International Association For Quantitative Finance

Welcome

5:45 pm
Registration

6:00 pm
Program

7:30 pm
Reception

For more information on the IAQF visit www.iaqf.org

IAQF & Thalesians Seminar Series:

Agency MBS Prepayment Model using Neural Networks

A Talk by
Joy Zhang

AGENCY MBS PREPAYMENT MODEL USING NEURAL NETWORKS

January 2019

Joy Zhang

MSCI Securitized Products Research

MSCI

SUMMARY: NEURAL NETWORKS AGENCY MBS PREPAYMENT MODEL

Why a machine learning model for Agency MBS?

- Prepayment is a highly complex and non-linear process with idiosyncratic nature
- Recent development in computational hardware enable us to complete large amount of computation in short time
- Machine learning models have excelled in many areas, such as image recognition, natural language processing, fraud detection, etc.

What is the model and what have we learned?

- Deep neural network model applied to pool level agency MBS prepayment data, compared with MSCI1 (the human model)
- Preliminary results show the deep learning model is able to capture very complex prepayment patterns and signals with extremely high computational efficiency

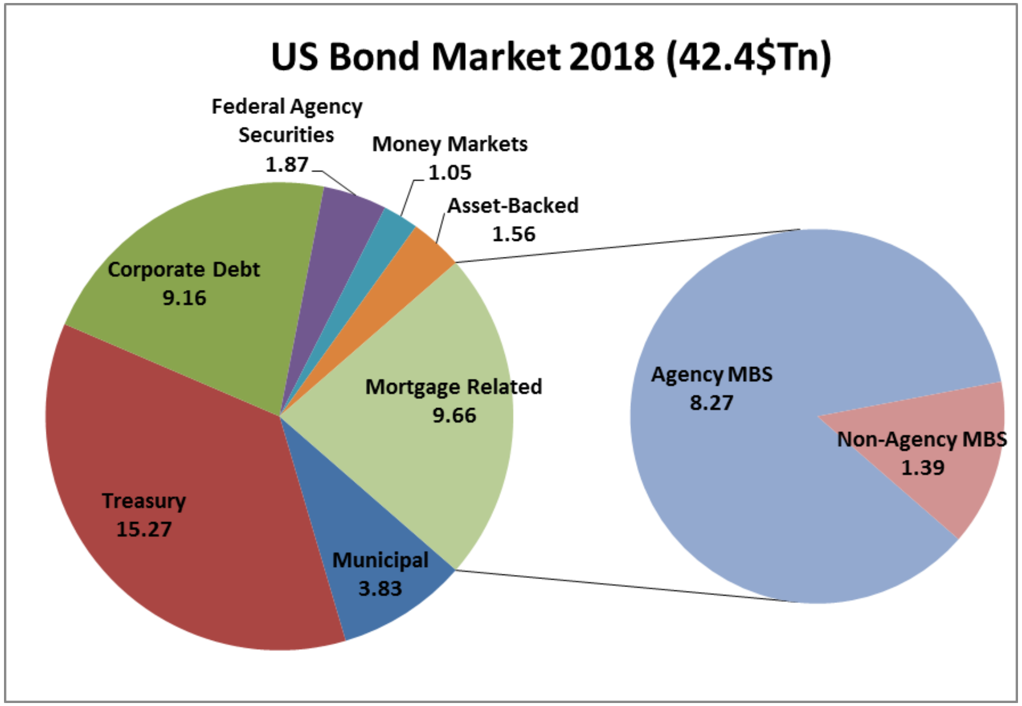
Conclusion and next step

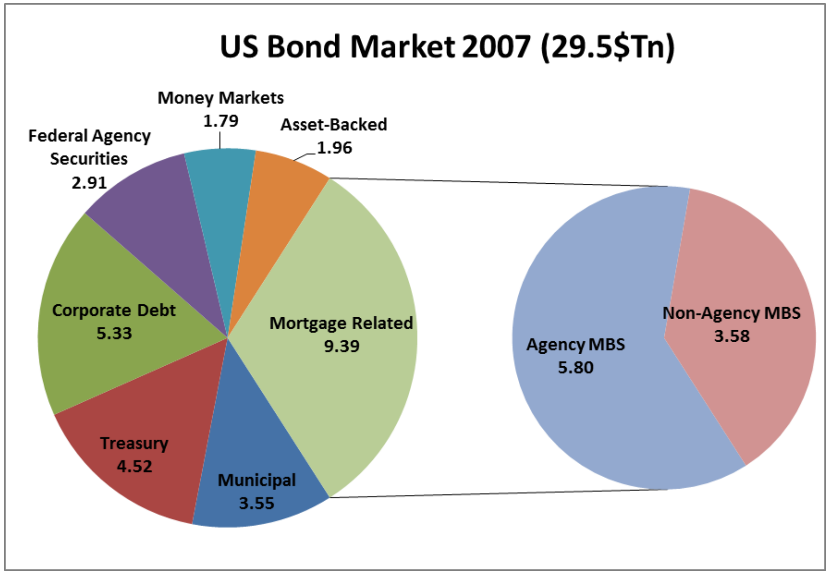
MACHINE LEARNING IN FINANCE

- Consumer credit risk models via Machine-Learning Algorithms (Dr. Andrew Lo, 2010)
  - Using machine-learning model for consumer credit delinquency and default
  - Generalized classification and regression trees
  - Accurately forecasted credit events 3 to 12 months in advance
- Risk and risk management in credit card industry (Dr. Andrew Lo, 2016)
  - Analyzed very large dataset consisting of credit card data from six large banks.
  - Decision trees and random forests model perform better than logistic regression at short time horizon
- Deep learning for mortgage risk (Dr. Kay Giesechke, 2015-2018)
  - Using deep neural network to model mortgage prepayment, delinquency and foreclosure
  - Loan level data
  - Compared NNM with a logit model
- Machine Learning and Alternative Data Approach to Investing (JPM,2017)
  - Comprehensive guide for apply machine learning to solve financial problem

MSCI

US BOND MARKET





MSCI

MODELING OBJECTIVE

Forecast prepayment rate for agency RMBS pools

SMM : Single Monthly Mortality Rate

CPR: Conditional Prepayment Rate

Agencies report previous month's prepayment speed on the 4^{th} business day of each month.

Prepayment types:

- Rate refinance
- House turnover
- Cash-out
- Curtailment
- Buyout

Difficulties with mortgage prepayment modeling

- Large data sets: ~20-2000 G data, Agency MBS covers ~400,000 pools/100^{+}mm loans performance over 20-30 years, pool/loan variables ~30-100
- Multiple, highly non-linear and interactive risk drivers ("layered risk")
  - Loan size vs. prepayment is function of moneyness
  - Age vs. prepayment is function of past moneyness history
  - Loan purpose (refi vs purchase) vs. prepayment is function of origination year
  - ….
- Regime changes
  - Mortgage credit and borrower risk appetite cycles, business practice and policies can all affect absolute level and risk drivers for prepayment/default

MORTGAGE PREPAYMENT MODELING: SCIENCE AND CRAFTSMANSHIP

Exhibit 4: Agency MBS prepayment regimes since 2003

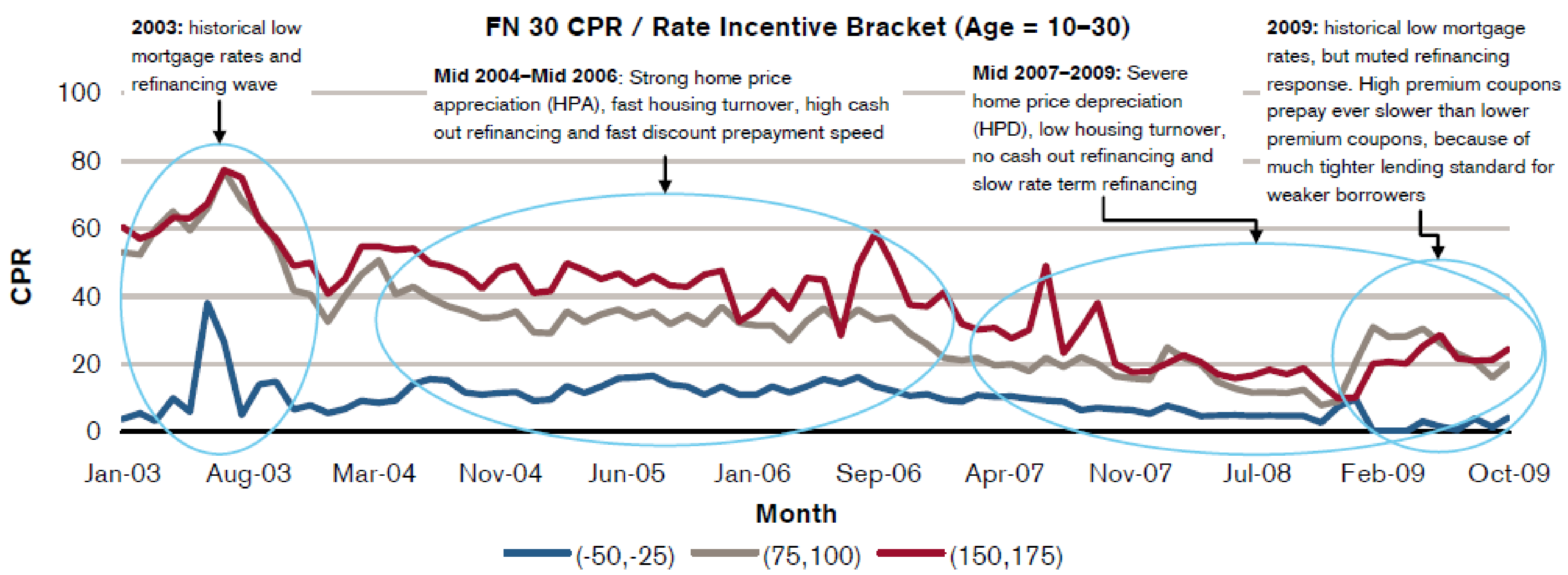
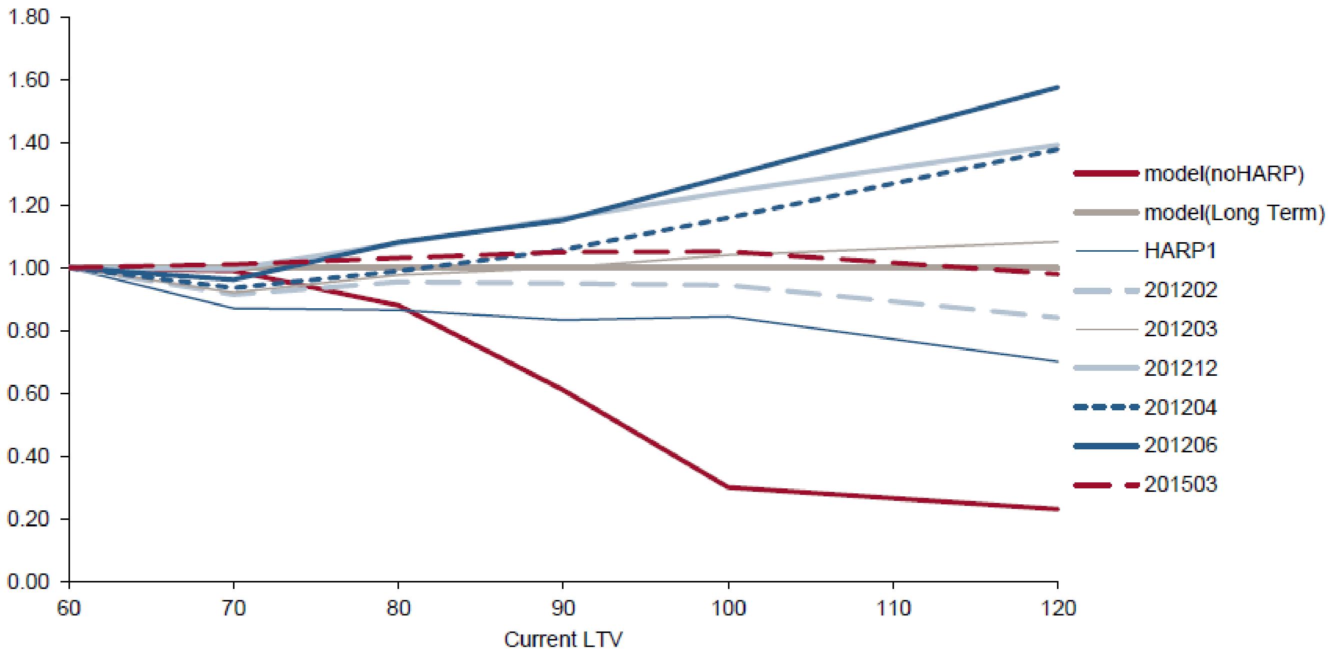


