

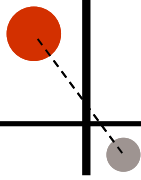


MULTIHORIZON SURVIVAL MODELS

Joseph L. Breeden, Prescient Models

Jonathan Crook, U. of Edinburgh





IFRS 9 and CECL

- Both loss reserve standards mandate long-range forecasting. For CECL it's everything.
- Ideally done at the loan-level
 - Need to incorporate delinquency and behavioral factors.
- Challenges:
 - Vintage models are good at long range forecasting, but not account level
 - Delinquency-based methods are good short-term, but not loan-term
 - Cox Ph models as with all methods will have a multicollinearity problem.



Loan-level Cash Flows

If we do this right, our CECL or IFRS 9 models can be used more widely.

We are already seeing applications to:

- Pricing loans at origination
- Valuing loans to sell or purchase
- Integration with product up-sell / cross-sell
- Integration with collections



Start with Vintages

We need lifetime and we know vintage (Age-Period-Cohort) are good at lifetime forecasting

$$F(a) = \sum_a \alpha_a \delta(a) \quad \text{Lifecycle vs. age, } a$$

$$G(v) = \sum_v \beta_v \delta(v) \quad \text{Credit risk vs. vintage, } v$$

$$H(t) = \sum_t \gamma_t \delta(t) \quad \text{Environment vs. date, } t$$

$$\delta(a) = \begin{cases} 1, & \text{if } j = a \\ 0, & \text{if } j \neq a \end{cases}$$

The target variable (default) is predicted as the sum of these effects. Interaction terms usually solved via segmentation.

$$Def(i, a, v, t) \sim F(a) + H(t) + G(v)$$



Estimation

To estimate α_a , β_v , and γ_t , we need the constraint

$$\sum_v \beta_v \delta(v) = \sum_t \gamma_t \delta(t) = 0. \quad \text{i.e. one constant term.}$$

Also, because $a = t - v$, we have a specification error in the linear terms.

We can project the constrained F , G , and H onto an orthogonal basis, remove one of the linear terms, estimate coefficients, and project back to the original functions – the Bayesian Age-Period-Cohort approach.



Loan-level Estimation

Using dummy variables for the vintage function, $G(v)$ provides forecast accuracy but not explanation.

Therefore, we replace $G(v)$ with scoring coefficients and new vintage dummies to capture adverse selection by vintage.

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

In this formulation,

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



Combining with Behavior Scoring

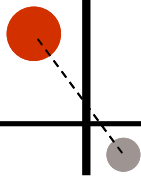
- For existing accounts, we need to incorporate delinquency. Using indicator variables, $D_i^k(t - L)$ where L is lag and k is delinquency state,

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

where

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

- If we simultaneously estimate $F(a)$, $H(t)$, c_j , d_k , and β_v'' , we would have a collinearity problem.
- By incorporating $F(a)$ and $H(t)$ from the APC analysis, the c_j and d_k can be estimated without the collinearity problem.



Multiple Horizons

- We can forecast defaults at horizon, h , using information lagged L months (no future information allowed).
- Equivalent to saying

$$Def(a, v, t_0 + h) = F(a) + H(t) + \sum_j c_j S_{ij} + \sum_k d_k D_i^k(t_0) + \sum_v \beta_v'' \delta(v)$$

