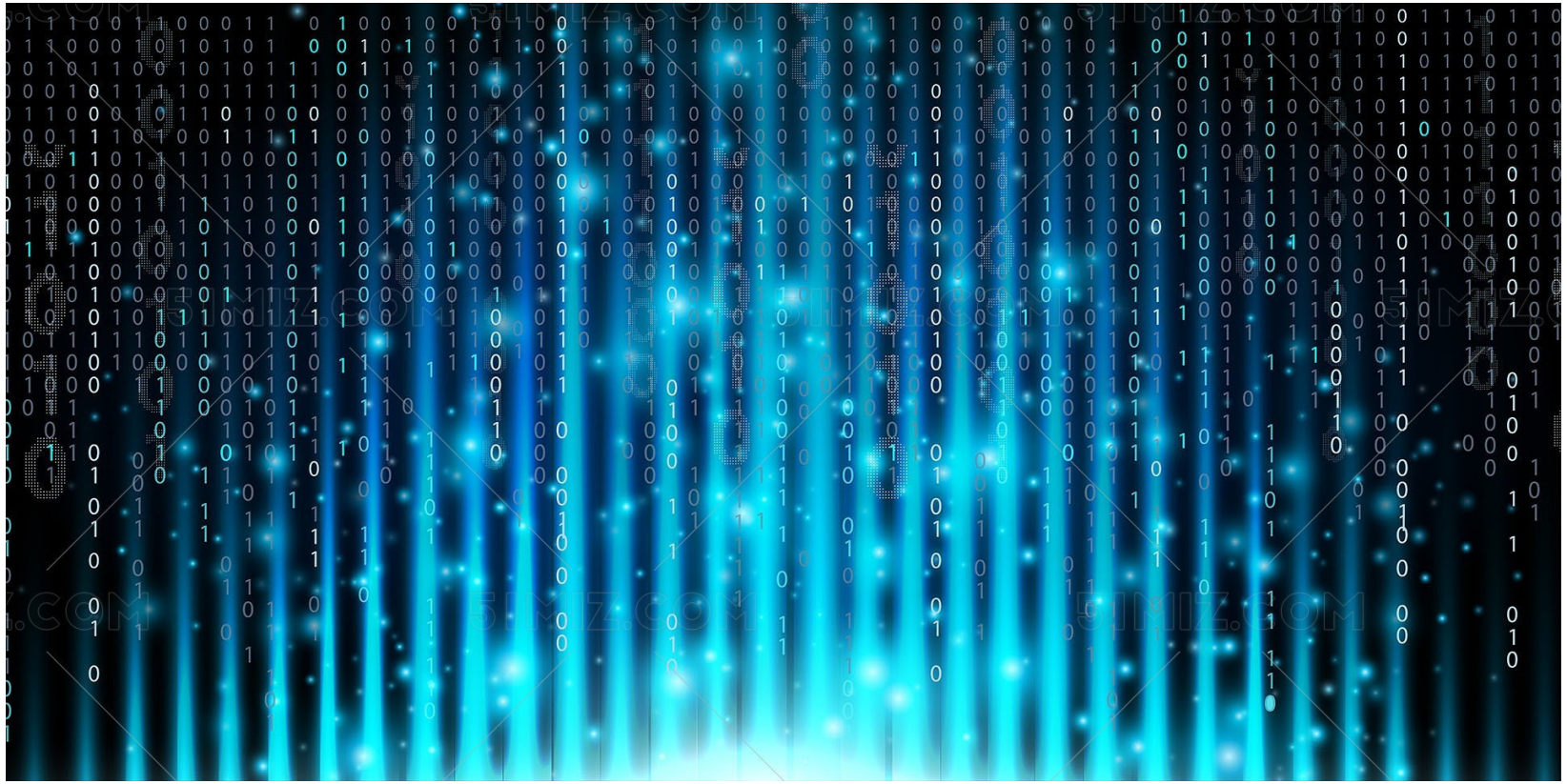
Stabilizing Machine Learning Models with Age-Period-Cohort Inputs for Scoring and Stress Testing

Joseph L. Breeden and Yevgeniya Leonova

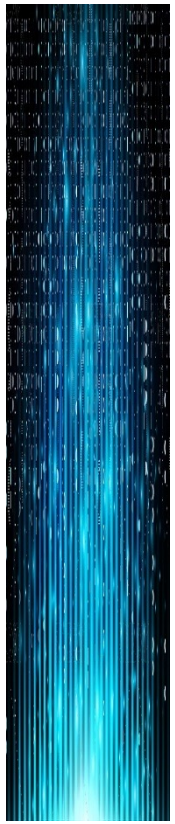
Deep Future Analytics LLC

[Breeden, Leonova]@deepfutureanalytics.com

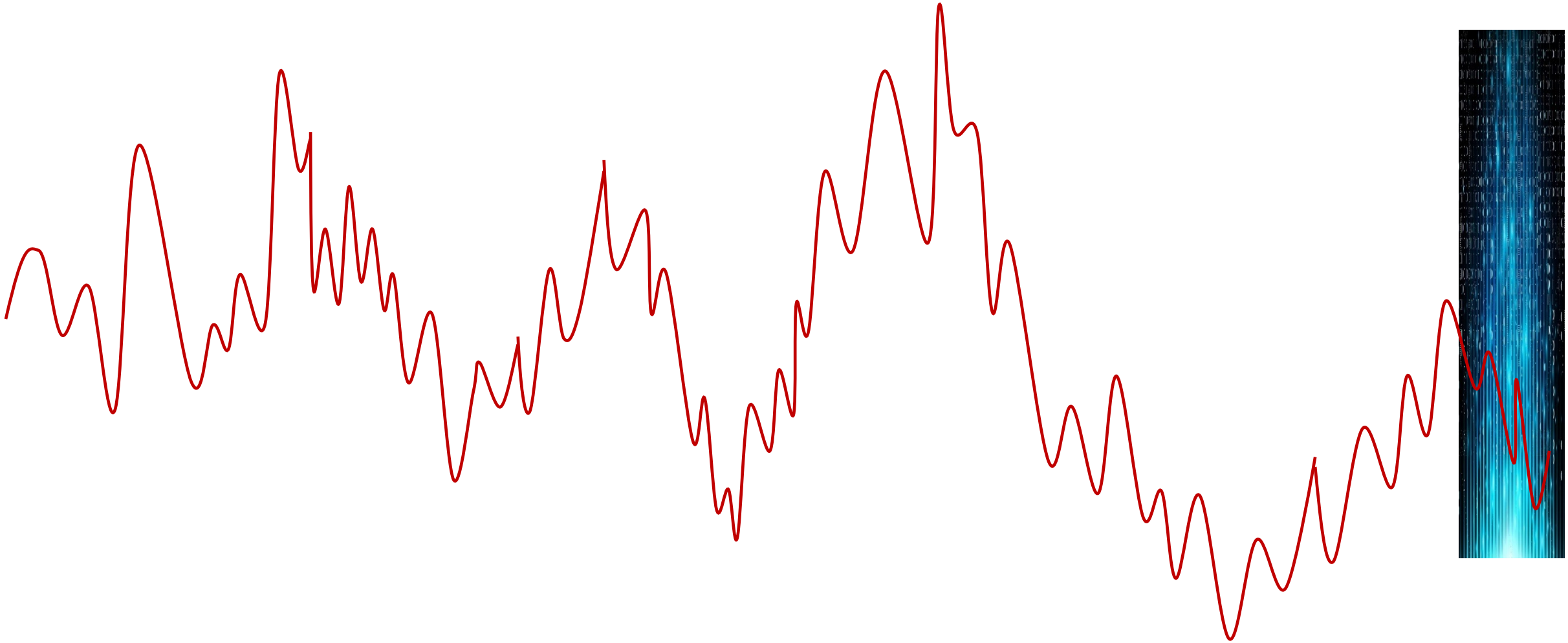
Big Data is enabling advances in Artificial Intelligence and Machine Learning