Exhibit 5: HARP CLTV curve history and long term model assumptions

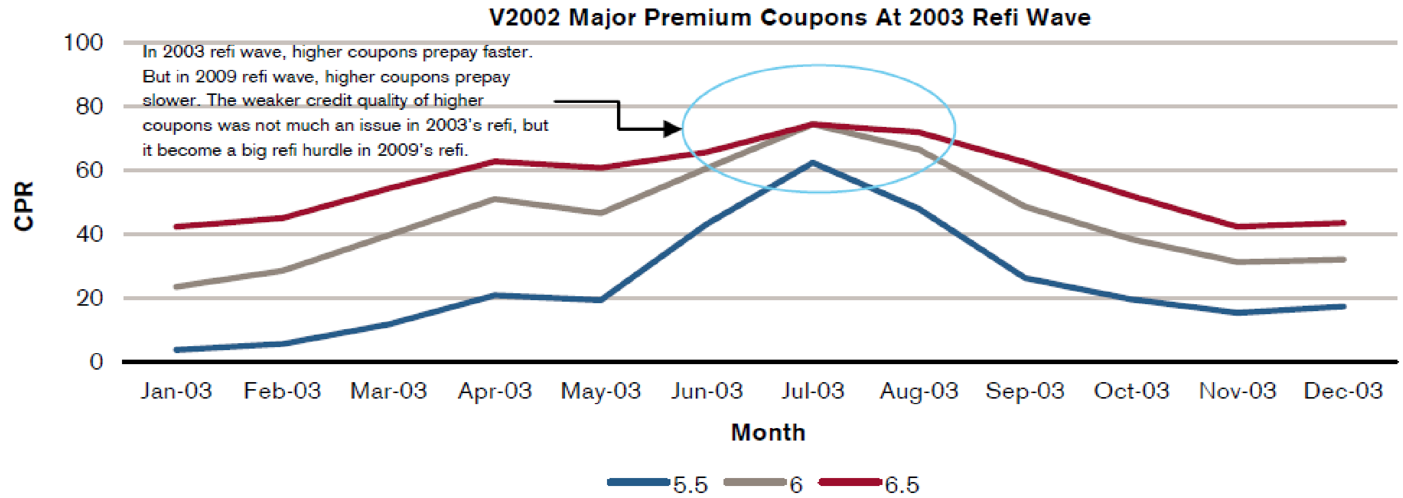
The CLTV curve represents the ratio of refinance speeds across CLTV spectrum, using sub-50 CLTV cohort as benchmark, with all other pool variables (for example, loan size, moneyness, FICO, etc.) holding constant

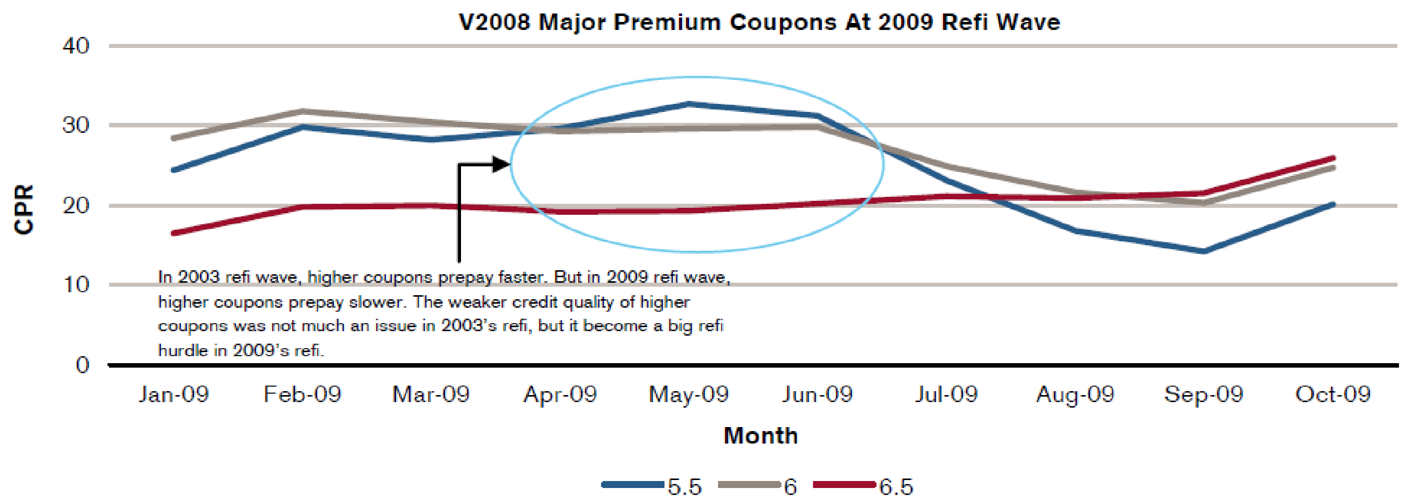


The HARP program caused temporary inversion of the CLTV Prepayment Curve

HARP: Home Affordable Refinance Program

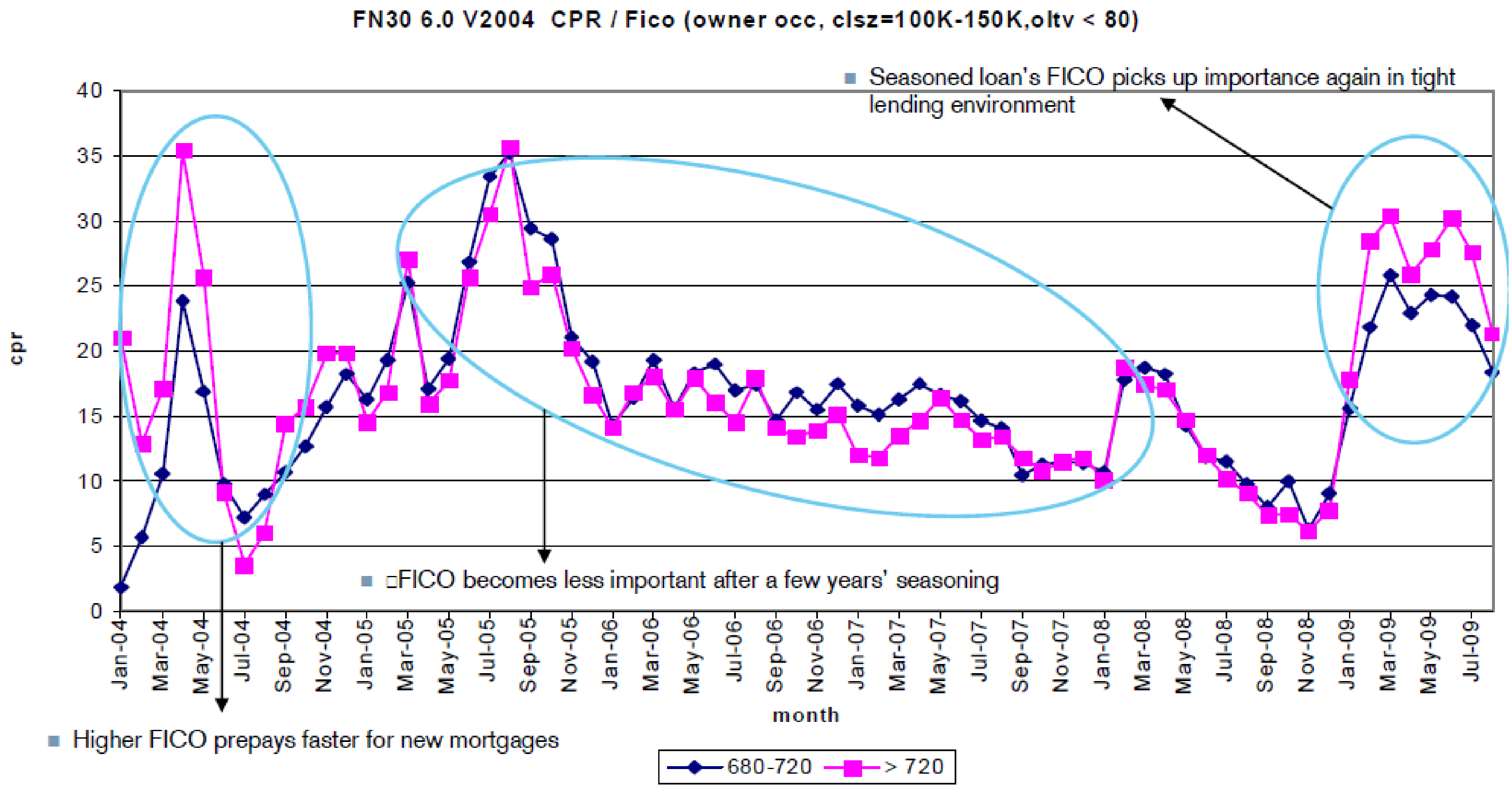
MORTGAGE PREPAYMENT MODELING: SCIENCE AND CRAFTSMANSHIP





MSCI

MORTGAGE PREPAYMENT MODELING: SCIENCE AND CRAFTSMANSHIP



Example of modeling:

Assume ppm (pool, time) = f(X1, X2, X3,…. Xn) …

start by assuming separable risk factor: ppm= f1(x1)*f2(x2)…. Until (often) proven incorrect…

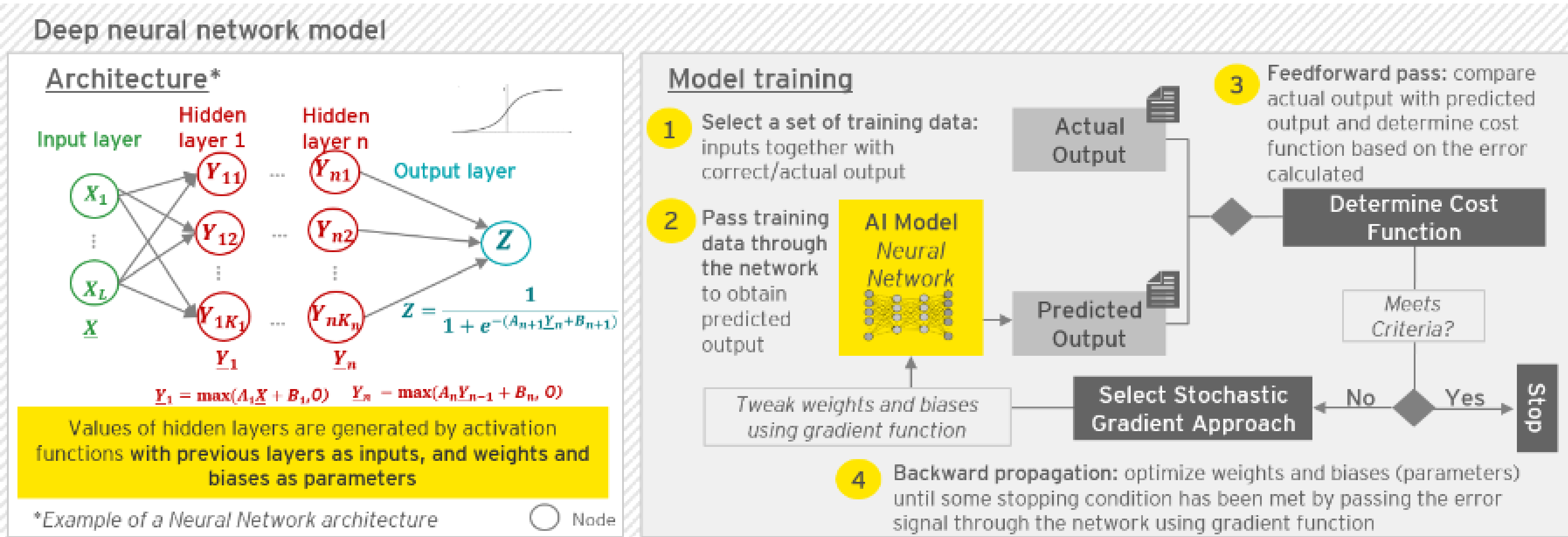
estimating f1(x1) by "building cohort", by bucketing loans/pools for groups of x1, but similar x2, x3….
(this further assumes quasi linear property of x2, x3…. Average(f2(x2) f3(x3)…)= f2(ave(x2))* f3(ave(x3))….