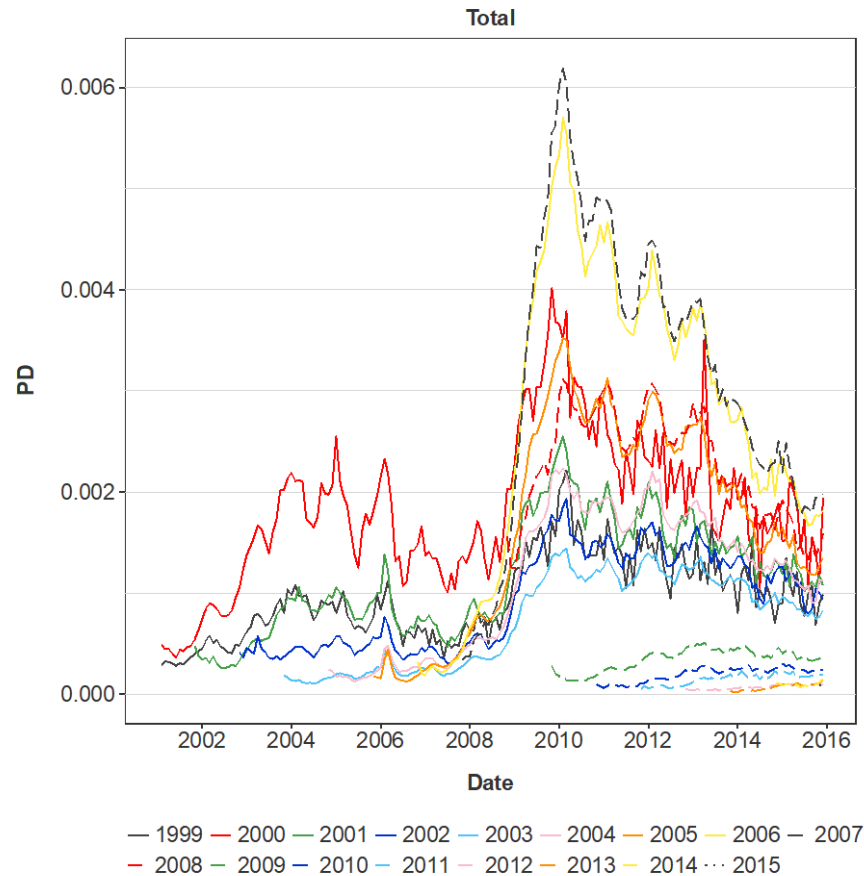
- The coefficients d_k or for any behavioral information will depend strongly on forecast horizon (lag L).
- Coefficients for each horizon are solved independently.



EXAMPLE FOR US MORTGAGE



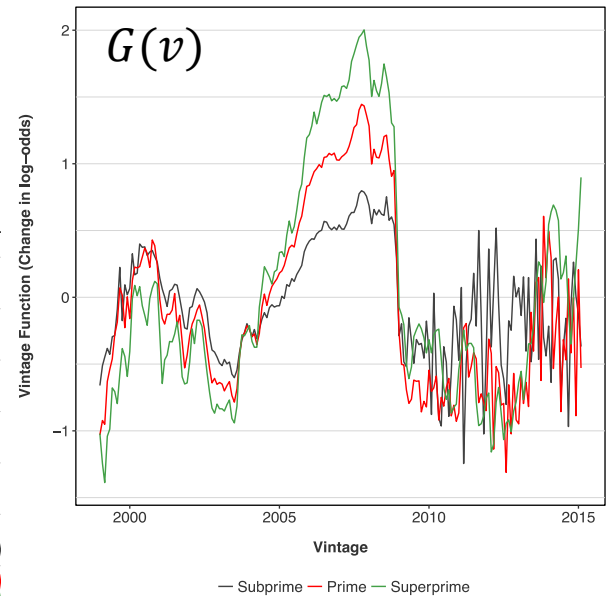
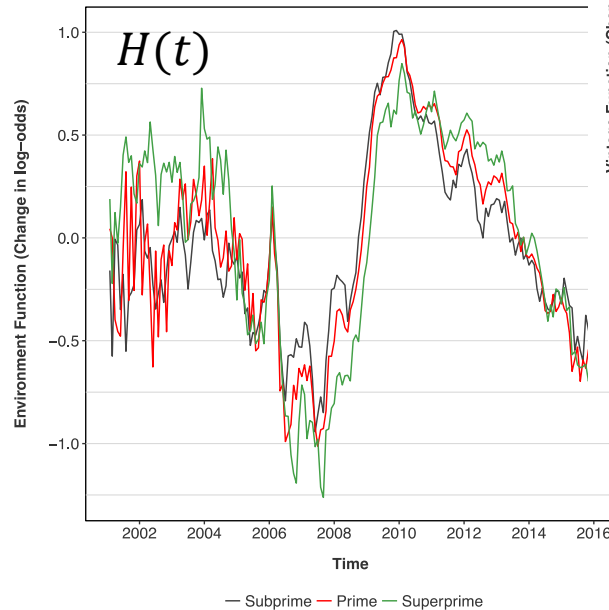
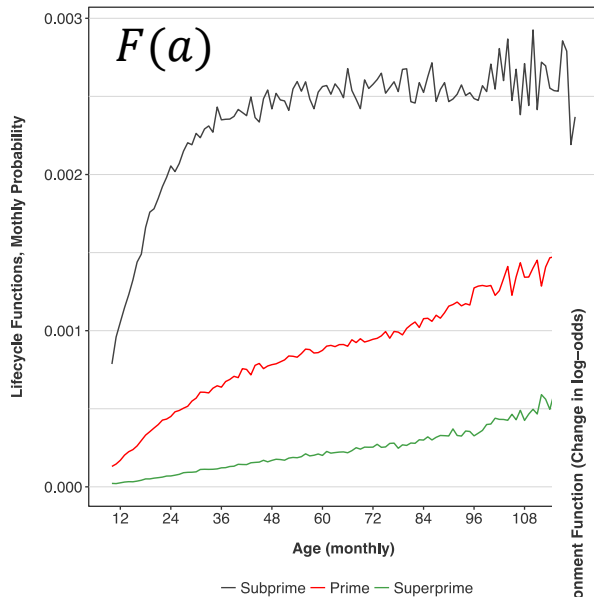
Vintage Models: Vintage-aggregate Data