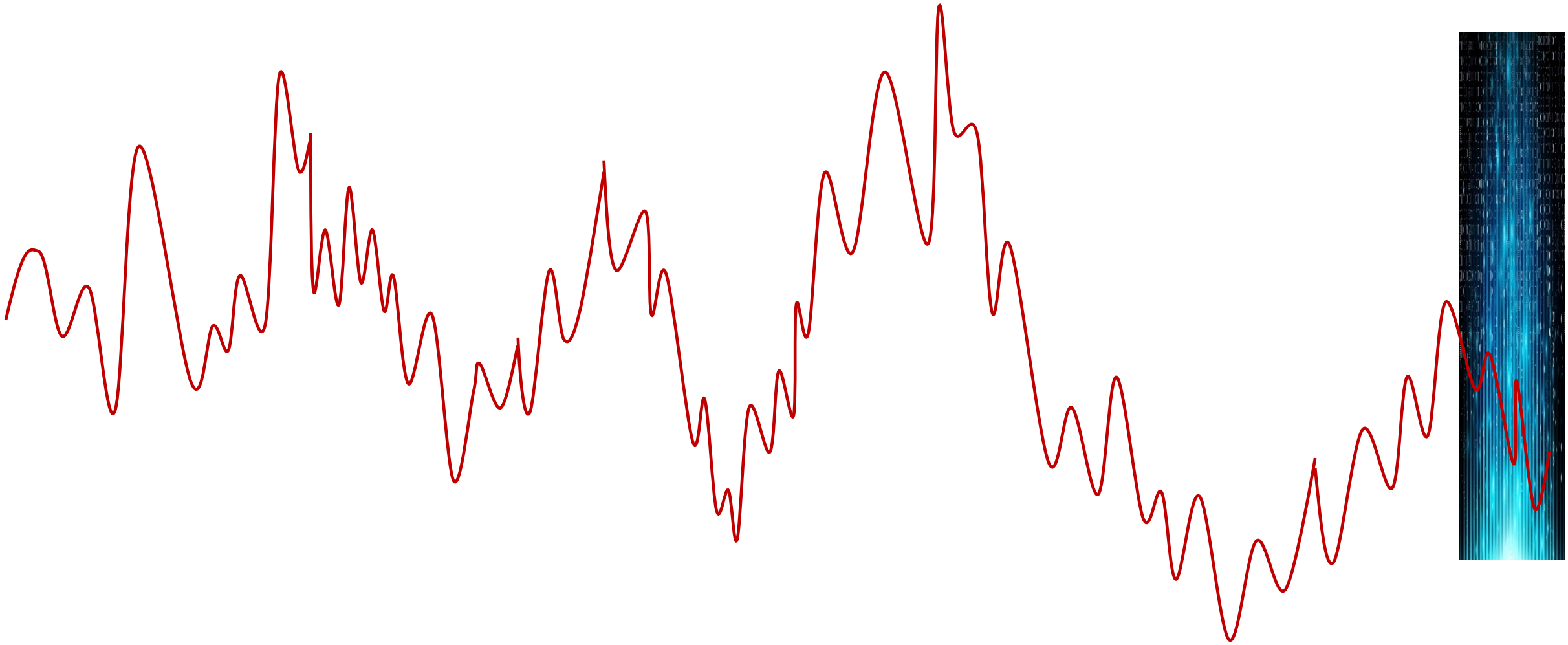
But we only have Big Data for a short slice of history



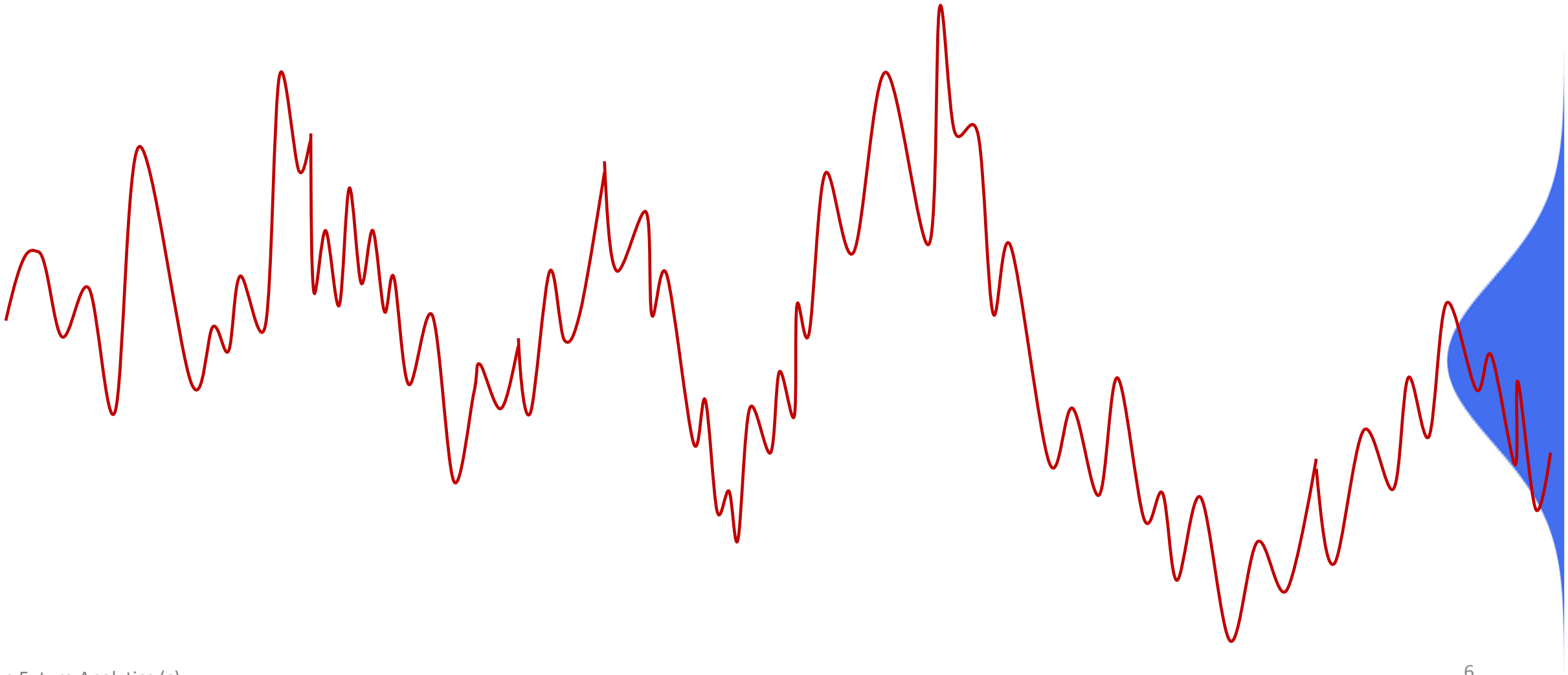
Critical information is available only in long, thin data



