Moving From Rankings to Ratings

Joseph L. Breeden
President & COO
breeden@strategicanalytics.com

Robert L. Parker
Director, Professional Services
rparker@strategicanalytics.com

The Problems with Scores

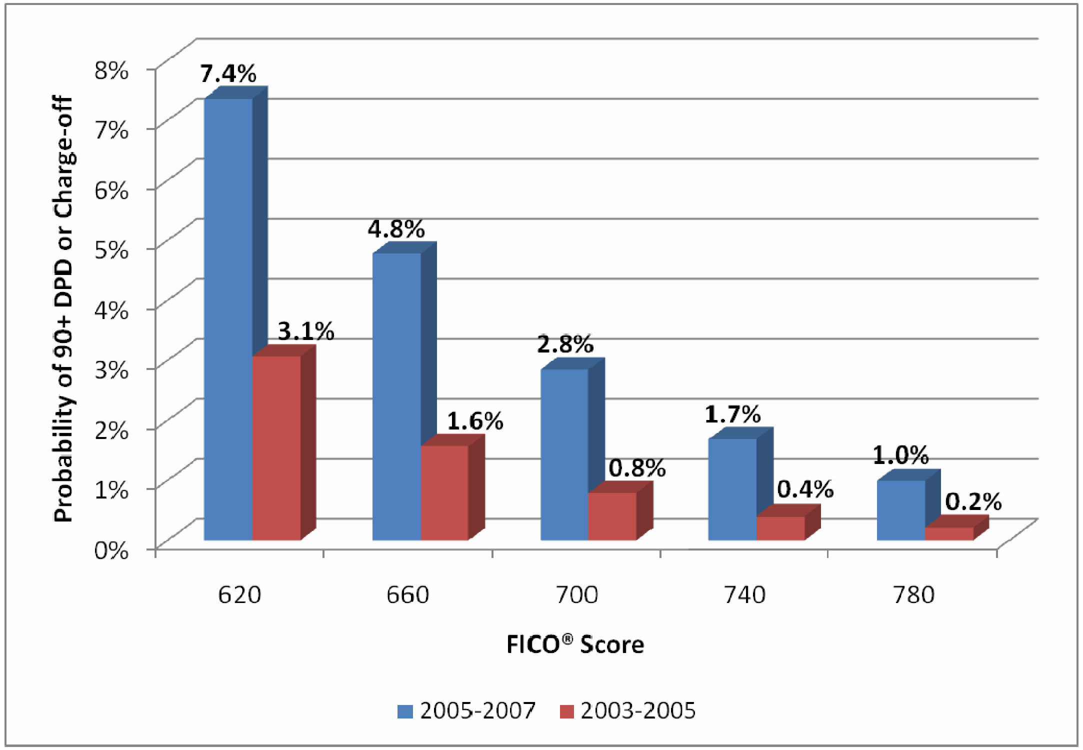
1. Difficulty calibrating scores to odds as the accounts age and the environment changes
2. Because score distributions drift and PDs vs. score change, use of behavior scores conflicts with the Through-the-Cycle goals of Basel II.
3. Using score migration as the basis of stress testing is problematic, also because both the distribution drifts and the score-odds calibration changes.

The Problem with Scores

- Existing scoring methods DO rank order, but they DO NOT predict probabilities.
- Proper pricing requires probabilities of going bad, not just rankings.
- Calibrating to probabilities is more complicated than assumed, because the scores confuse consumer behavior with lifecycle and environmental effects.
- In getting the ratings correct, we can also improve the rankings, because we know the difference between accounts that go bad in a good economy versus a bad economy.

Score Drift

- Score distributions have been observed to move significantly through economic cycles, but an equal portion of this drift is attributable to adverse selection and account maturation.