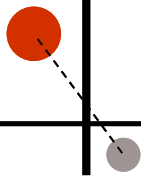
Vintage-aggregate data is created for default rate, attrition rate, outstanding balance, and default balance ratio.

Vintage Models: Default Rate Decomposition

Each key rate is decomposed into a lifecycle versus age of the loan, credit quality by vintage, and environmental impacts by calendar date.



The environment is correlated to macroeconomic factors so that economic scenarios may be used to predict each vintage.



Modeling Economic Sensitivity

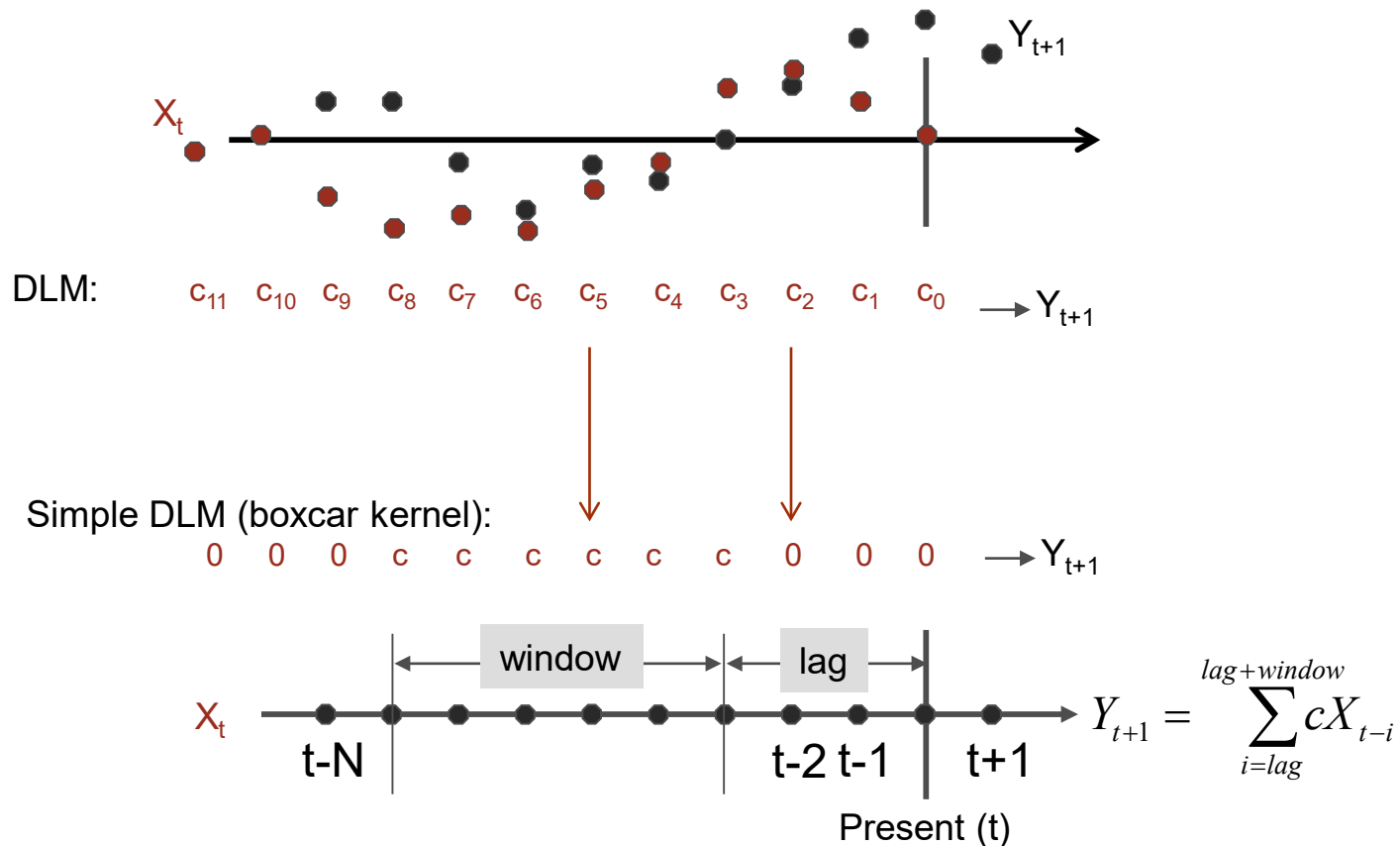
The default rate and pay-down rate time series were modeled with macroeconomic factors.