….. Checking overall fit after all Xn are fitted, adding extra variables to deal with non-linear and interactive variables… this often does not lead to convergence …

- Time consuming and non-standard approaches
- Experience and step-by-step / regime-by-regime progress are valued

- Can new techniques in machine-learning modeling provide the much needed disruption?

FEED FORWARD NEURAL NETWORK



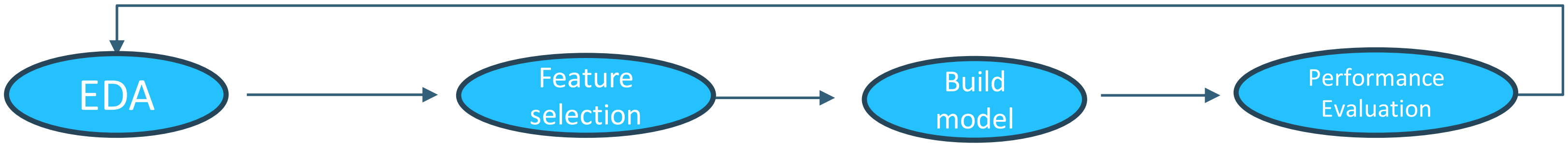
Network architecture:

- Hyper-parameters
  - Number of layers, batch size, learning rate, max-norm constraint, dropout rate, …

Ensemble techniques:

- Bagging: minimum MSE of different realizations and neural networks
- Boosting: Fine tune a neural network via changing a few hyper-parameters

BUILDING NEURAL NETWORK MODEL



EDA
1, sanity check
2. data cleansing
3. data transformation

Feature selection
1,information value
2.correlation matrix
3.domain knowledge

Build model
1. link weights
2. hyper-parameters

Performance Evaluation
1,error tracking
2.sensitivity
3. transparency tool (deeplift)

Deep neural network fitting

2003-2018 30yr agency MBS data (~25G data)

30+ input variables: pool attributes, macro-economic variables

To reduce complexity of machine model, we added incentive, 1 regime indicators, and 1 policy indicator (HARP)

1 round of fitting can be completed in ~ 3 hours on a GPU machine

MSCI

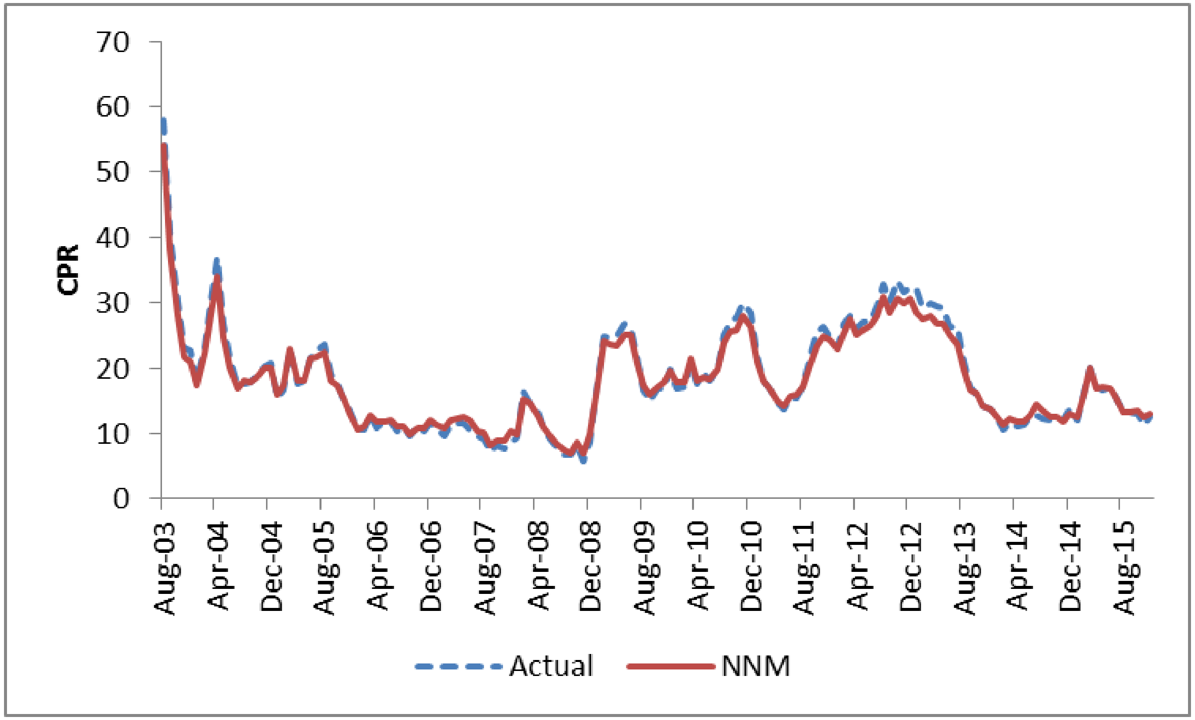
MODEL DRIVERS

Independent variables	
WALA	Weighted Average Loan Age
WAC	Weighted Average Coupon
CLNSZ	Current Average Loan Size
OLTV	Original Loan to Value
Refi%	Percentage of Refinanced Loans by UPB
SecHome%	Percentage of Second Home Loans by UPB
MultiFamily%	Percentage of Muti Family Loans by UPB
Investor%	Percentage of Investor Loans by UPB
TPO%	Percentage of Third party origination by UPB
AOL	Original Average Loan Size
LNSZ_Q4	Max original loan size
LNSZ_Q3	Max original Loan Size - 3rd Quartile
LNSZ_Q1	Max original Loan Size - 1st Quartile
Geo_CA%	Percentage of California Loans by UPB
Geo_FL%	Percentage of Florida Loans by UPB
Geo_TX%	Percentage of Taxas Loans by UPB
Geo_NY%	Percentage of New York Loans by UPB
Geo_NE%	Percentage of New England Region Loans by UPB
Geo_NO%	Percentage of North Region Loans by UPB
Geo_SO%	Percentage of South region Loans by UPB
Geo_PC%	Percentage of Pacific region Loans by UPB
Geo_AT%	Percentage of Atlantic region Loans by UPB
Geo_NONUS%	Percentage of non-US region Loans by UPB
Seasonality	Calendar month

Derived Variables	
Incentive	WAC - Mortgage Rate(t)
Rolling Incentive	Average Incentive (20month) $\sum_{t=1}^{t=\min(20,wala)} Incentive/\min(20,w$
Loan size dispersion	(LNSZ_Q3-LNSZ_Q1)/AOL
SATO	Spread-at_origination = WAC - Mortgage Rate(0)
HPA	House Price Appreciation (HPI(t)/HPI(0)-1)
HARP-able	and Dec. 2011 2: IssueMonth <= Jun. 2009 and factor date > Dec. 2011
HARP-ed	Refi% = 100 and OLTV > 80 and issueMonth > Jun. 2009
Underwritting standard	0: before 2008, 1: after 2008
Weight	
cBal	Current Balance
Dependent Variable	
Prepayment speed	Prepayment speed in SMM

MSCI

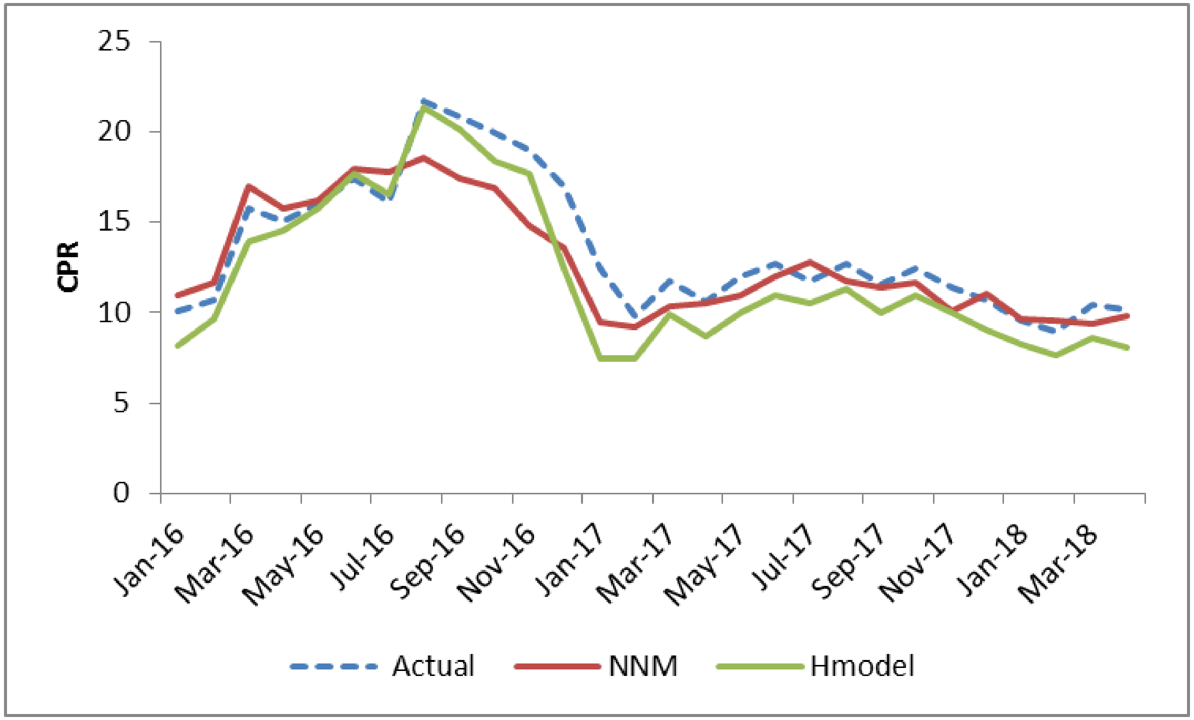
AGENCY 30YR UNIVERSE SPEEDS ERROR TRACKING



- Training Data: 2003 – Dec. 2015. Random sample 10% pools.
- Error tracking is generated using out-of-the-sample pools.

