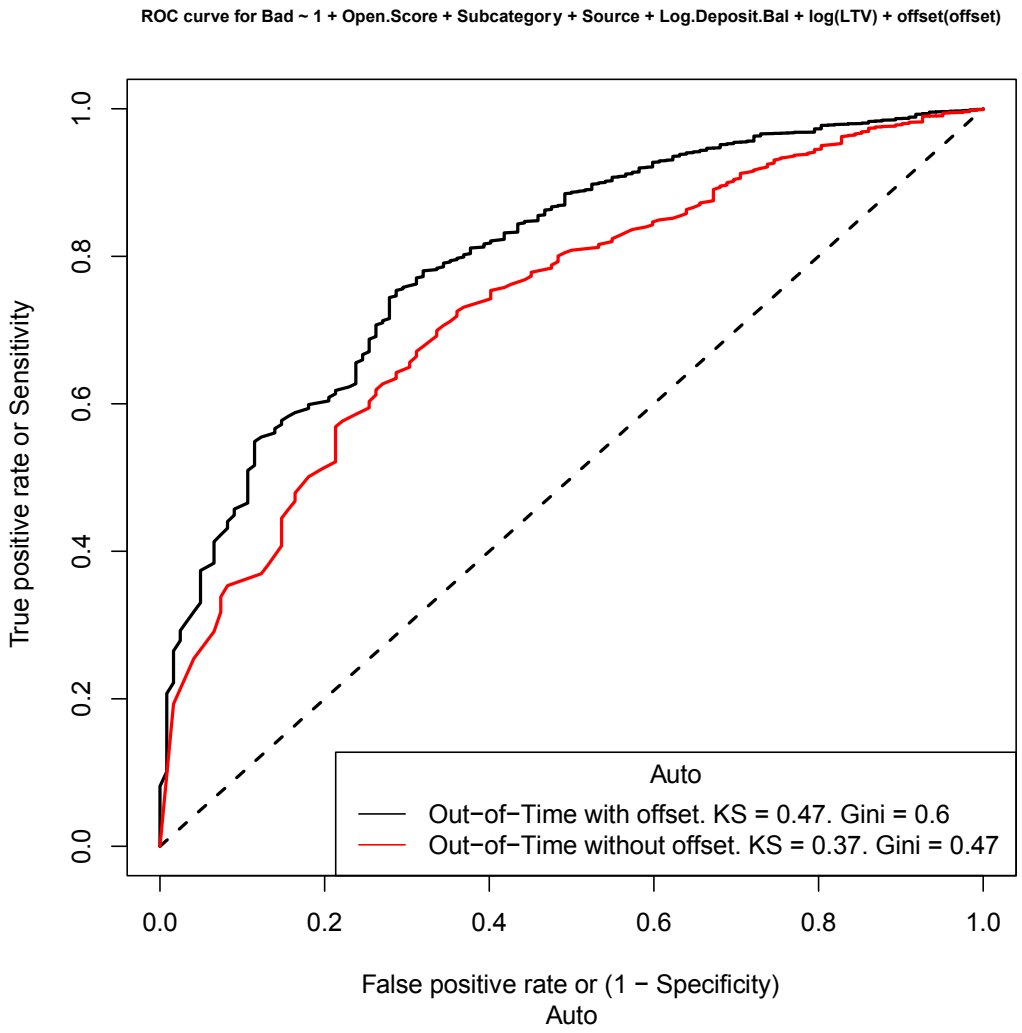
Out-of-Sample Validation, with and without Offset



Out-of-Time Model Validation



Out-of-Time Validation, with and without Offset



Forecasting

- Use lifecycle and an environmental scenario to create the offset.
- The model predicts loan-level, monthly probabilities.
- Integrate over the product lifetime for pricing or a fixed window for account management.
- Also works for behavior scoring.

Stress Testing & Capital

- No additional modeling is required for stress testing.
- Insert an extreme scenario for the future $H(t)$ and aggregate the probabilities over the desired forecast horizon.
- Through the cycle PDs can be obtained the same way, by setting $H(t) = H_{TTC}$.
- No difference between stress testing and capital.

Contact Us

- For additional information, contact:

Joseph L. Breeden, CEO
Prescient Models LLC
300 Catron St., Suite B
Santa Fe, NM 87501
+1-505-670-7670
breeden@prescientmodels.com