The following list of macroeconomic variables was chosen from the government-published DFAST scenarios as having the strongest relationships to mortgage performance.

Variable	Transformation
Real GDP	Log ratio
Nominal Disposable Income	Log ratio
Unemployment rate	Moving average
Unemployment rate	Difference
Unemployment rate	Log ratio
House Price Index	Log ratio
Mortgage Interest rate	Difference
DJIA	Log ratio

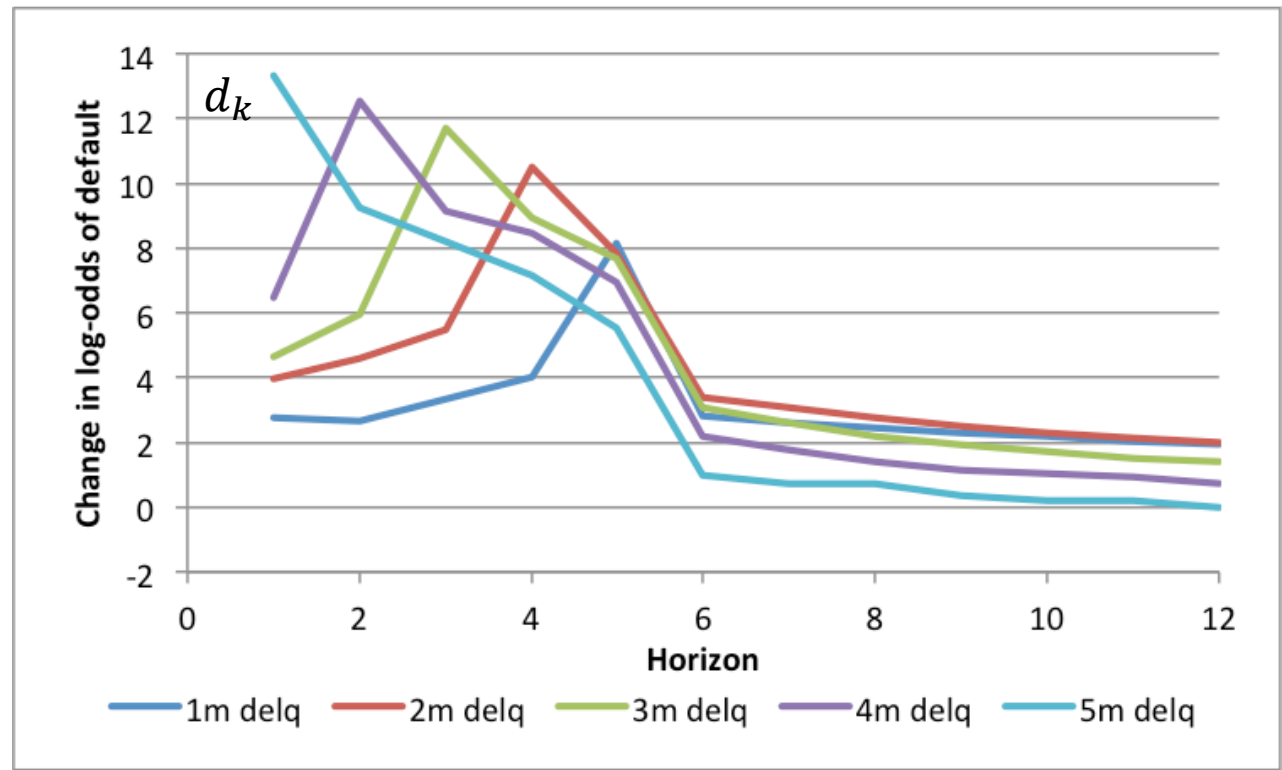
Illustration of Simplified Distributed Lag Model