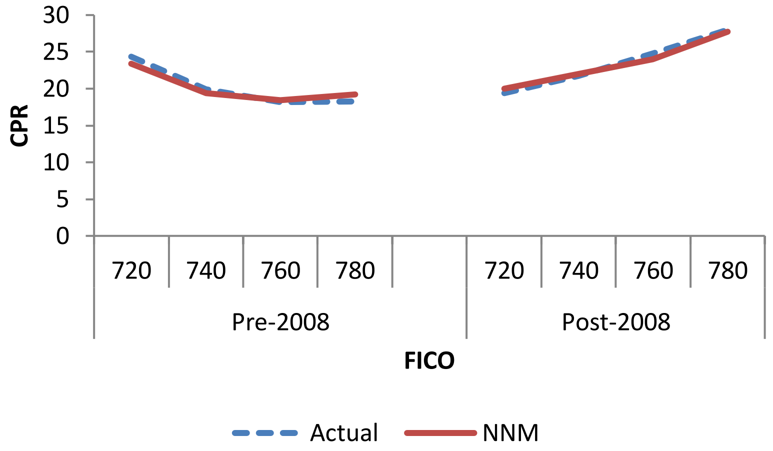
MSCI

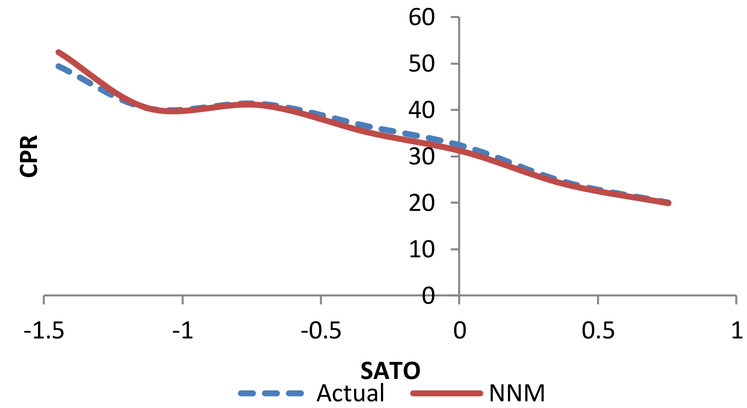
OUT-OF-SAMPLE FORECASTS

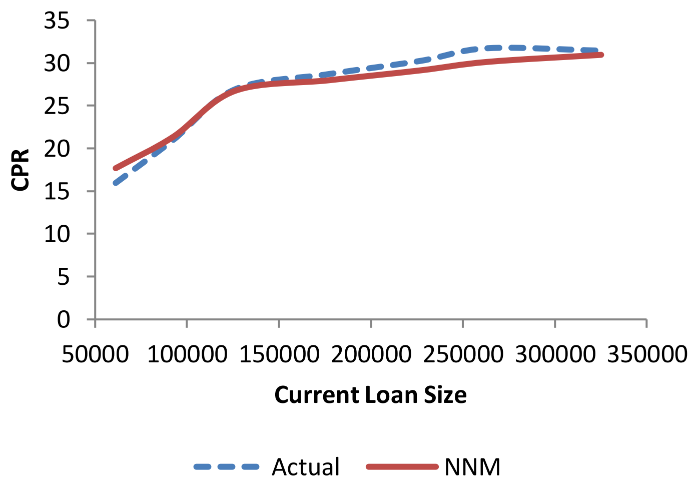


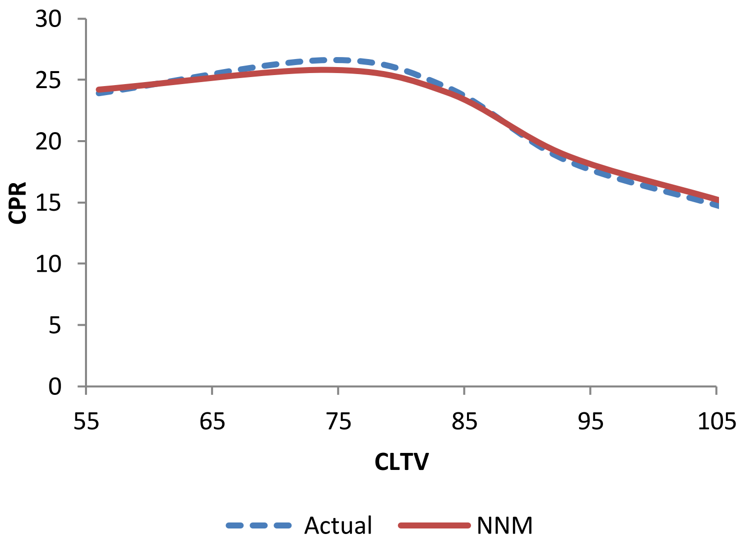
- True out-of-time and out-of-sample test.
- Overall fitting is good in out-of-sample test
- Missed the refi wave in second half of 2016