We need different types of models for these different types of data



If we start with a model of the long, thin data, then the AI / ML model can learn around that



Short-term and Long-term Structure

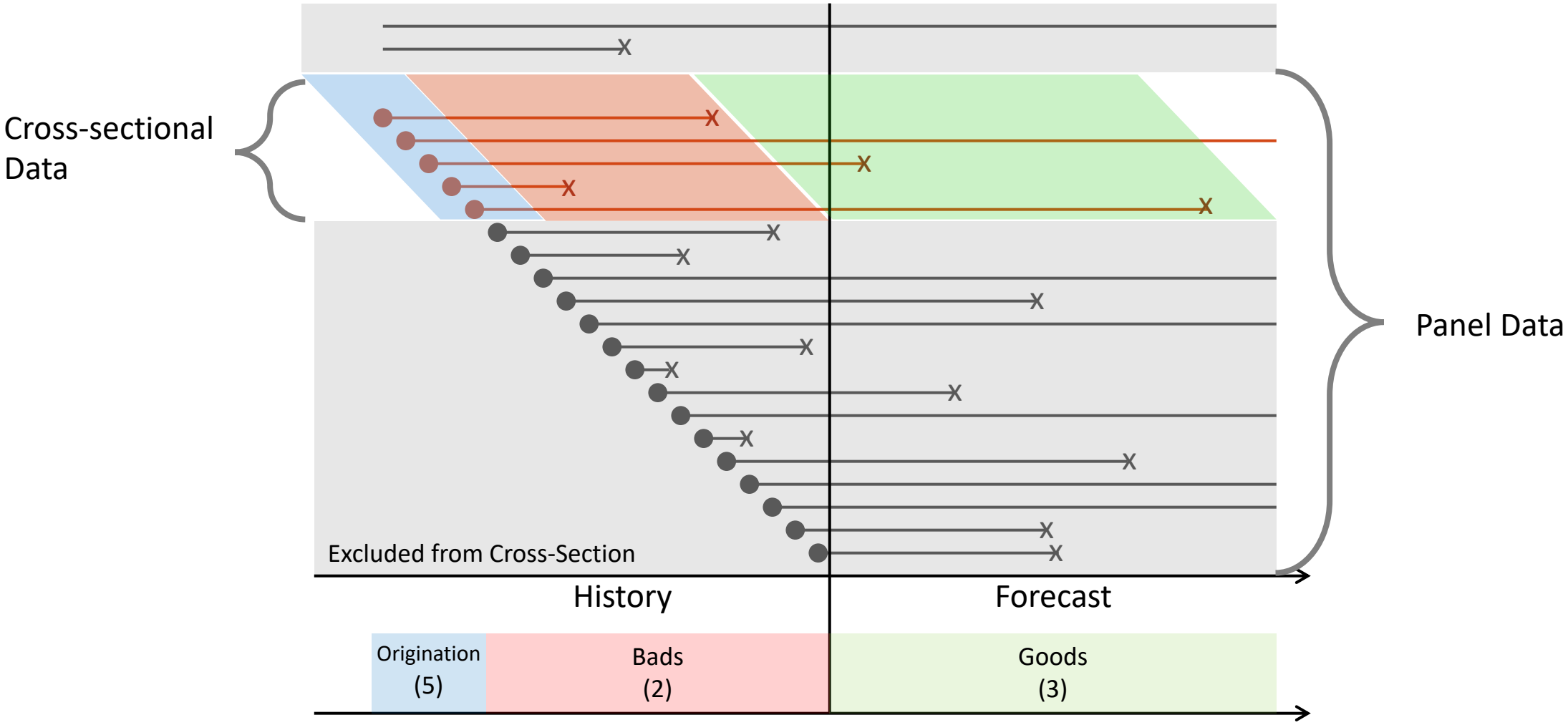
DEEP FUTURE ANALYTICS

"…the risk differentiation can be based on shorter time series, whereas risk quantification has to be based on long-run averages that require extended historical observation periods, which might not yet be available for new data sources which could be used by ML techniques." -- MACHINE LEARNING FOR IRB MODELS: FOLLOW-UP REPORT FROM THE CONSULTATION ON THE DISCUSSION PAPER ON MACHINE LEARNING FOR IRB MODELS, EBA/REP/2023/28 AUGUST 2023

- Most organizations create their short-term model first (credit score), then calibrate to long-term trends.
- The risk is that the short-term model misassigns the long-term trends with short-term patterns before the long-term modeling gets a chance.

- Solution: Build the long-term model first.