STRATEGIC ANALYTICS

Finding a Solution

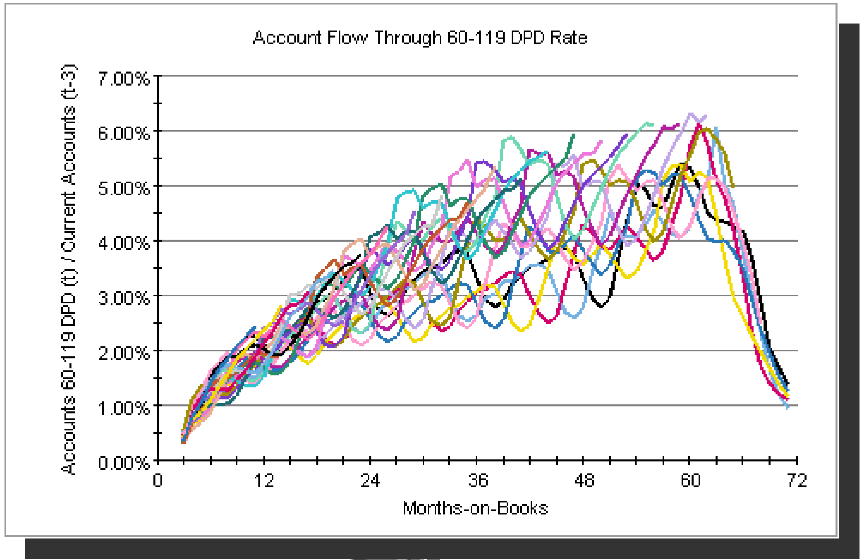
Portfolio Drivers

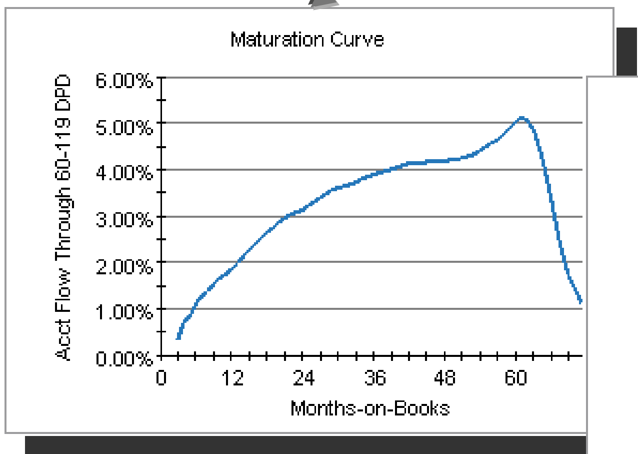
- At a portfolio level, we understand the drivers.
- These drivers need to be incorporated into the account-level modeling.
- Exactly the same drivers (functions) can be used at both the portfolio and account level.

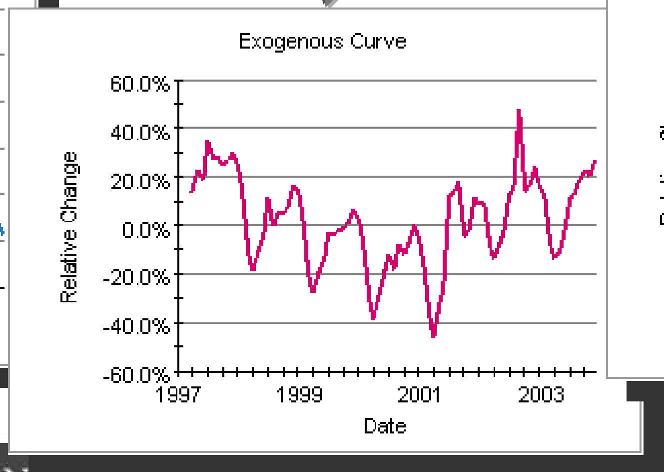
Dual-time Dynamics (DtD)

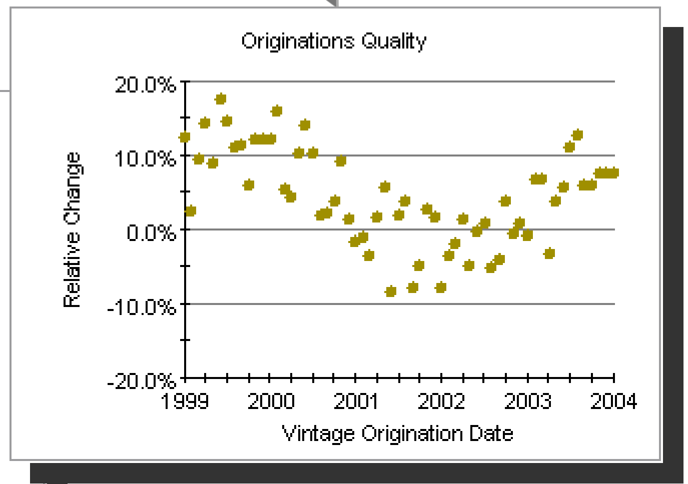
- **Vintage-level data is decomposed into functions of months-on-books (maturation), calendar date (exogenous), and vintage (quality).**