MODEL RISK FACTORS



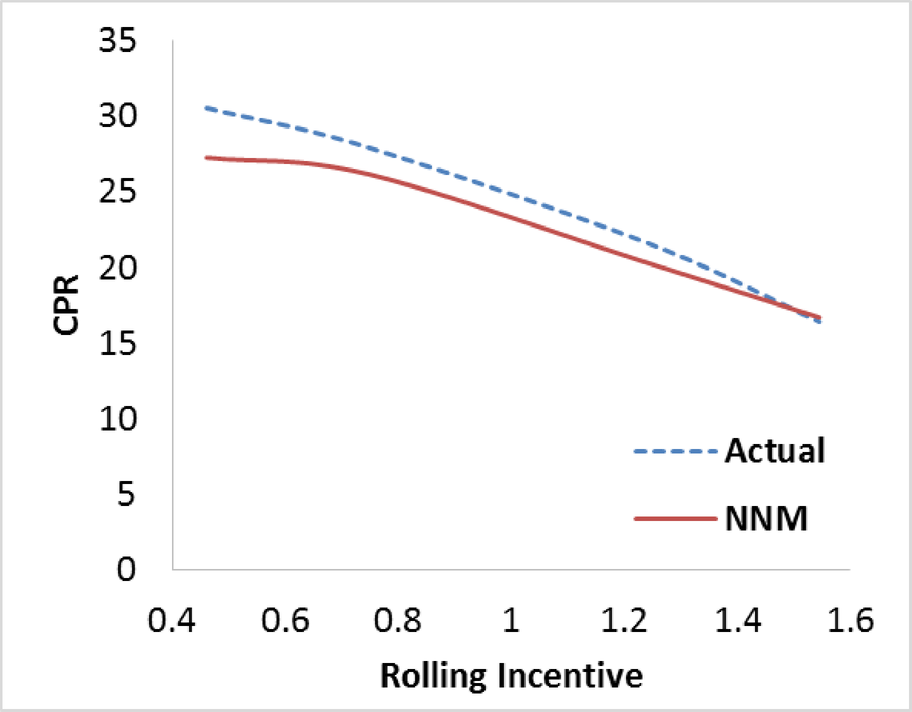






MSCI

MODEL BURNOUT

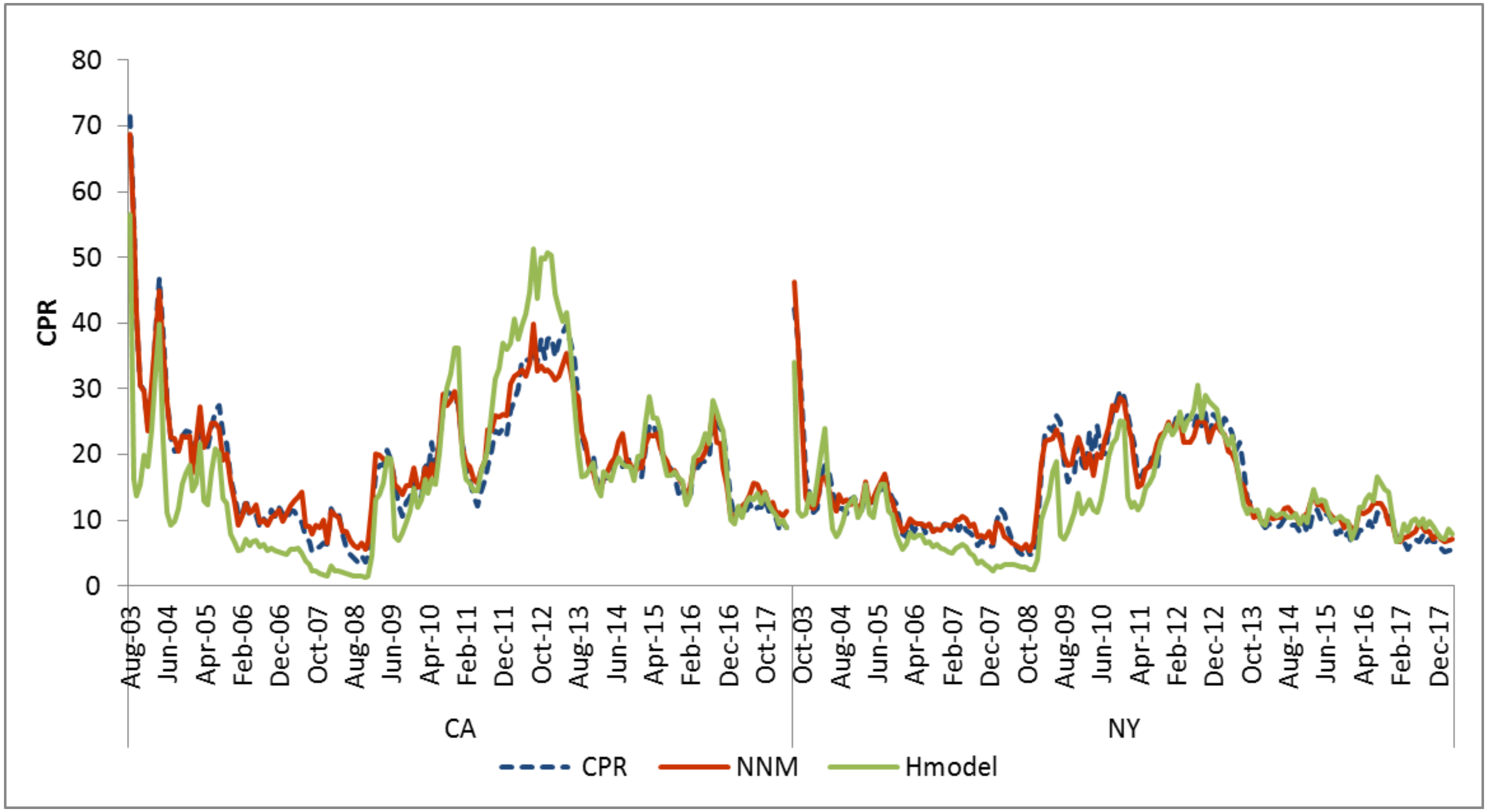


NNM and actual prepayment speeds against average incentive in prior 20 months

MSCI

MODEL POOL VARIABLES VS "HUMAN"MODEL

NNM and Hmodel Error Tracking against State Variables

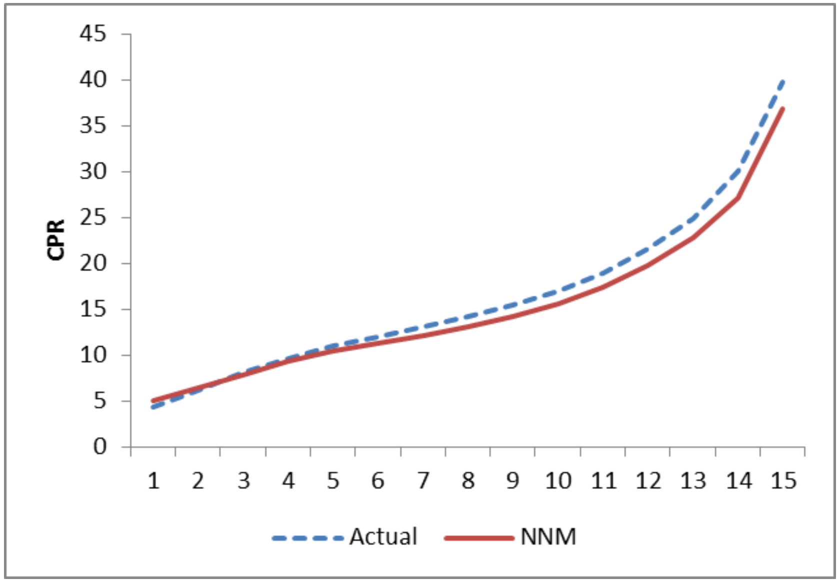


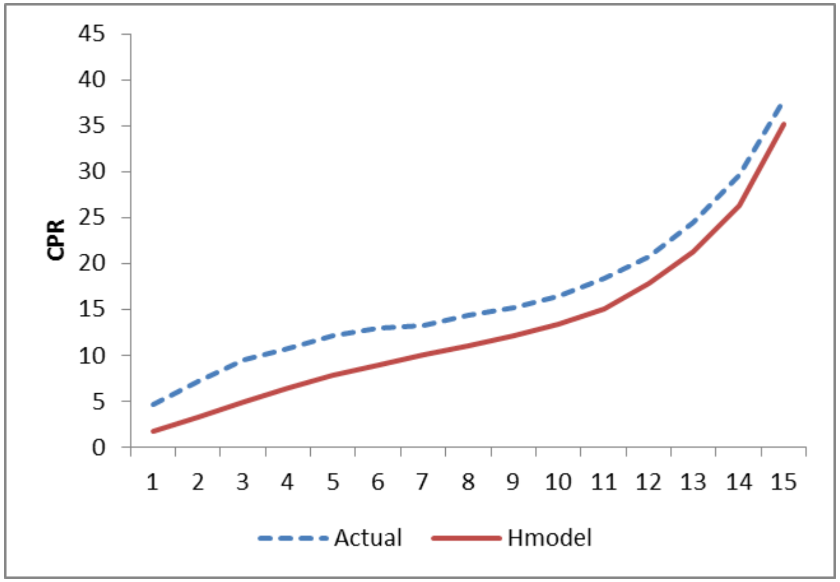
NNM accurately captured state-level prepayment behaviors

MSCI

MODEL POOL VARIABLES VS HUMAN MODEL

Ranking-Based Sample Error Tracking for Coupon 4s

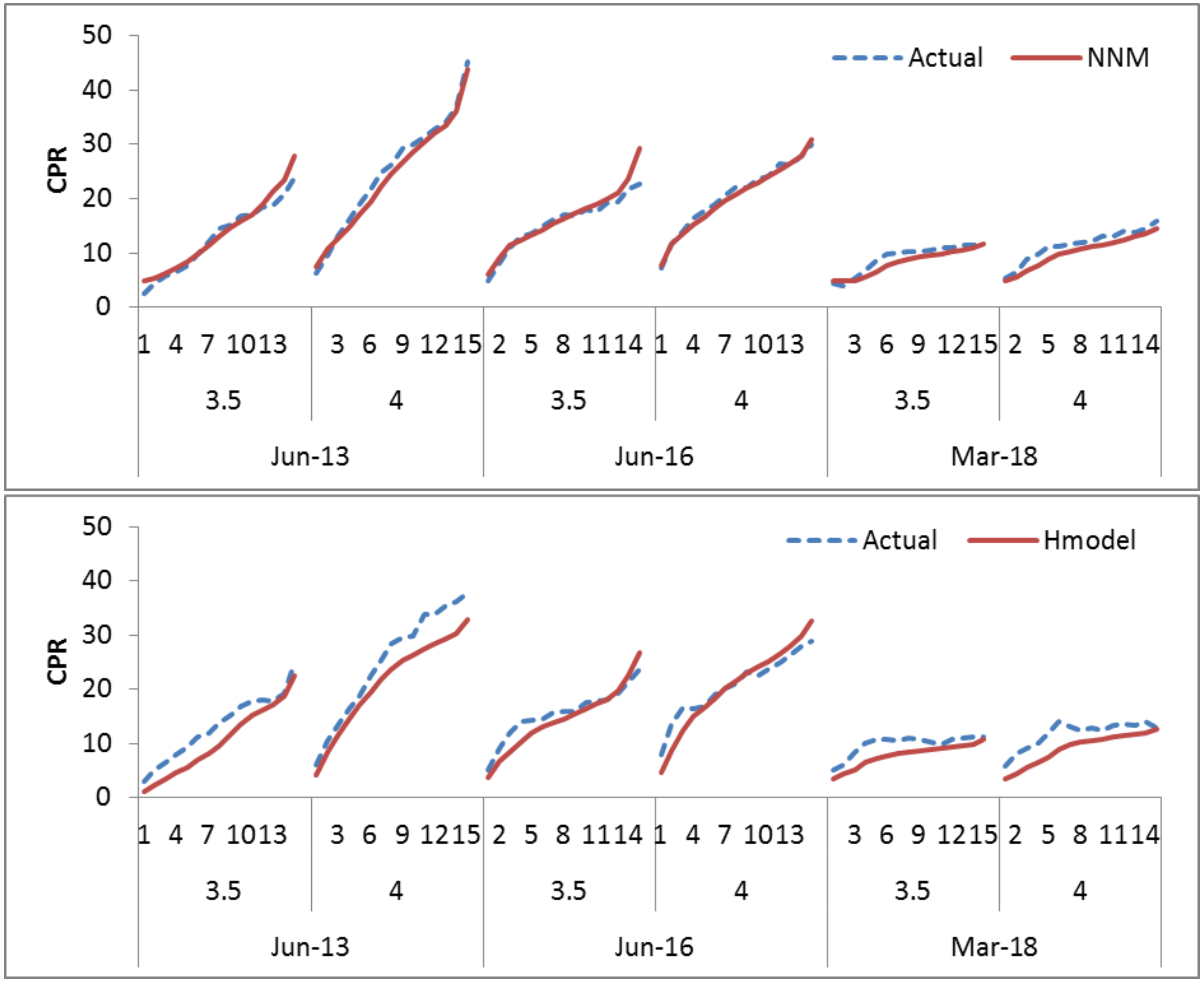




- Ranking based error tracking methodology provides a comprehensive measure of model accuracy across all pool variables
- NNM performed better than Hmodel

MSCI

MODEL POOL VARIABLES VS HUMAN MODEL

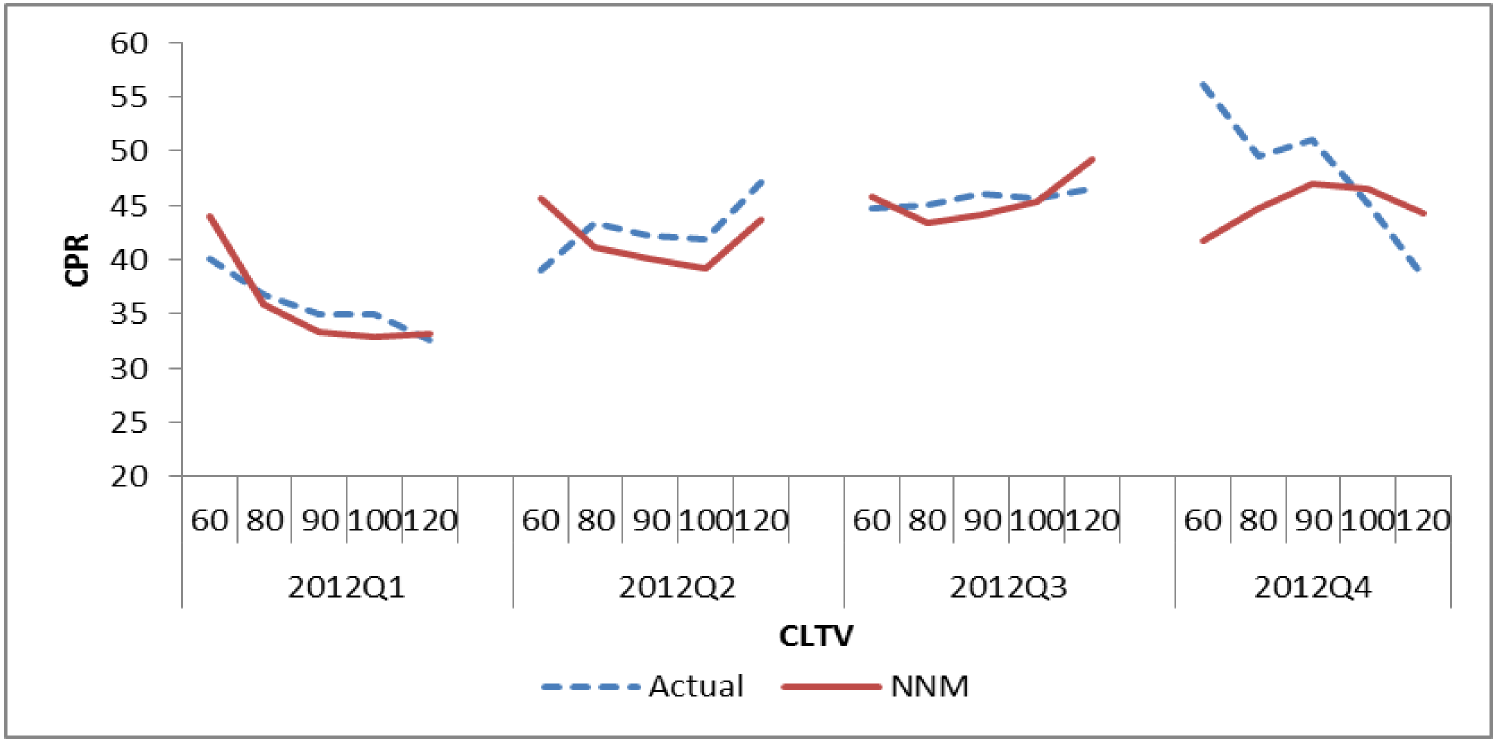


Sample ranking-based error tracking at different time point

MSCI

MODEL HARP EFFECTIVENESS

Error Tracking against HARP effectiveness across CLTV Cohorts



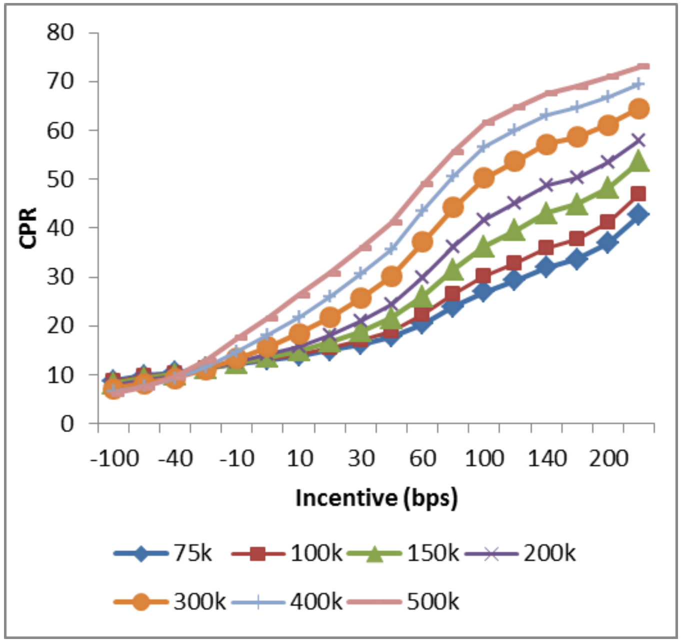
NNM is able to pick up the general trend of HARP effectiveness but missed the complexity of its revolution