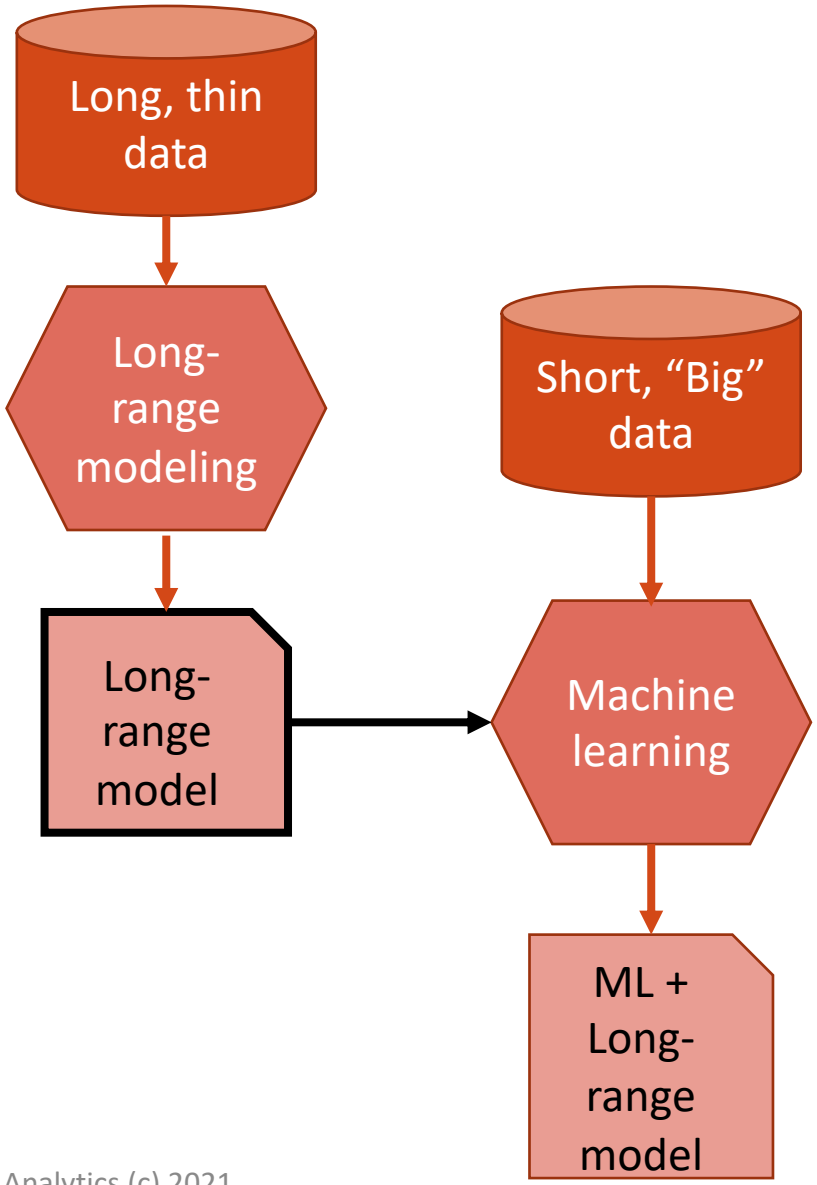
Panel or Cross-sectional Data



With a fixed outcome window, cross-sectional data does not explicitly incorporate temporal information.

A Two-Step Solution

- Thin, long historic data feeds into an existing method for making long-range forecasts.
- The long-range model is a fixed input to the ML model development on short, big data.



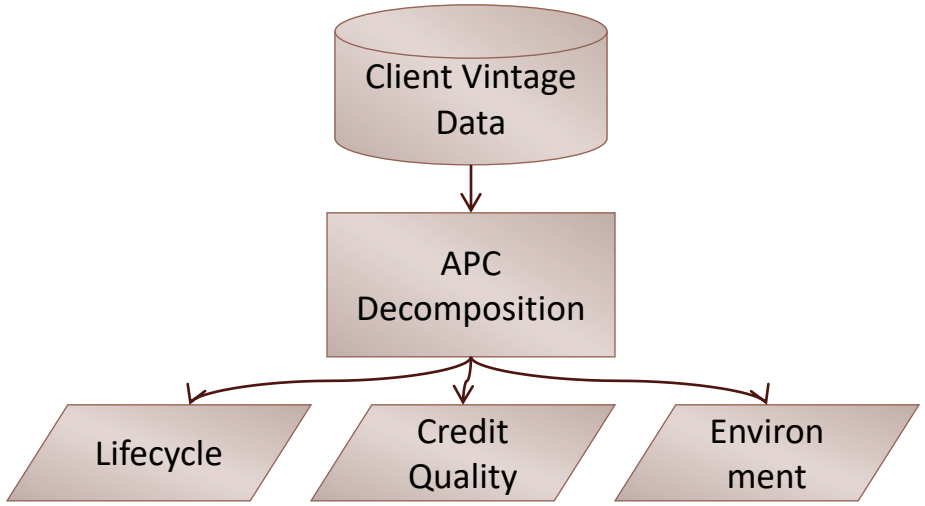
Estimating Age and Calendar Effects

Decomposition:

An APC model is applied to the client's data to measure lifecycle, credit quality, and environment.

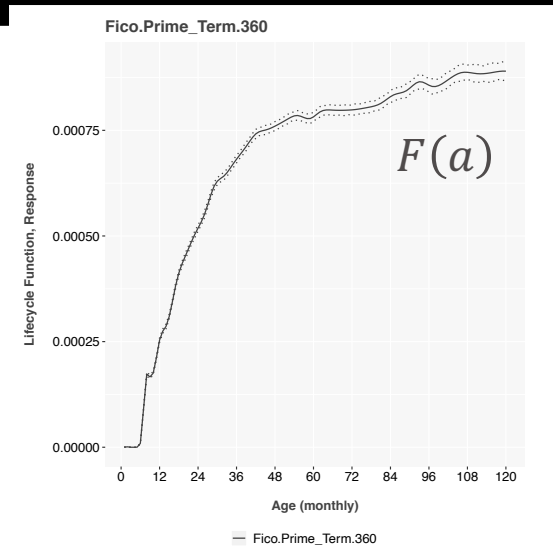
This step uses only aggregate performance data, no behavioral attributes, so it can use older data.

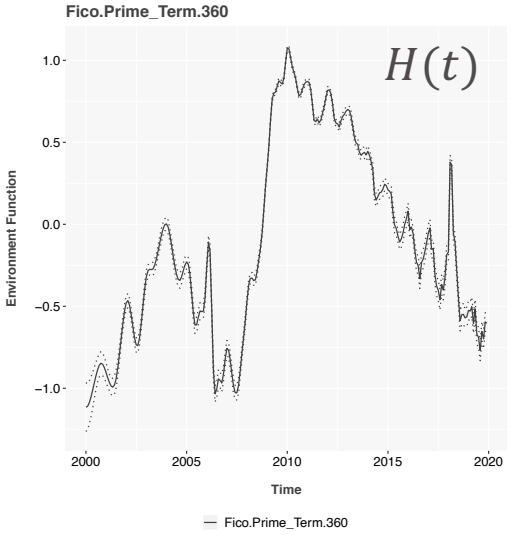
This is done in parallel for PD, PA, EAD, and LGD; although it can be applied to many model structures.