$$r(a,v,t) = e^{f_m(a)} e^{f_g(t)} e^{f_Q(v)}$$



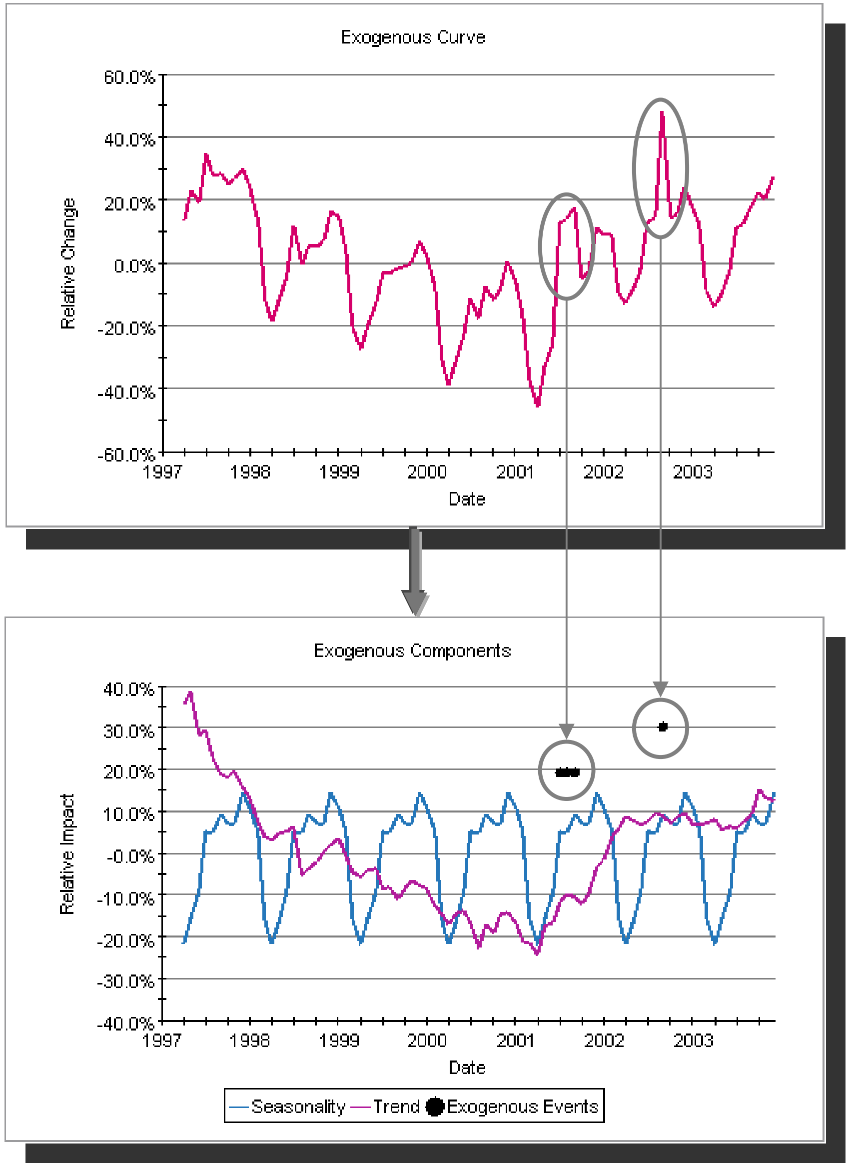






Decomposing the Exogenous Curve

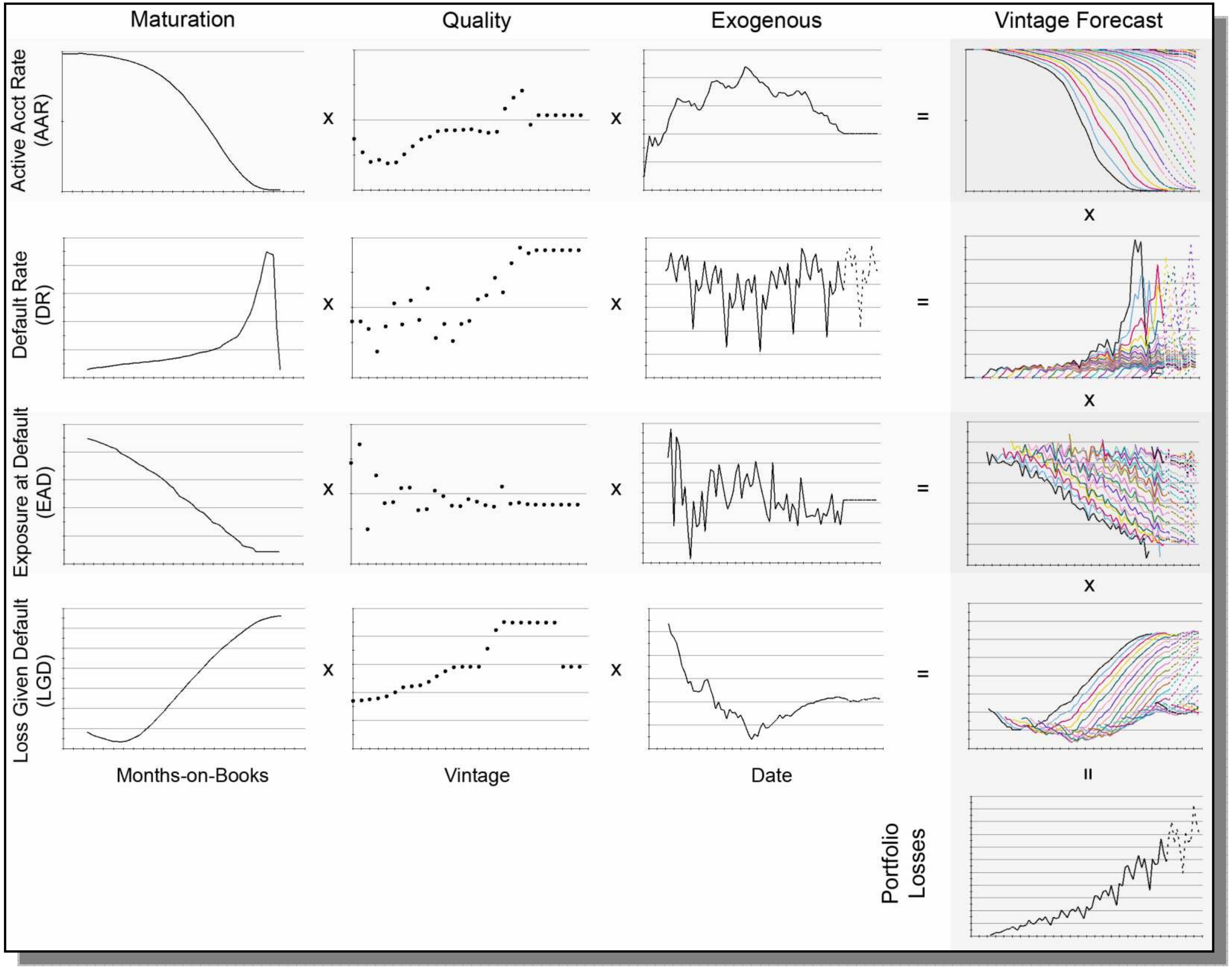
- The exogenous curve measures the relative impact of external factors upon intrinsic consumer dynamics
- e.g. “20% higher delinquency than would have been expected from the maturation process”
- To ascertain cause-and-effect, the exogenous curve is further decomposed into seasonality, trend (usually macroeconomic impact), and events (management actions).