- A structure is assumed between the coefficients to reduce the degrees of freedom.



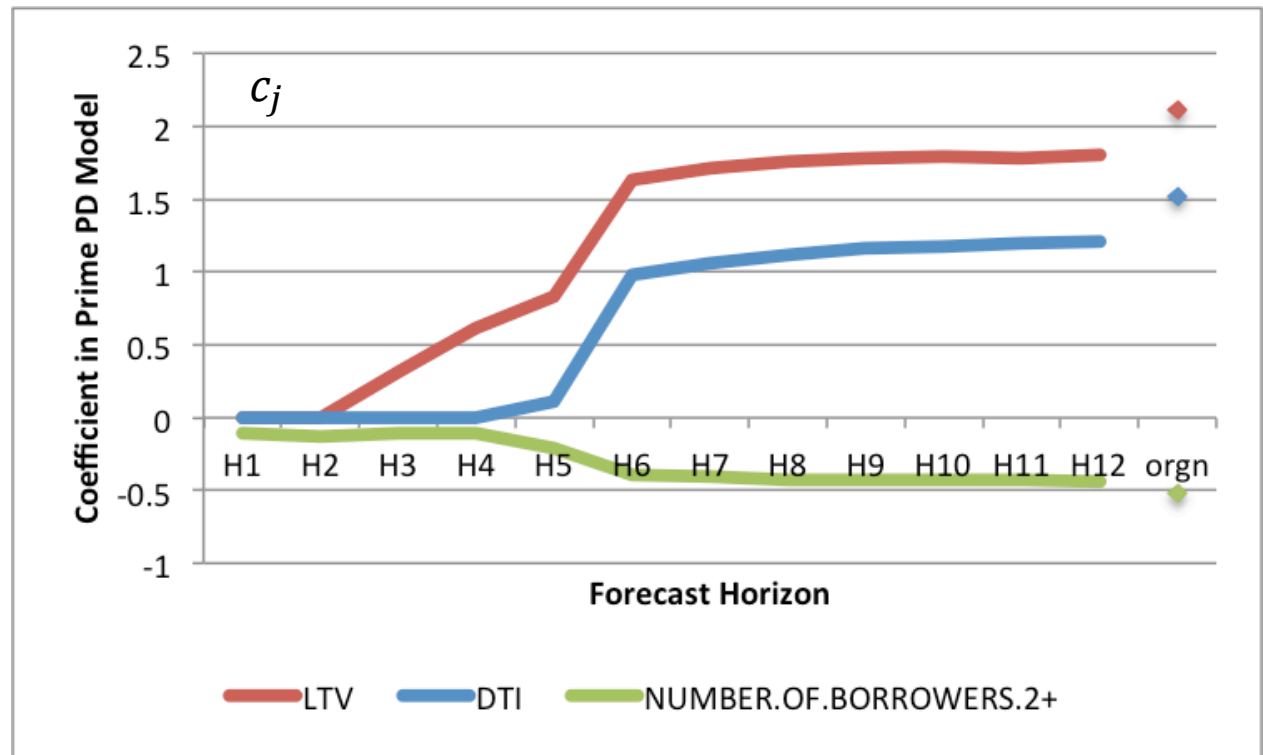
Multihorizon Survival Model: Delinquency Coeffs

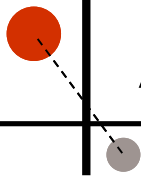
Changing the coefficients by forecast horizon captures the unique “information decay rate” for different variables.