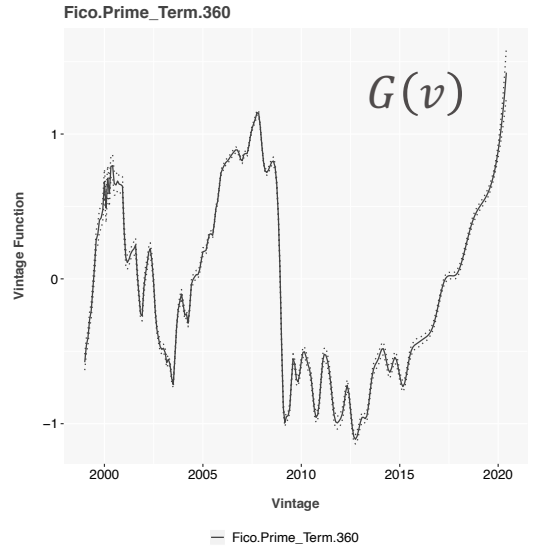


Age-Period-Cohort Analysis of PD

First an Age-Period-Cohort model is created to extract the lifecycle versus account age F(a), credit quality versus vintage date G(v), and environment H(t) versus calendar date.



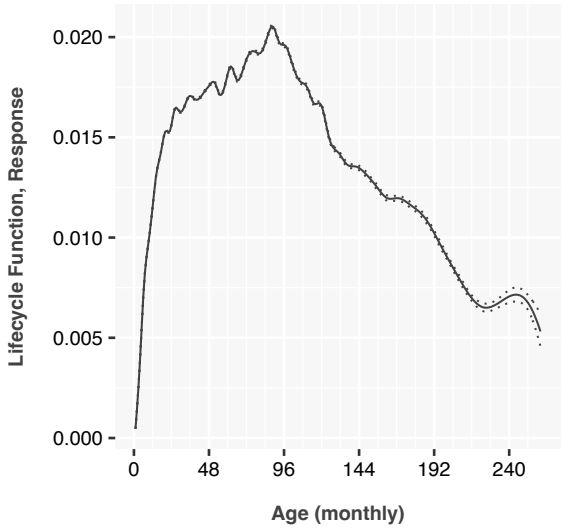


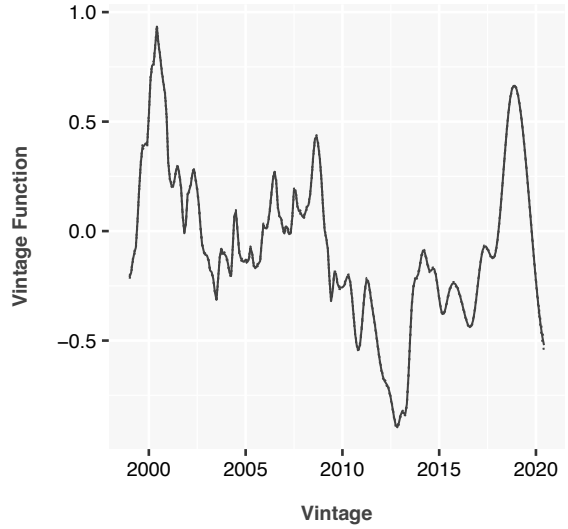


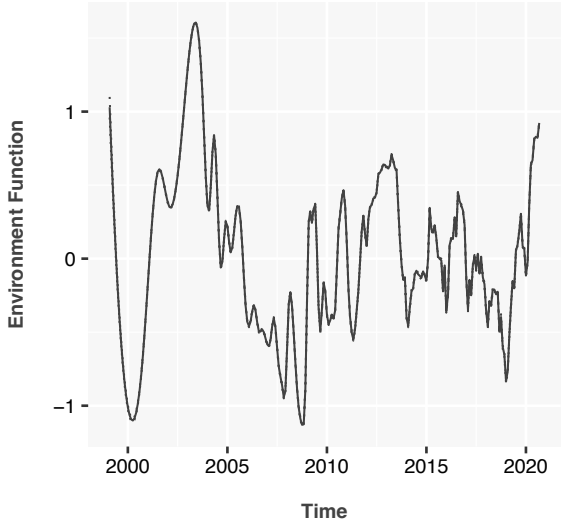
$$\log\left(\frac{PD}{1-PD}\right) = F(a) + G(v) + H(t)$$

APC Decomp of PA

- Follow the same process for prepayment rates







Estimating Scoring Effects

Scoring w/ Decomposition

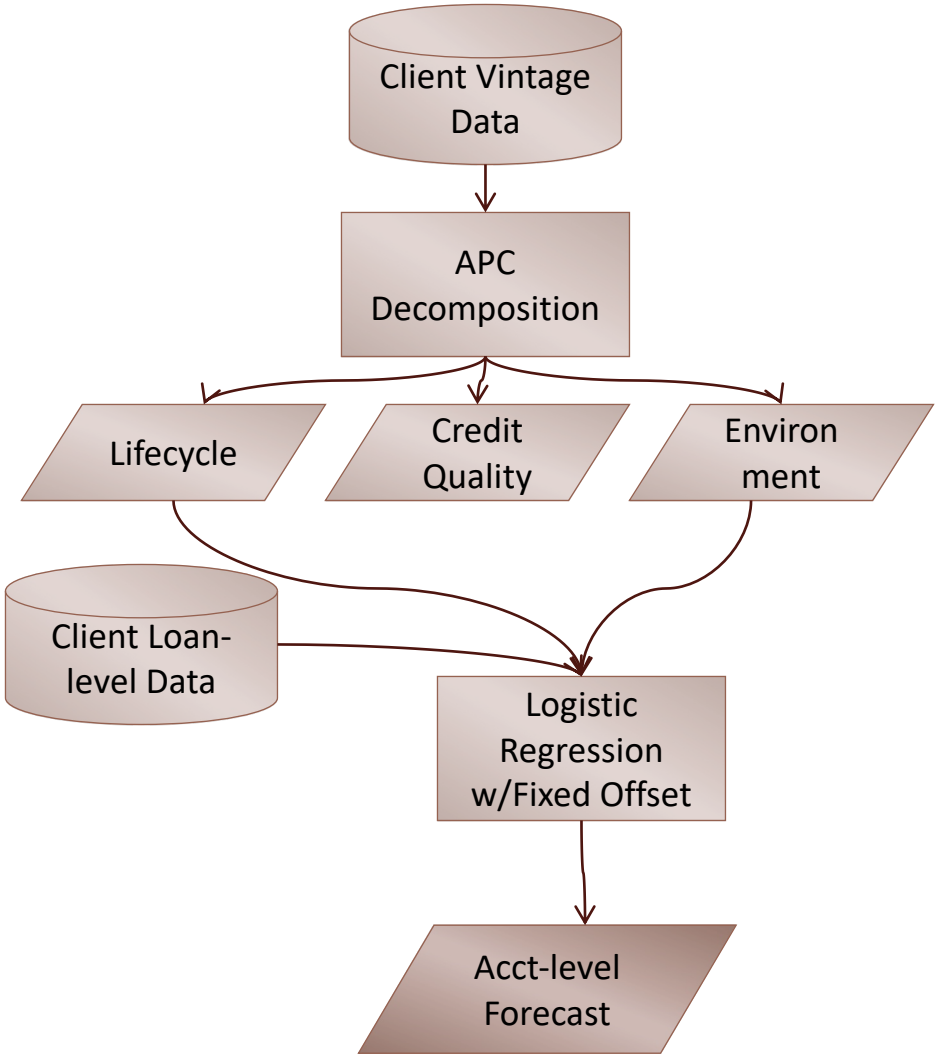
The known lifecycle and environment are taken as inputs (fixed offsets) to a “scoring” model.