Properties of DtD

- Does not presume scores or economic factors are the only explanations of portfolio performance.
- After measuring environmental and credit quality functions, traditional factors like scores and economics can be considered as possible explanatory variables.
- Philosophically the same as Age-Period-Cohort modeling in Demography and Tree Ring Analysis.
- Can be computed at the vintage-level for efficient calculation. Also applicable at the account level.
- Applies to any rate variable: PD, LGD, EAD, delinquency rates, attrition rates, balance ratios, etc.

Creating the Portfolio Forecast



DtD Scoring

DtD Scoring

- By explicitly including the environment, “bad” is no longer binary. We can tell the difference between being “bad” in a good environment (really bad), or being “bad” in a bad environment (not so bad).
- Augmented survival models help with this, but they often use existing scores and fixed economic factors. This approach would miss the problems faced during the mortgage crisis.
- Using the non-parametric functions from DtD allows robust analysis of short data sets, and measurement of adverse selection.

DtD Scoring

- The same basic framework, with some additions…

$$p(i,a,t) = e^{f_m(a)} e^{B \cdot X(i,t)} e^{f_g(t)}$$

Probability of "Bad" **Maturation** **Quality** **Environment**

- Vintage-level quality is now converted to an account-level "score".
- Environment is separable into economic and management factors, if desired, or left as a single exogenous curve.
- $e^{B \cdot X(i,t)}$ is the "score" with the usual rank-order properties and adverse action possibilities.

DtD Scoring

$$p(i,a,t) = e^{f_m(a)} e^{B \cdot X(i,t)} e^{f_g(t)}$$

- Environment and maturation can be inserted directly for the portfolio-level analysis without needing to re-estimate, because they are independent of the score construction – *long vintage-level data sets can be used to estimate maturation and exogenous while short account-level data sets are used to create the score*.
- Forecasts are relative to the current environment, or a specific scenario.

Adverse Selection

- Adverse Selection changes through marketing campaigns, so it is best measured by vintages.
- What cannot be explained by score is then “adverse selection”.

$$f_{AS}(v) = f_Q(v) - \sum_i B \cdot X(i)$$

- When predicting the probability of bad for new originations, we should extrapolate the adverse selection factor forward until it can be explained and solved, or proven to be gone:

$$p(i,a,t,v) = e^{f_m(a)} e^{b \cdot x(i,t) + f_{AS}(v)} e^{f_g(t)}$$

Origination versus Behavior Scores

- The DtD Scoring framework applies for both origination and behavior scores.
- When the environment is trending worse, many accounts will show increased stress.
- This aggregate environmentally-driven stress will be captured in the exogenous curve, so both origination and behavior scores will answer the question, “Is this account showing more stress than expected due to the economy?”

Validation

- The score component can be validated with the usual measures: Gini, KS, etc.
- But we can also do temporal tests on the scores.
  - Is rank ordering preserved over different time periods?
  - Is the score distribution stable for different time periods?
- And we can test the maturation and exogenous components for stability across the sample.

Comparison to Survival Models

- Survival Models and Proportional Hazards Models can be used for this analysis, with a few modifications.
- Most practitioners use short (~2 yr) data sets of account-level data for their Survival Models. Trying to incorporate macroeconomic factors explicitly is problematic at best.
- Instead, they either need to learn the lifecycle and macroeconomic components from longer time series and insert into their scoring.
- Or, they need to use a non-parametric function for the exogenous curve instead of trying to use specific macroeconomic factors.