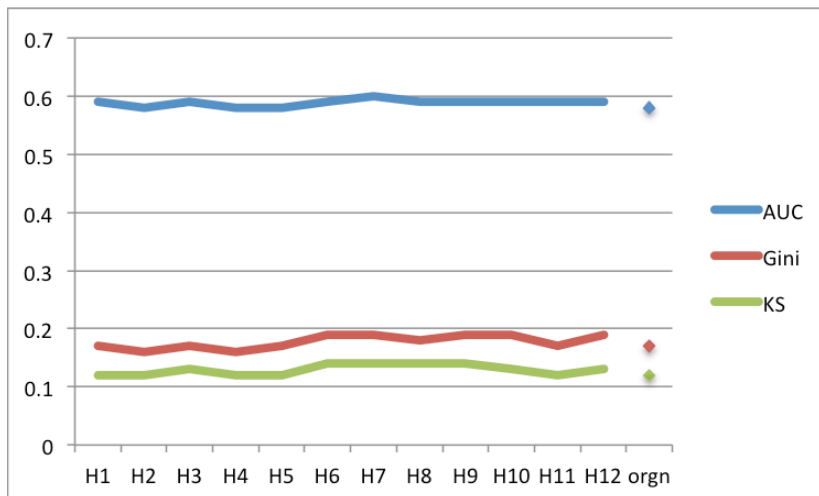
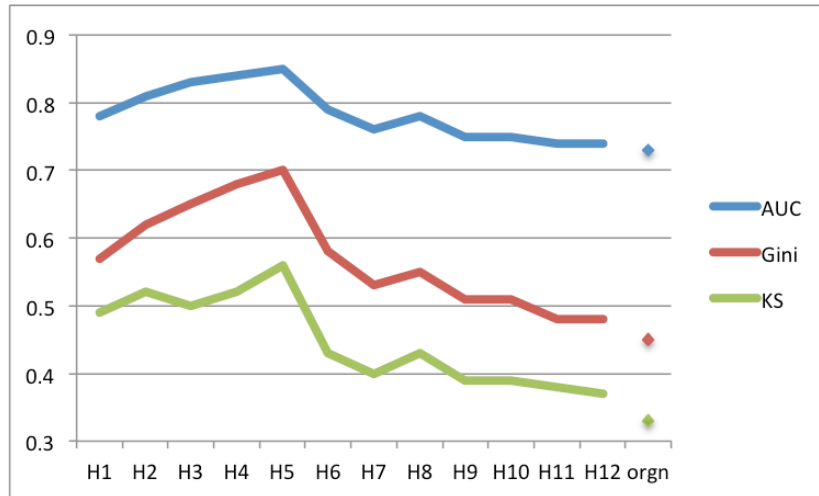
Multihorizon Survival Model: Origination Coeffs

For long horizons, the coefficients converge back toward the coefficients at origination, because behavioral information loses value quickly.