The account-level forecast includes everything we know from observable data

Follows the algorithms of

Breeden, J.L. “Incorporating lifecycle and environment in loan-level forecasts and stress tests”, *European Journal of Operational Research*, 255(2), 2016, pp. 649-658.

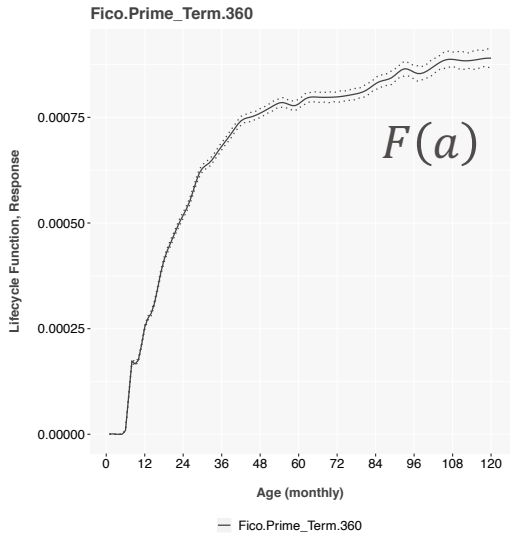
J. Breeden & J. Crook, “Multihorizon discrete time survival models”, *J of the Operational Research Society*, 2020.

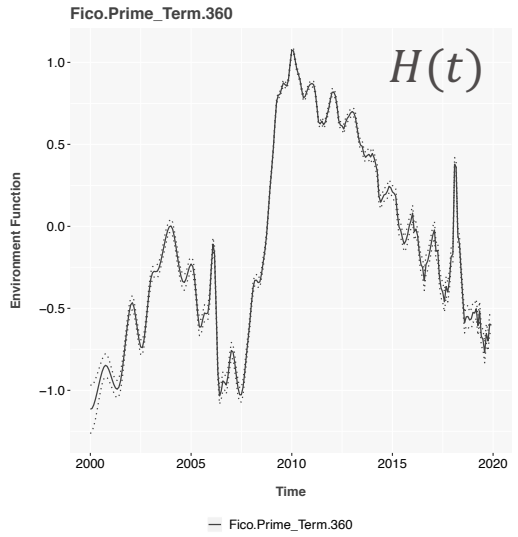


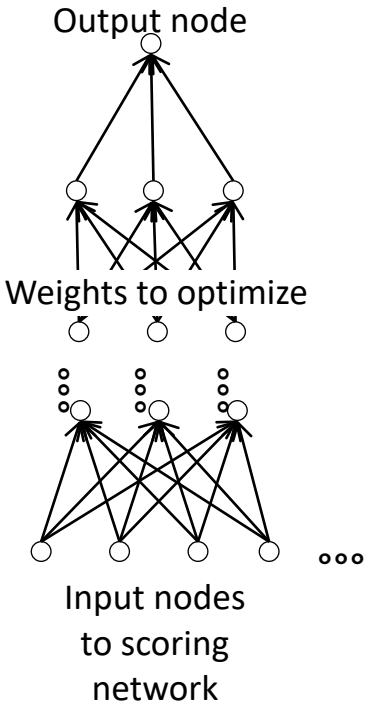
NN + APC Panel Models

$$\log\left(\frac{PD}{1-PD}\right) = F(a) + H(t) + ML(\boldsymbol{x})$$

The second step is to discard the vintage function G(v) and replace it with a neural network built on "Big Data".