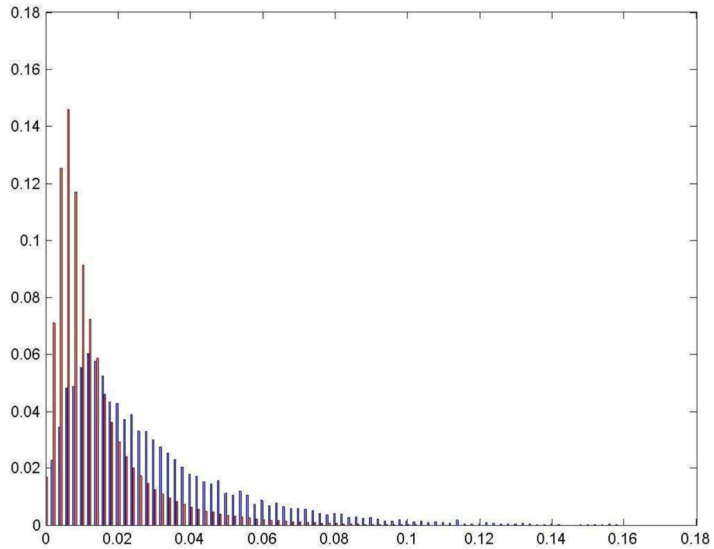
Results

- DtD Scoring model: In a 12-month window,

$$\text{P(Bad)} = \sum_{t=1}^{12} p(i,a,t) = e^{B \cdot X(i,tOrig)} \sum_{t=1}^{12} e^{f_m(a)} e^{f_g(t)}$$

- Logistic Regression scoring model for 12-month Bad:

$$\text{P(Bad)} = 1 \big/ \left(1 + e^{-B \cdot X(i,tOrig)}\right)$$



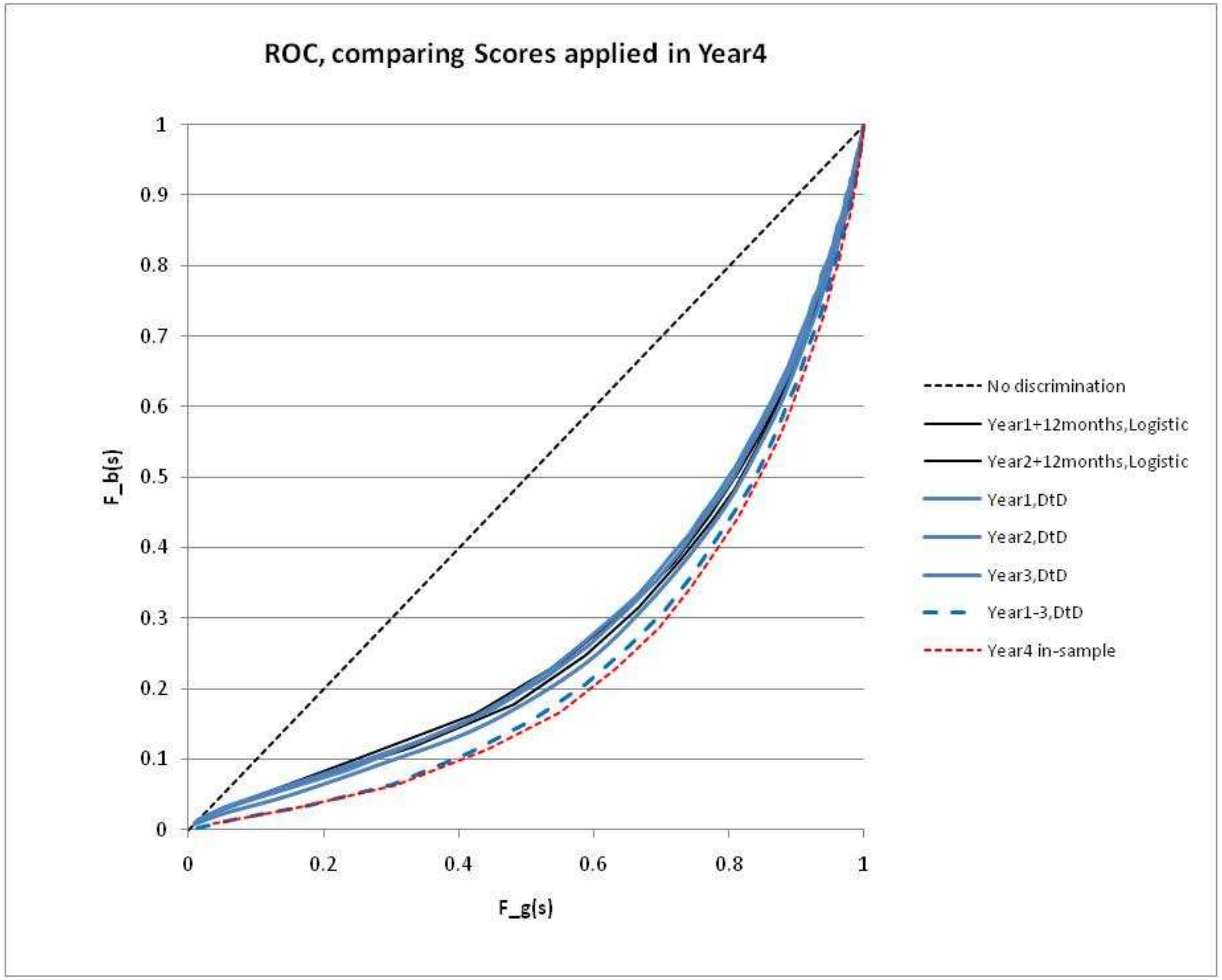
- Plot of Good/Bad distributions is reversed from the usual view: The “Score” is the P(Bad).
- --- Goods
- --- Bads