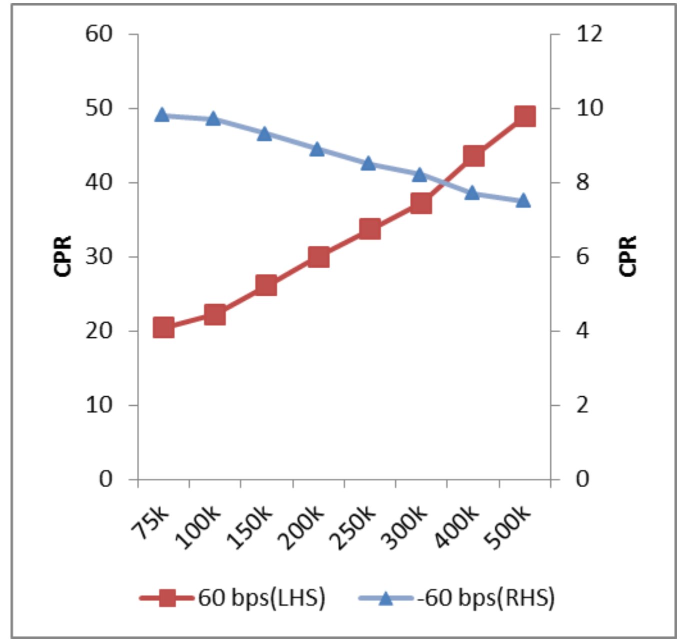
MSCI

Back

MODEL SENSITIVITY

Model prepayment sensitivity to loan sizes and refinance Incentives





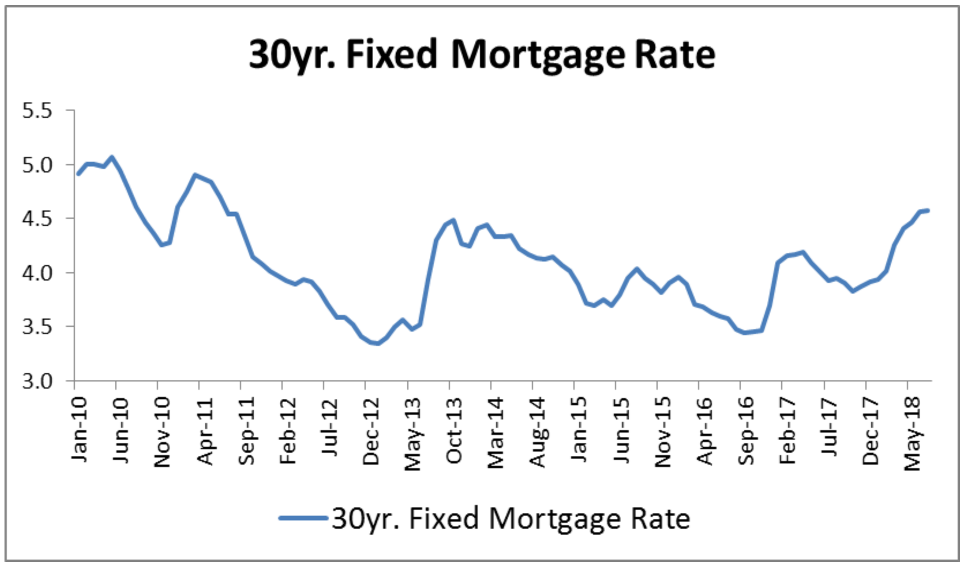
NNM captured the prepayment behavior for loan size

"MEDIA EFFECT"

FH 2011 3.5 vs 2010 4 comparisons, across TPO/Retail and Refi/Purchase combinations									
Cohort	Observation Range	CPR	WALA	SATO	CLTV	CLNSZ	Incentive	FICO	Avg.UPB(bn)
		Purchase/Retail							
FH 3.5 2011	Jul.12 - Dec. 12	16.1	13	-5	77	212258	52	770	2.91
FH 4 2010	Nov. 11 - Feb. 12	13.9	15	3	78	201901	45	767	6.26
		Purchase/TPO							
FH 3.5 2011	Jul.2012 - Dec. 12	21.9	12	-3	76	235847	50	770	4.04
FH 4 2010	Nov. 11 - Feb. 12	16.4	16	3	78	224734	45	765	8.66
		Refi/Retail							
FH 3.5 2011	Jul.12 - Dec. 12	29.2	12	-2	66	216270	54	771	7.31
FH 4 2010	Nov. 11 - Feb. 12	15.3	15	11	70	208962	52	766	30.89
		Refi/TPO							
FH 3.5 2011	Jul.12 - Dec. 12	46.1	12	-8	64	269298	46	773	9.58
FH 4 2010	Nov. 11 - Feb. 12	26.2	15	2	69	245496	44	767	23.02

2011 3.5s and 2010 4s prepayment speeds are compared across loan attributes, loan purpose and origination channel

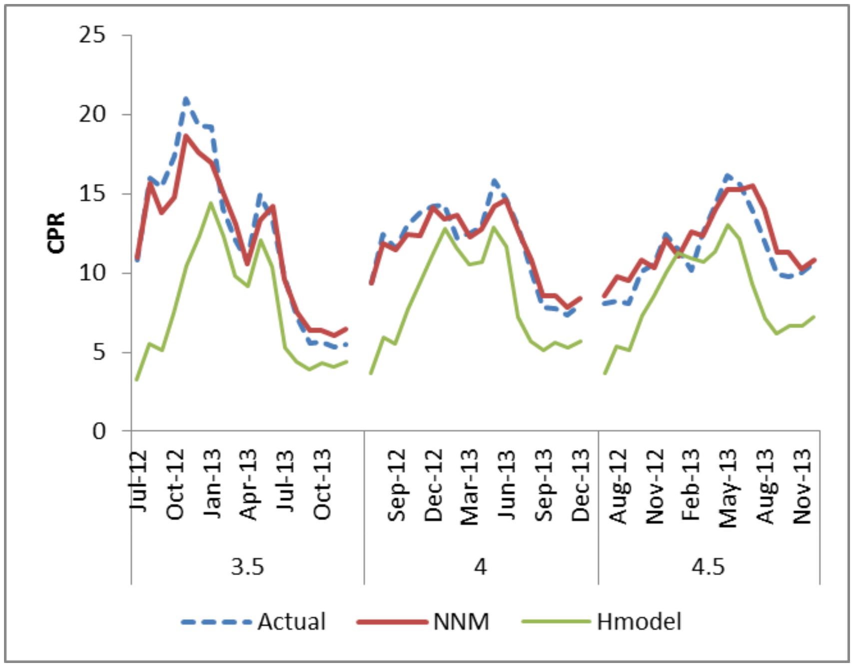
3.5s is much faster than 4s given similar loan attributes and incentive



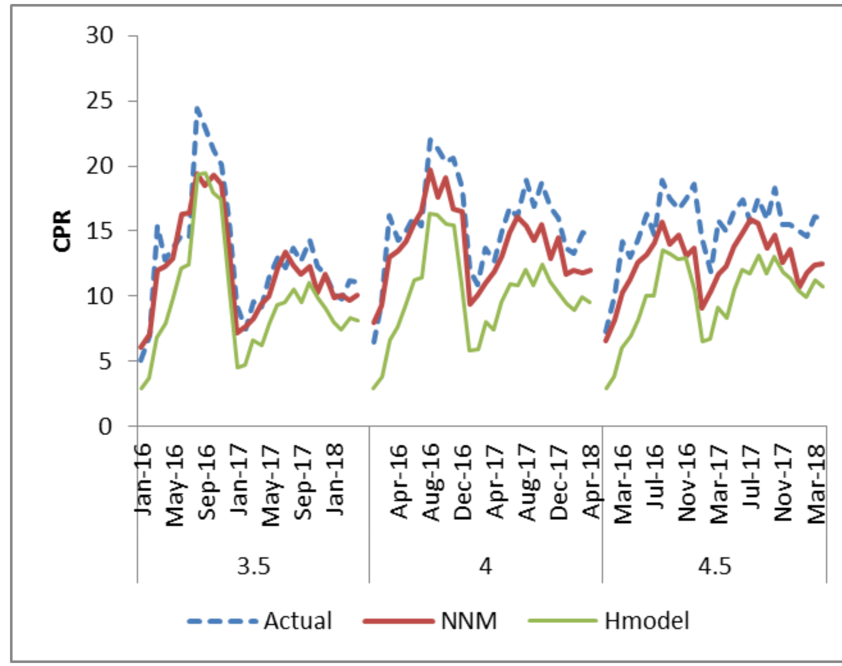
MSCI

MODEL "MEDIA EFFECT"

2012 Vintage in 2012 refinance wave



2015 Vintage in 2016 refinance wave



MSCI

CONCLUSION

NN model vs. "Human Model"

- Accurate forecast and prepayment anomalies flag
- Accurate model of large numbers of risk factors
- Accurate model of highly non-linear and interactive risk factors
- Highly efficient modeling process - hundreds times of increases in modeling efficiency

Next step

- Apply state-of-the-art methodology/techniques to gain model transparency
- Addressing model overfitting, true out-sample and model specification risk issues
- Incorporate neural networks in Agency MBS modeling practice

ACKNOWLEDGEMENT

David Zhang:

Managing Director, Securitized Product Research, MSCI

Jan Zhao:

Principal, Advanced Analytics, Ernst & Young

Fei Teng:

Senior Quantitative Analysts, Quantitative Advisory Services , Ernst & Young

Siyu Lin:

Senior Quantitative Analysts, Quantitative Advisory Services , Ernst & Young

Henry Li:

Executive Director, Quantitative Advisory Services, Ernst & Young

NOTICE AND DISCLAIMER

This document and all of the information contained in it, including without limitation all text, data, graphs, charts (collectively, the “Information”) is the property of MSCI Inc. or its subsidiaries (collectively, “MSCI”), or MSCI’s licensors, direct or indirect suppliers or any third party involved in making or compiling any Information (collectively, with MSCI, the “Information Providers”) and is provided for informational purposes only. The Information may not be modified, reverse-engineered, reproduced or redisseminated in whole or in part without prior written permission from MSCI.

The Information may not be used to create derivative works or to verify or correct other data or information. For example (but without limitation), the Information may not be used to create indexes, databases, risk models, analytics, software, or in connection with the issuing, offering, sponsoring, managing or marketing of any securities, portfolios, financial products or other investment vehicles utilizing or based on, linked to, tracking or otherwise derived from the Information or any other MSCI data, information, products or services.

The user of the Information assumes the entire risk of any use it may make or permit to be made of the Information. NONE OF THE INFORMATION PROVIDERS MAKES ANY EXPRESS OR IMPLIED WARRANTIES OR REPRESENTATIONS WITH RESPECT TO THE INFORMATION (OR THE RESULTS TO BE OBTAINED BY THE USE THEREOF), AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, EACH INFORMATION PROVIDER EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES (INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTIES OF ORIGINALITY, ACCURACY, TIMELINESS, NON-INFRINGEMENT, COMPLETENESS, MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE) WITH RESPECT TO ANY OF THE INFORMATION.

Without limiting any of the foregoing and to the maximum extent permitted by applicable law, in no event shall any Information Provider have any liability regarding any of the Information for any direct, indirect, special, punitive, consequential (including lost profits) or any other damages even if notified of the possibility of such damages. The foregoing shall not exclude or limit any liability that may not by applicable law be excluded or limited, including without limitation (as applicable), any liability for death or personal injury to the extent that such injury results from the negligence or willful default of itself, its servants, agents or sub-contractors.

Information containing any historical information, data or analysis should not be taken as an indication or guarantee of any future performance, analysis, forecast or prediction. Past performance does not guarantee future results.