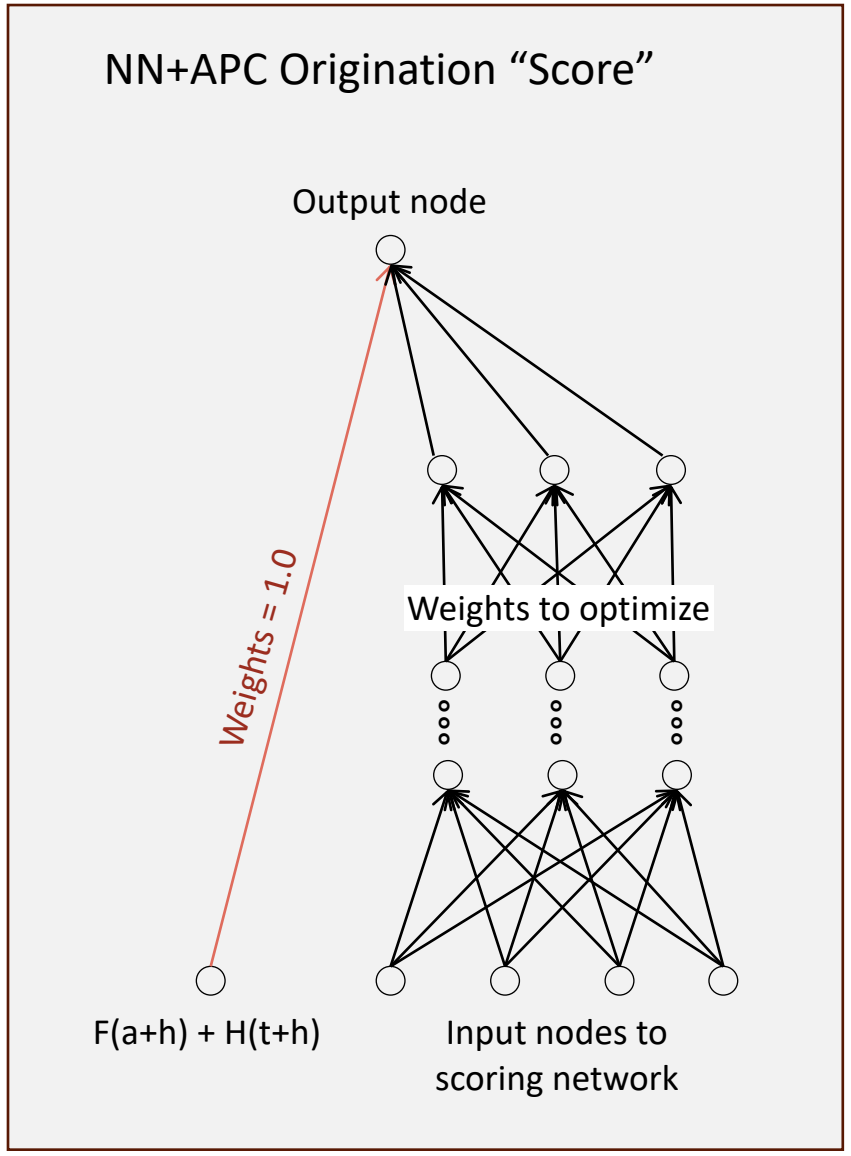


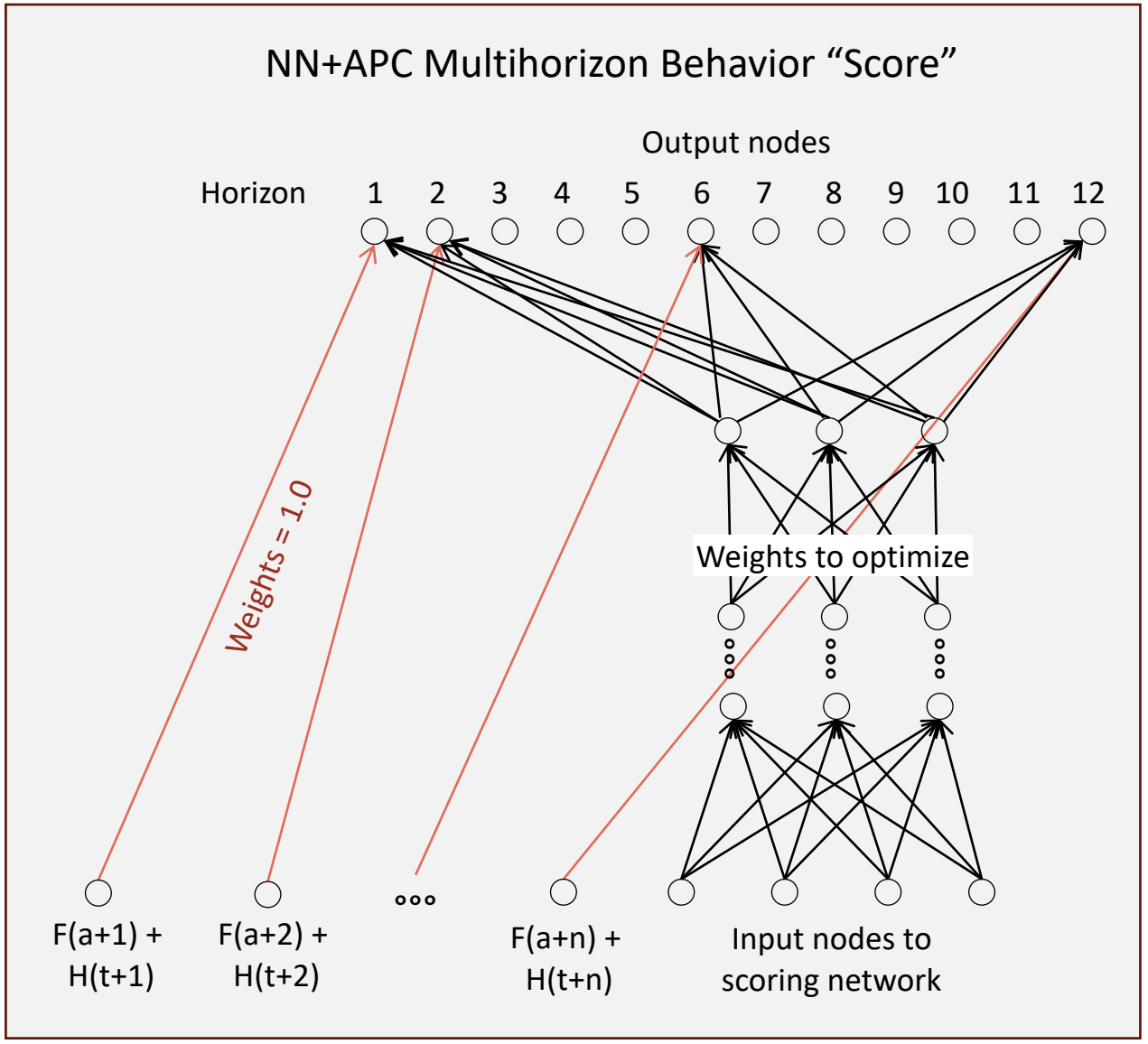




Follows the algorithm of Breeden, J.L. and Y. Leonova, "When Big Data Isn't Enough: Solving the long-range forecasting problem in supervised learning", International Conference on Modeling, Simulation, Optimization and Numerical Techniques, Shenzhen, China, 2019.