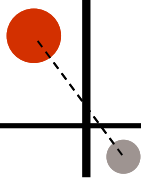


Account-level Accuracy

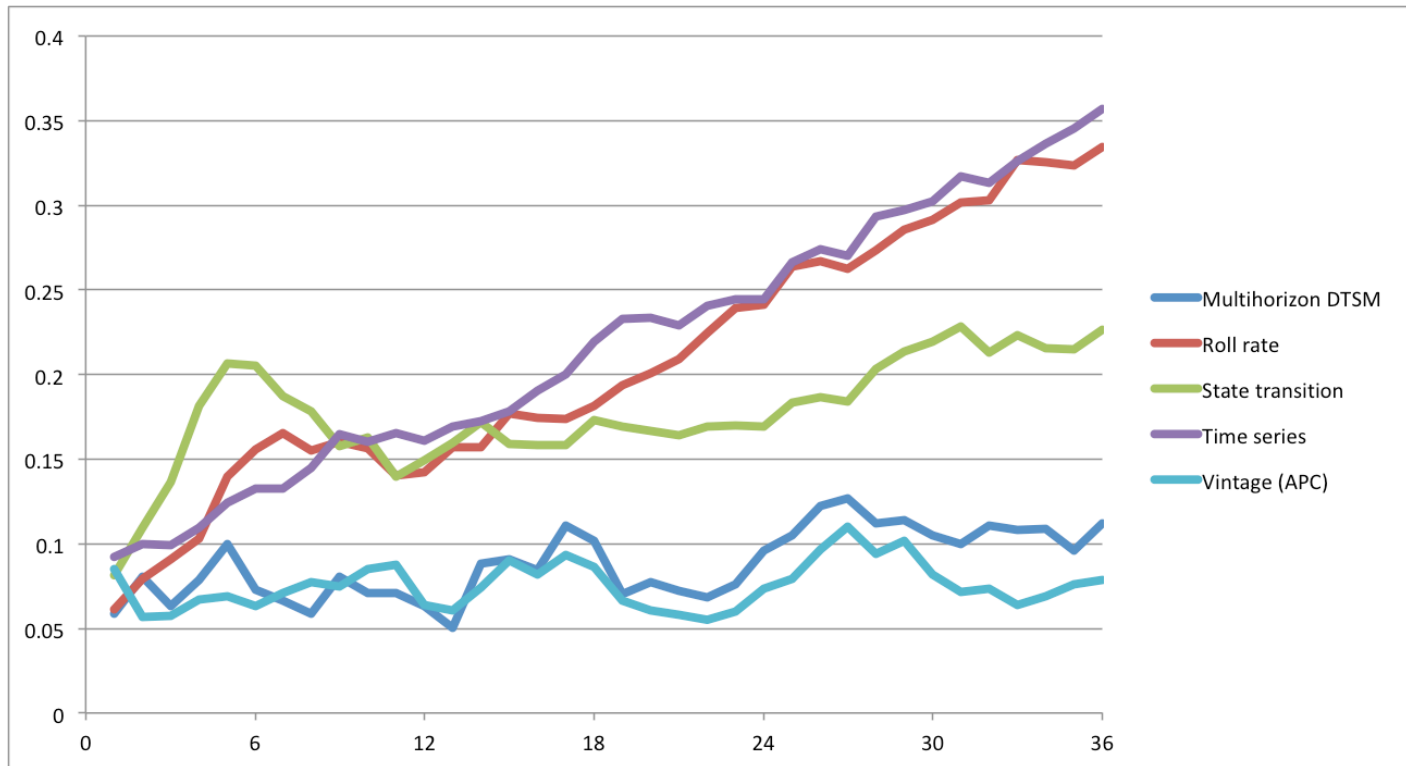


Accuracy is higher in the first few months because of the importance of delinquency in predicting initial defaults.

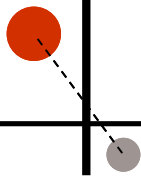
In later months, the accuracy converges to that of the origination score.



Forecast Accuracy by Horizon



- Vintage and APC Scoring models have better long-term performance.
- Time series and roll rate models start ok by trend strongly to increased error.



Contact & References

Joseph L. Breeden, Prescient Models LLC, breeden@prescientmodels.com
Jonathan Crook, U. of Edinburgh, j.crook@ed.ac.uk

Related work:

T. Bellotti and J. Crook. Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting*, 29(4):563 – 574, 2013.

Joseph L. Breeden and Jose J. Canals-Cerdá. Consumer risk appetite, the credit cycle, and the housing bubble. *Journal of Credit Risk*, 14(2):1–30, 2018.

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.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.

T R Holford. The estimation of age, period and cohort effects for vital rates. *Biometrics*, 39(2):311–324, 1983.

Volker Schmid and Leonhard Held. Bayesian age-period-cohort modeling and prediction - bamp. *Journal of Statistical Software, Articles*, 21(8):1–15, 2007.