Results - Discrimination

Year4: Comparing DtD Scoring to Logistic Regression: Good/Bad Discrimination

- 5-year account-level data set, Year1-Year5
- Models built in Year1-Year3,
- Each Score is applied to 12 months of new originations in Year4, with each account followed for a 12-month Good/Bad determination
- The DtD Scoring model, with coefficients B estimated over Year1-Year3, approaches the level of discrimination of the *in-sample* Year4 logistic regression model

Note: The logistic regression models are not improved by using a longer training period

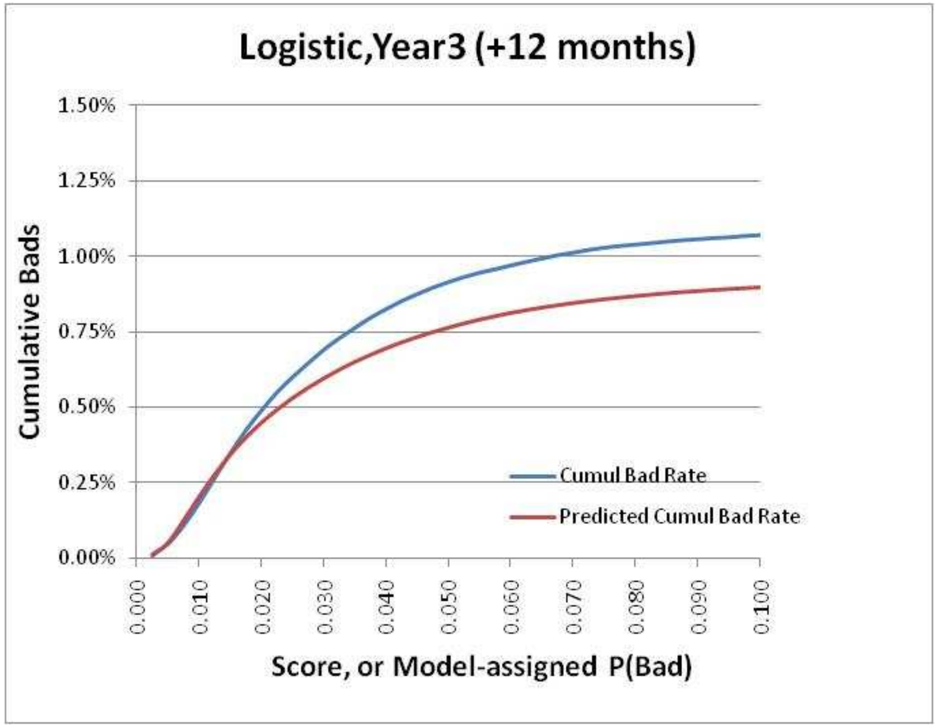


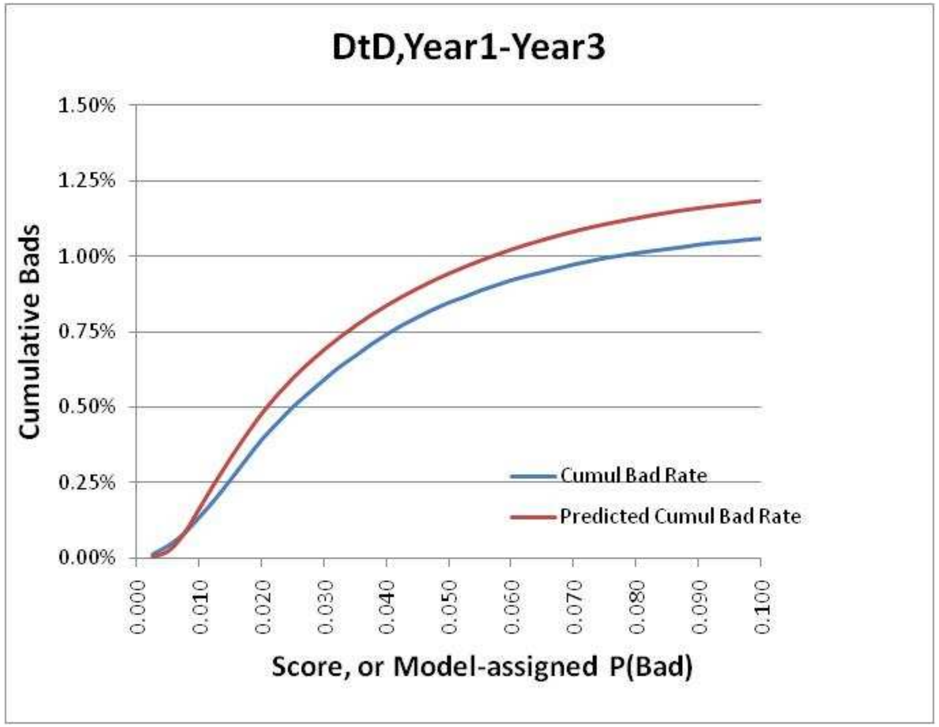
Results - Prediction

Year4: Comparing DtD Scoring to Logistic Regression: Prediction of Bad Rate

- Logistic models built in Years 1-3 under-predict the Bads rate
- Year4 environment worsened, and
- Model drifted as time passed
- Note that Year3 Logistic model “peeks” into future

- DtD Scoring models built in Years 1-3 may over- or under-predict the Bads rate, depending on goodness of economic forecast
- With DtD Scoring, a model built over a longer time period improves prediction





Conclusions

Implications for Basel II

- No net migration of accounts will occur across scores (grades), i.e. the score distribution will be stable through the economic cycle.
- Score distribution stability is what is needed to make Basel II function as intended. Long-run PDs as a function of score will then be stable through the economic cycle – no procyclicality.
- So, DtD Scores can be used for Basel II without modification on either side. The results will be much better than standard scores.

Implications for Stress Testing

- Most practitioners try to segment by behavior score and then stress each segment.