NN+APC Architectures

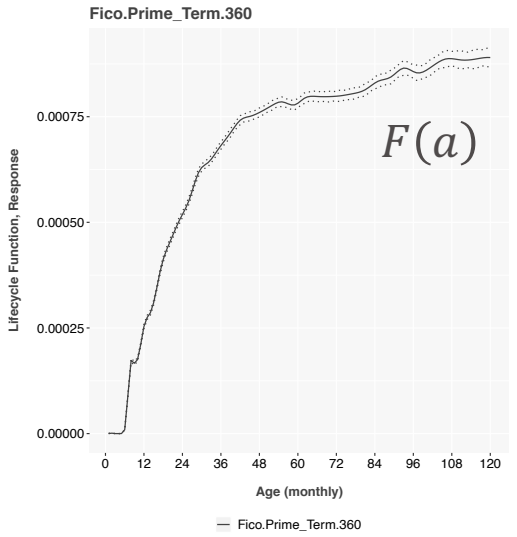


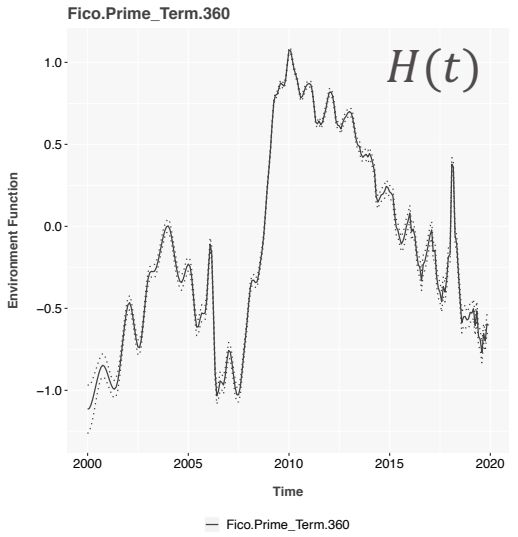


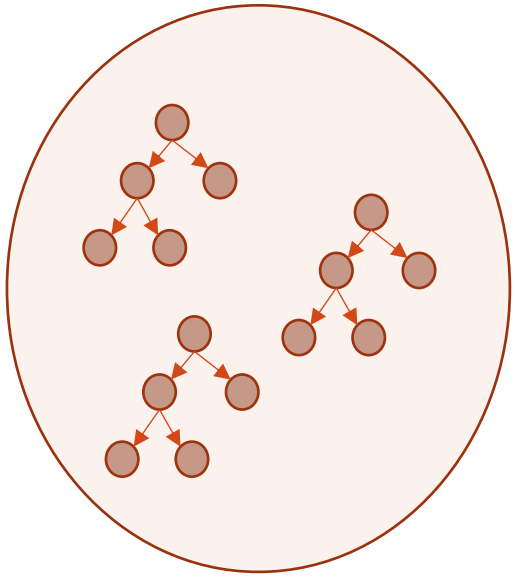
SGBT + APC Panel Models

$$\log\left(\frac{PD}{1-PD}\right) = F(a) + H(t) + ML(\boldsymbol{x})$$

The second step is to discard the vintage function G(v) and replace it with an XGBoost scoring algorithm built on “Big Data”.







Ensemble of Gradient Boosted Trees

Predictive Factors

Variable	Relative Influence
Bureau.Score	29.57
Channel	26.19
Number.Of.Borrowers	12.18
Origination.Balance	10.89
CLTV	7.49
DTI	6.87
PPM.Flag	3.99
Loan.Purpose	0.84
Property.Type	0.82
LTV	0.58
First.Time.Homebuyer.Flag	0.32
Occupancy.Status	0.13
Mortgage.Insurance.Pct	0.11
Number.Of.Units	0.02

- Including APC as inputs did not significantly change the importance of the scoring factors.
- It does change the coefficients.

Discrimination Accuracy – Gini Coefficient

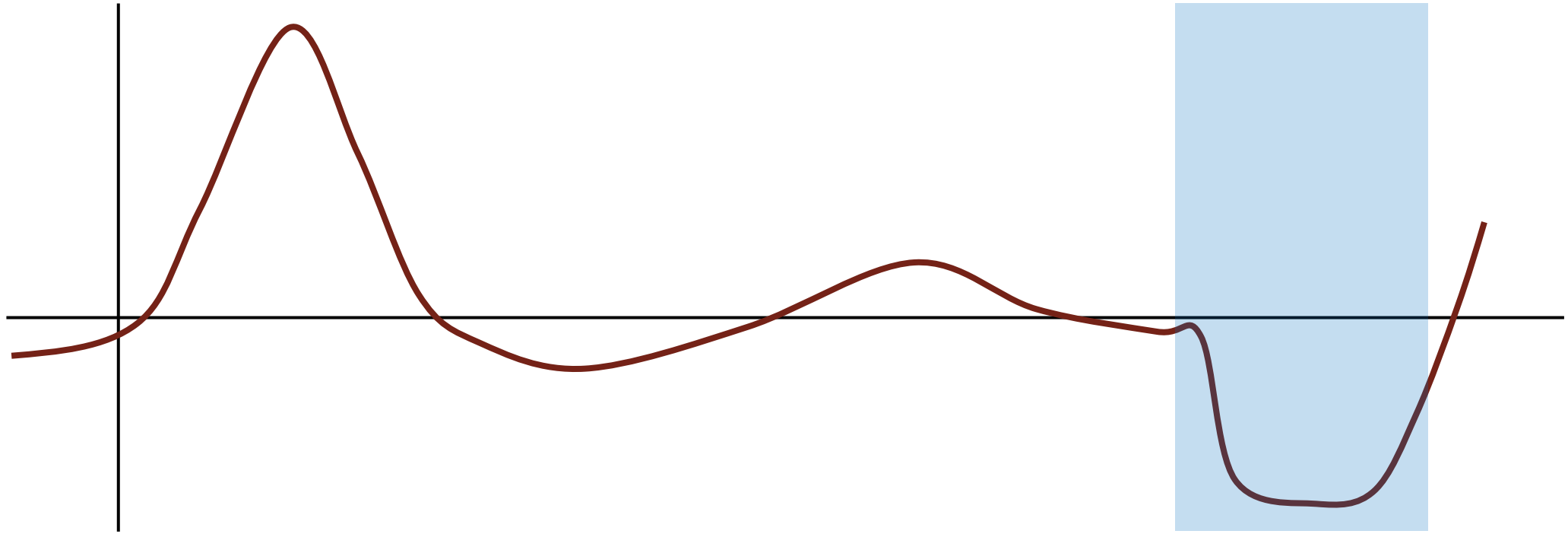
- Data was split in Development / X-Validation / OOS
- The Gini coefficient was computed for each model.
- A better Gini coefficient does not imply better performance through time or by vintage.

Full Period Tests	Date Range	Multihorizon Survival	SGB Trees	SGBT+APC	Neural Network	NN+APC
20% In-Sample	Jan 2015 — Dec 2016	0.64	0.61	0.70	0.48	0.56
80% Out-of-Sample	Jan 2015 — Dec 2016	0.66	0.28	0.69	0.48	0.54
Out-of-Time	Jan 2017 — Dec 2019	0.76	0.40	0.71	0.42	0.59
Annual Tests						
20% In-Sample	Jan 2015 — Dec 2016	0.89	0.85	0.90	0.74	0.65
80% Out-of-Sample	Jan 2015 — Dec 2016	0.89	0.68	0.85	0.74	0.67
Out-of-Time	Jan 2017 — Dec 2019	0.89	0.64	0.86	0.64	0.69

Scoring with Pandemic Data

Options

1. Skip over 2020-2021
2. Normalize 2020-2021



Conclusions

1. Incorporating APC inputs to ML allows scoring methods to be used as time series forecasts.
2. ML + APC has systematically better out-of-sample performance than ML alone.
3. Modeling the long time dynamics first and providing as input to ML stabilizes the ML model and reduces overfitting.

Contact Us

Joseph L. Breeden

breeden@deepfutureanalytics.com

+1-505-670-7670

Deep Future Analytics

1600 Lena St, Ste. E3

Santa Fe, NM 87505

USA