The Information should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. All Information is impersonal and not tailored to the needs of any person, entity or group of persons.

None of the Information constitutes an offer to sell (or a solicitation of an offer to buy), any security, financial product or other investment vehicle or any trading strategy.

It is not possible to invest directly in an index. Exposure to an asset class or trading strategy or other category represented by an index is only available through third party investable instruments (if any) based on that index. MSCI does not issue, sponsor, endorse, market, offer, review or otherwise express any opinion regarding any fund, ETF, derivative or other security, investment, financial product or trading strategy that is based on, linked to or seeks to provide an investment return related to the performance of any MSCI index (collectively, “Index Linked Investments”). MSCI makes no assurance that any Index Linked Investments will accurately track index performance or provide positive investment returns. MSCI Inc. is not an investment adviser or fiduciary and MSCI makes no representation regarding the advisability of investing in any Index Linked Investments.

Index returns do not represent the results of actual trading of investible assets/securities. MSCI maintains and calculates indexes, but does not manage actual assets. Index returns do not reflect payment of any sales charges or fees an investor may pay to purchase the securities underlying the index or Index Linked Investments. The imposition of these fees and charges would cause the performance of an Index Linked Investment to be different than the MSCI index performance.

The Information may contain back tested data. Back-tested performance is not actual performance, but is hypothetical. There are frequently material differences between back tested performance results and actual results subsequently achieved by any investment strategy.

Constituents of MSCI equity indexes are listed companies, which are included in or excluded from the indexes according to the application of the relevant index methodologies. Accordingly, constituents in MSCI equity indexes may include MSCI Inc., clients of MSCI or suppliers to MSCI. Inclusion of a security within an MSCI index is not a recommendation by MSCI to buy, sell, or hold such security, nor is it considered to be investment advice.

Data and information produced by various affiliates of MSCI Inc., including MSCI ESG Research Inc. and Barra LLC, may be used in calculating certain MSCI indexes. More information can be found in the relevant index methodologies on www.msci.com.

MSCI receives compensation in connection with licensing its indexes to third parties. MSCI Inc.’s revenue includes fees based on assets in Index Linked Investments. Information can be found in MSCI Inc.’s company filings on the Investor Relations section of www.msci.com.

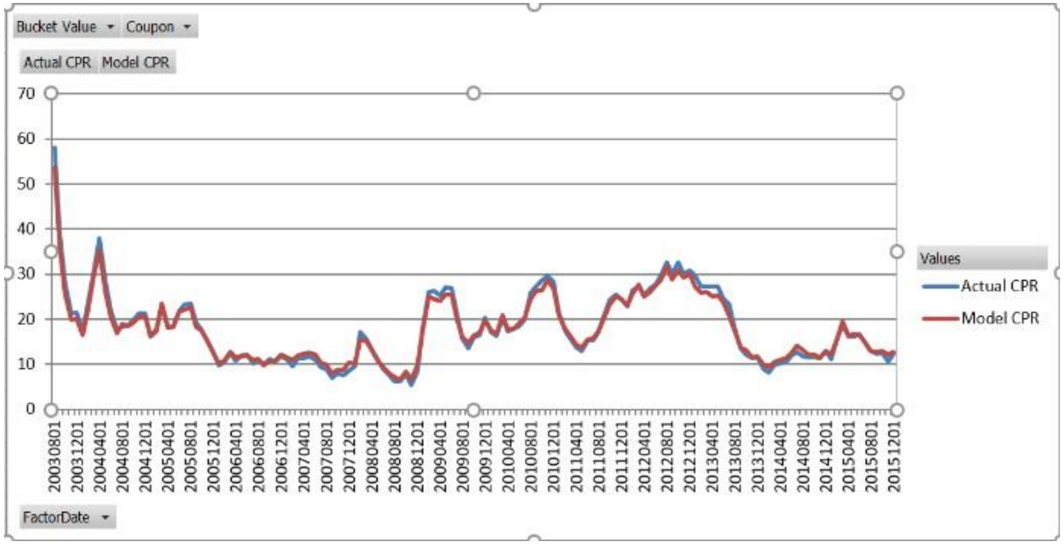
MSCI ESG Research Inc. is a Registered Investment Adviser under the Investment Advisers Act of 1940 and a subsidiary of MSCI Inc. Except with respect to any applicable products or services from MSCI ESG Research, neither MSCI nor any of its products or services recommends, endorses, approves or otherwise expresses any opinion regarding any issuer, securities, financial products or instruments or trading strategies and MSCI’s products or services are not intended to constitute investment advice or a recommendation to make (or refrain from making) any kind of investment decision and may not be relied on as such. Issuers mentioned or included in any MSCI ESG Research materials may include MSCI Inc., clients of MSCI or suppliers to MSCI, and may also purchase research or other products or services from MSCI ESG Research. MSCI ESG Research materials, including materials utilized in any MSCI ESG Indexes or other products, have not been submitted to, nor received approval from, the United States Securities and Exchange Commission or any other regulatory body.

Any use of or access to products, services or information of MSCI requires a license from MSCI. MSCI, Barra, RiskMetrics, IPD, FEA, InvestorForce, and other MSCI brands and product names are the trademarks, service marks, or registered trademarks of MSCI or its subsidiaries in the United States and other jurisdictions. The Global Industry Classification Standard (GICS) was developed by and is the exclusive property of MSCI and Standard & Poor’s. “Global Industry Classification Standard (GICS)” is a service mark of MSCI and Standard & Poor’s.

MSCI

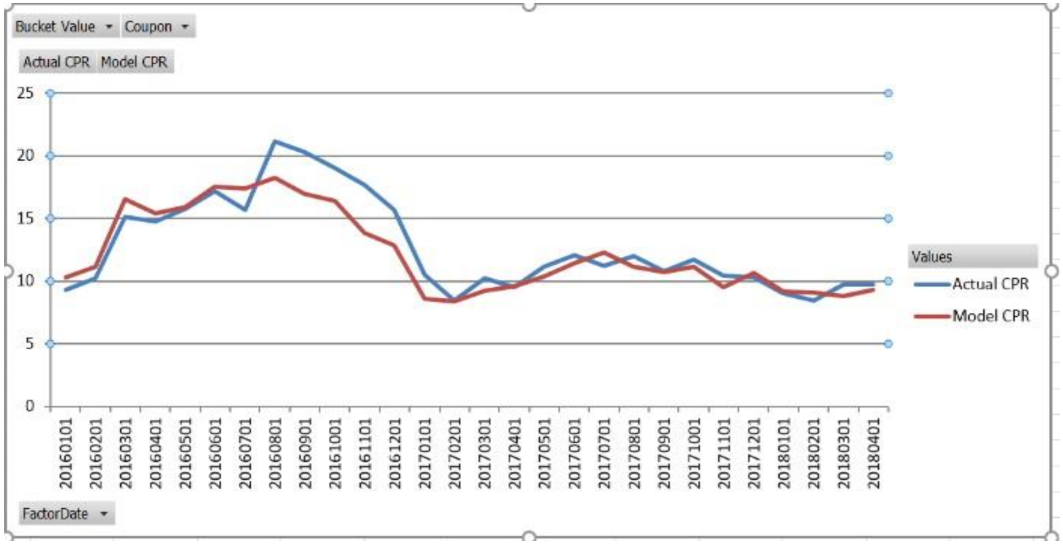
MODEL ERROR TRACKING

In-time
Out-of-sample
(1/2003-12/2015)



1. All attributes statistics are very close on July and August 2016 except CPR.
2. Risk driver is missing, i.e., media effect or regime change

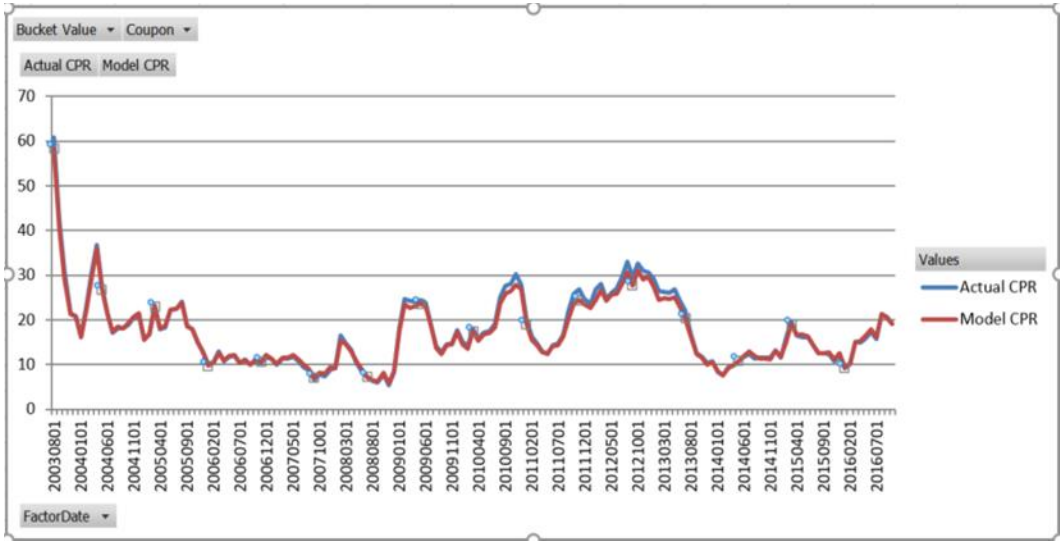
Out-of-time
Out-of-sample
(1/2016-4/2018)



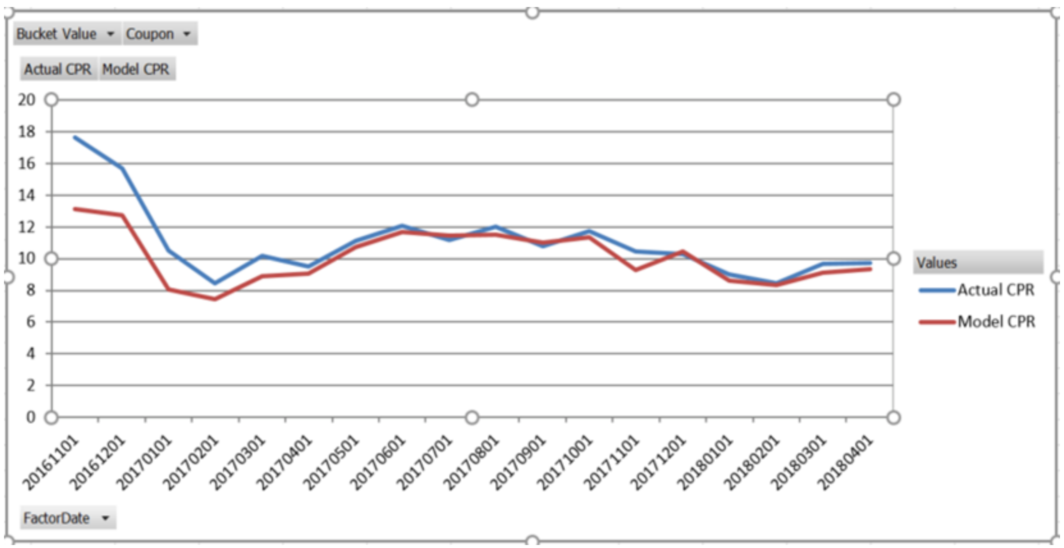
MSCI

MODEL ERROR TRACKING

In-time
Out-of-sample
(1/2003-10/2016)