- The problem is that there is net drift in scores across segments and changes in PD for each segment.
- With DtD Scores, there is no net drift across segments, and at any score value, we get a PD stress test simply by putting in a stressed economic scenario.
- Dual-time Dynamics has been used for stress testing for 10 years, so DtD Scoring allows behavior scores to become a segmentation dimension within that same framework.

Through-the-Cycle Economic Capital

- The Basel II regulatory capital formula does not recognize the "predictable volatility" arising from consumer lifecycles.

- Using the distributionally-stable DtD Scores, we can compute portfolio, segment, or account-level capital as

$$PD(v) = e^{b \cdot x(i,t) + f_{AS}(v)} e^{D^{-1}(s)} \sum_{t=1}^{12} e^{f_m(t-v)}$$

- Where D is the distribution of annual values of the exogenous curve, and s is the desired solvency level. The maturation curve is integrated over the time horizon desired in the capital calculation.

- This approach would not have failed in the US Mortgage Crisis.

The Scores Did Not Fail

- Through the recent economic crisis, scores did not fail to rank-order. Rather, the calibration between score and odds or probability of default failed badly.
- This approach succeeds at creating time series forecasts of odds or probability of default given a scenario for the future environment.
- The need to frequently “recalibrate” scores is largely eliminated. Longer data sets can be modeled.
- DtD Scores provide some additional lift relative to logistic regression, because we normalize for the environment in which the default occurred.
- Adverse selection can be quantified and included in pricing.