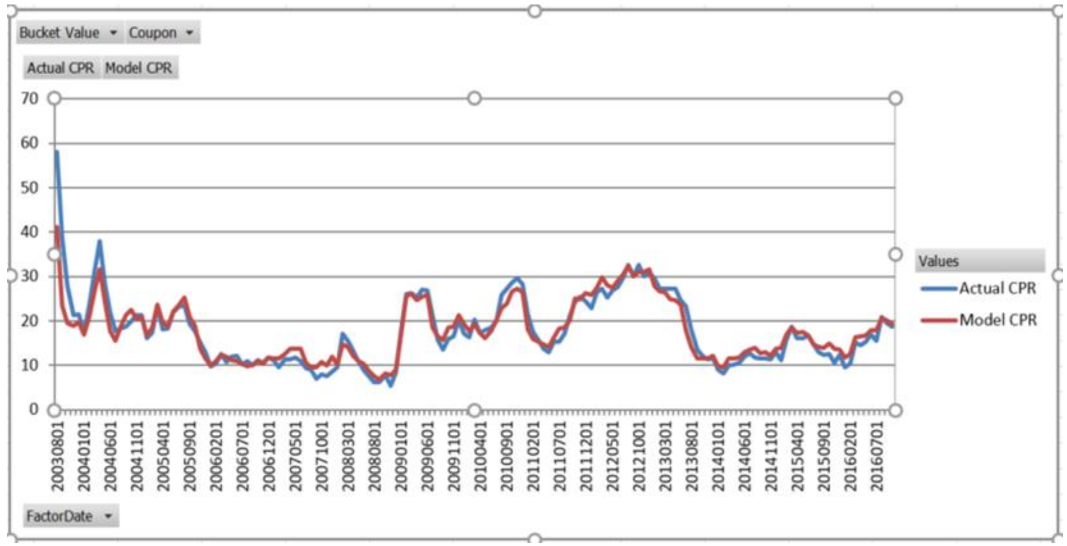
Out-of-time
Out-of-sample
(11/2016-4/2018)



MSCI

MODEL ERROR TRACKING

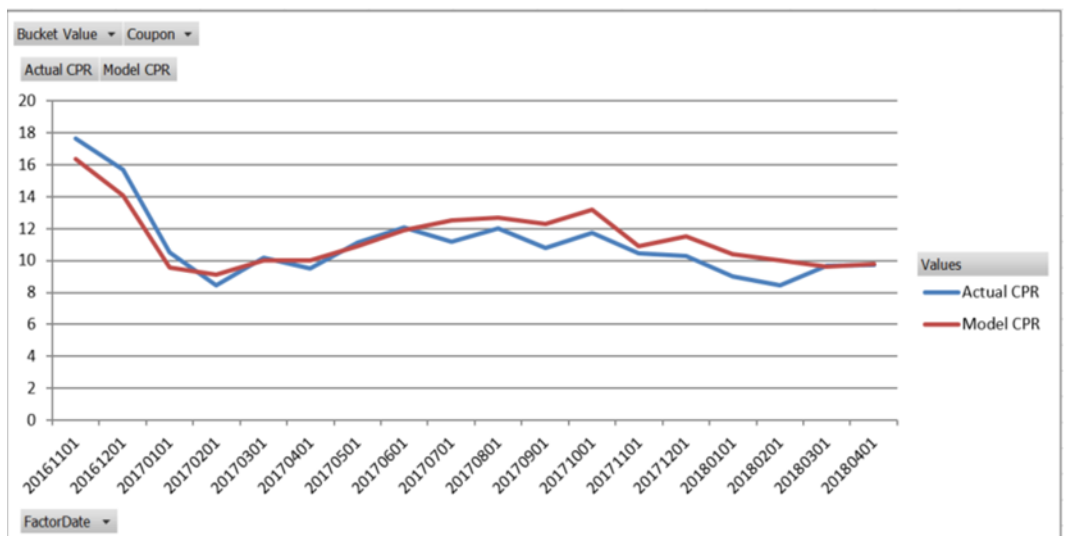
In-time
Out-of-sample
(1/2003-10/2016)



When Increase weights on 8/2016 – 10/2016 by 40 times in training:

1. Better in the early stage of out-of-time test
2. Sacrifice other period.

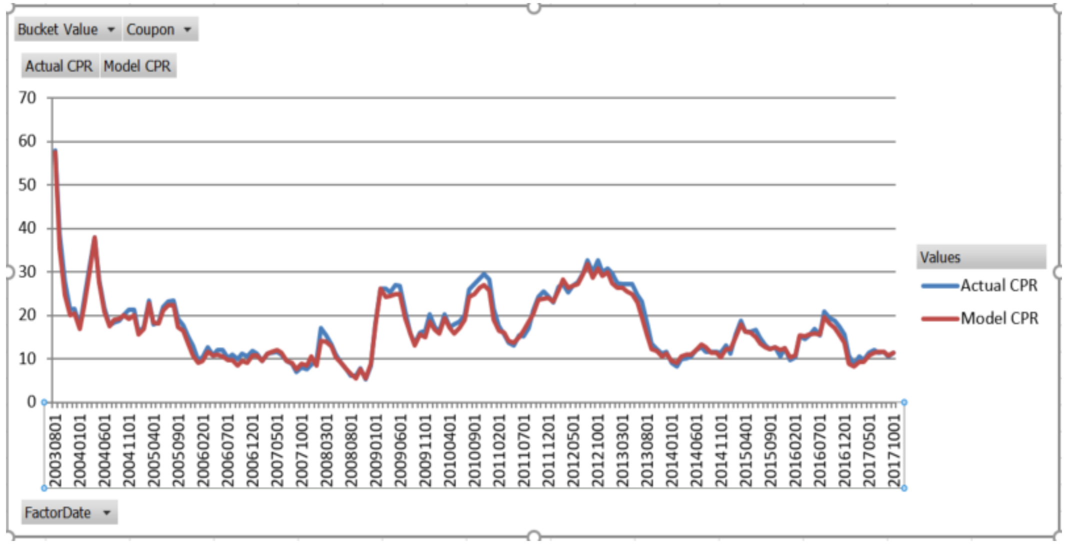
Out-of-time
Out-of-sample
(11/2016-4/2018)



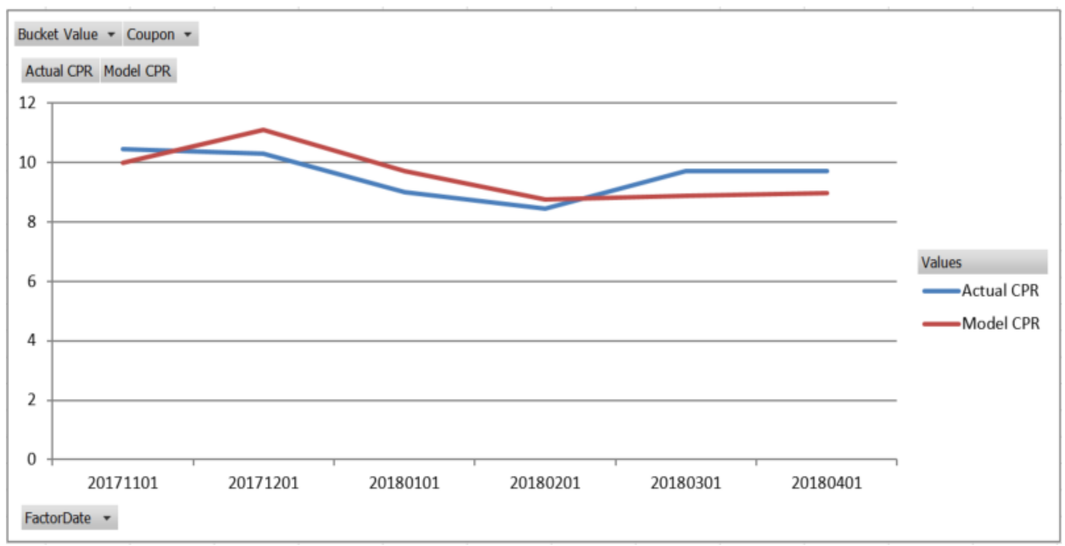
MSCI

MODEL ERROR TRACKING

In-time
Out-of-sample
(1/2003-10/2017)



Out-of-time
Out-of-sample
(10/2017-4/2018)

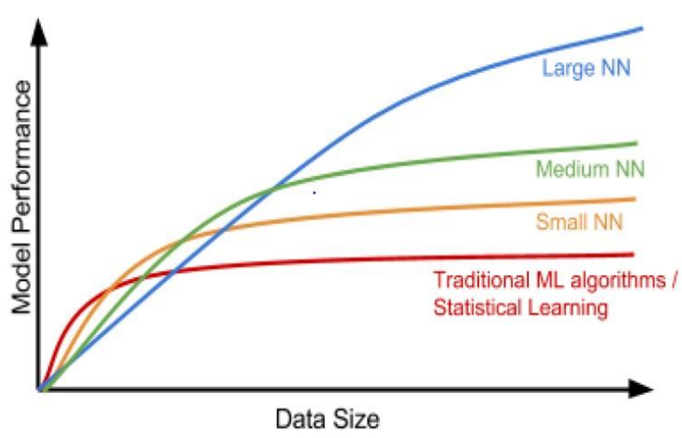


MSCI

TRADITIONAL VS.. DEEP MACHINE LEARNING

Identify problems → Set benchmarks → Train models → Compare **performance** → Choose model → Optimize **model** → Deploy

Traditional learning algorithm		**Deep Learning**	
Pros	Cons	Pros	Cons
Works better on smaller data	Hard to scale	state-of-the-art for certain domains, such as computer vision and speech recognition.	require large amount of data.
Financially and computationally cheap	Lack of variability	Perform very well on image, audio, and textual data, Easily updated with new data	Not suitable for classical machine learning problems.
Algorithms are easier to interpret, have more theories to back them up	Labor intensive model maintenance	Versatile architecture and low overhead maintenance	Computationally intensive to train, and they require much more expertise to tune



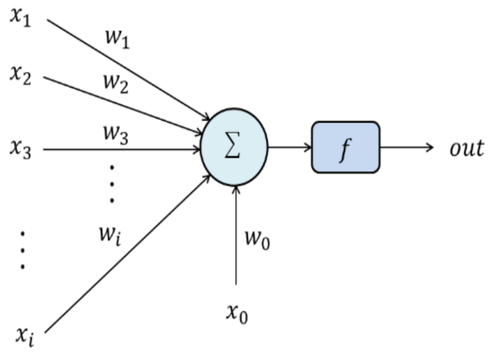
MSCI

NEURAL NETWORKS MODEL

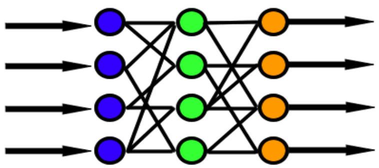
Feed forward neural network (FNN)

the information moves in only forward direction from the input nodes to the output nodes. There are no cycles or loops in the network.;

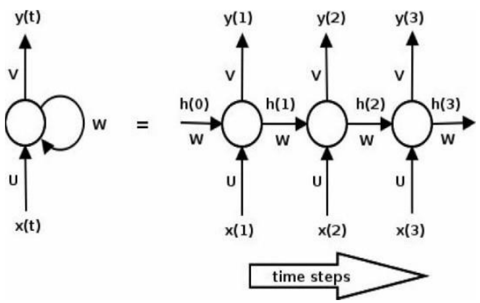
Deep FNN consists of tens of layers and thousands of nodes; the simplest kind of FNN is logistic model



Logistic Model



Feedforward Neural Network



Recurrent Neural Network

Recurrent Neural Network (RNN)

A class of neural networks exploit the historical input sequences. Such inputs could be text, speech, time series, and anything else where the occurrence of an input in the sequence is dependent on the inputs that appeared before it

Motivation: Not all problems can be converted into one with fixed length inputs and outputs, such as text translation, speech recognition or time-series; predictions require a system to store and use context information

The input at time t include both the attributes at t and the intermediate values containing history at t-1.

MSCI