Agency MBS Prepayment Model Using Neural Networks

JIAWEI "DAVID" ZHANG, XIAOJIAN "JAN" ZHAO, JOY ZHANG, FEI TENG, SIYU LIN, AND HONGYUAN "HENRY" LI

JIAWEI "DAVID" ZHANG
is a managing director in Securitized Product Research at MSCI in New York, NY.
jzhang@alumni.princeton.edu

XIAOJIAN "JAN" ZHAO
is a principal in Advanced Analytics at Ernst & Young LLP in New York, NY.
zhaoxiaojian@gmail.com

JOY ZHANG
is an executive director and director in Securitized Product Research at MSCI, New York, NY.
joy.jzhang@gmail.com

FEI TENG
is a senior quantitative analyst in Quantitative Advisory Services at Ernst & Young LLP in New York, NY.
feiteng2010@u.northwestern.edu

SIYU LIN
is a senior quantitative analyst in Quantitative Advisory Services at Ernst & Young LLP in New York, NY.
linsiyu2019@gmail.com

HONGYUAN "HENRY" LI
is an executive director in Quantitative Advisory Services at Ernst & Young LLP in New York, NY.
hongyuan.li@gmail.com

Mortgages and mortgage-backed securities (MBS) are among the largest financial sectors in the United States. They also serve as key levers for the central bank and federal government in managing monetary policy and stimulating the national economy, as shown during the recent financial crisis. Mortgage prepayment modeling is essential to investment and risk analysis for MBS. It is also among the most complex areas of financial modeling. The complexities include the following:

- *A large dataset*—for example, the agency MBS data cover about 450,000 pools and 100 million loans, over 20 years; the data volume is in the order of terabytes.
- *A large set of risk factors*—the number of risk factors, including loan/pool and borrower attributes, and macroeconomic drivers, ranges from tens to hundreds (Dunsky 2014).
- *Difficulties in model specification and estimation* because of
  - risk factors often being highly nonlinear and interactive.
  - regime changes in credit availability, borrower behavior, and business practice.

Some examples of nonlinearity and interactive risk factors and regime changes are as follows:

- Prepayment due to housing turnover and prepayment due to rate refinance have very different risk factor sensitivities (Yu 2018). Because of this, housing turnover and rate refinance are often estimated separately.
  - Rate refinance tends to increase with loan size, while housing turnover often decreases with loan size.
  - Prepayment sensitivities to loan age, seasonality, geographical location/state, house price appreciation, etc., also differ between housing turnover and rate refinance.
- Housing turnover and rate refinance each have their own subset of models, which include cash-out refinance, term refinance, "trade-up," etc. In addition, agency MBS prepayment can also be caused by curtailment and delinquency buyout. These submodels have their own patterns of risk factor dependencies. As such, they are often specified and estimated separately. However, prepayment data often do not include the reason for prepayment.
- Housing and mortgage policy changes and borrowers' behavior changes also lead to prepayment regime changes (Zhang 2018b). Technology changes and loose mortgage credit produced very fast prepayment speeds from 2003 to 2006. The subsequent housing

crash, financial crisis, and tight mortgage credit in 2009–2011 led to much lower levels of prepayment intensity. While overall prepayment intensity has been gradually increasing, it has not reached pre-crisis levels. In addition, sensitivities to risk factors have also changed with overall prepayment intensity. For example, pre-crisis, low-FICO borrowers often were faster than high-FICO borrowers, due to the need to leverage house price appreciation for cash-out and to reduce mortgage rates when their credit recovered after one or two years of payment. However, low-FICO borrowers have been significantly slower after the crisis due to very tight mortgage underwriting standards (Zhang 2017).

- High current loan-to-value ratio (CLTV) borrowers generally could not refinance before the HARP1 program in 2009–2010. The HARP1 (Home Affordable Refinance Program) program was marginally successful in refinancing high CLTV loans. The subsequent HARP2 program, which started in 2011, achieved great success, with high-CLTV loans refinancing faster than low-CLTV loans for many months after the program was launched. (In prepayment model lingo, the CLTV curve was inverted during these periods.) (Zhang 2014).
- Business practices by originators and servicers also tend to affect prepayment behavior in the short and long term. For example,
  - During periods of high refinance activity, resource-constrained servicers and originators often optimize their workforce by prioritizing their efforts. This often leads to faster prepayments for certain segments of the borrower universe.
  - Many servicers have evolved their business strategies and practices, leading to secular changes in their prepayment patterns (Zhang 2018a).

Due to these difficulties there is a lack of standard modeling methodology, especially in the area of agency pool-level prepayment modeling. Often, modelers need to "guesstimate" model specifications through a trial-and-error approach, in which various function specifications of several sets of risk factors are proposed and tested against actual prepayment data.

This modeling process tends to have many shortcomings. For example, *idiosyncratic characteristics*. Among the MBS modeling and investment community, it is well known that prepayment model structures, forecasts, and valuation results span a wide spectrum across the industry, even though these models are often based on the same data source and a similar understanding of prepayment drivers and dynamics. This "craftsman" nature of prepayment modeling is not common among financial modeling.

Lack of transparency. Given the empirical and idiosyncratic nature of prepayment modeling, it is generally difficult to reproduce the results even of those prepayment models commonly used in the industry. The investment community often refers this as the "black-box" nature of prepayment modeling.

Model updating process. The trial-and-error approach is not only time consuming, but it also prone to large changes when new data are made available. Major model updates often happen months after the actual prepayment phenomenon that triggered the model revision.

Recent advances in artificial intelligence (AI) and computing power provide opportunities for a new prepayment modeling approach that may overcome many of these shortcomings.

DEEP NEURAL NETWORKS MODEL

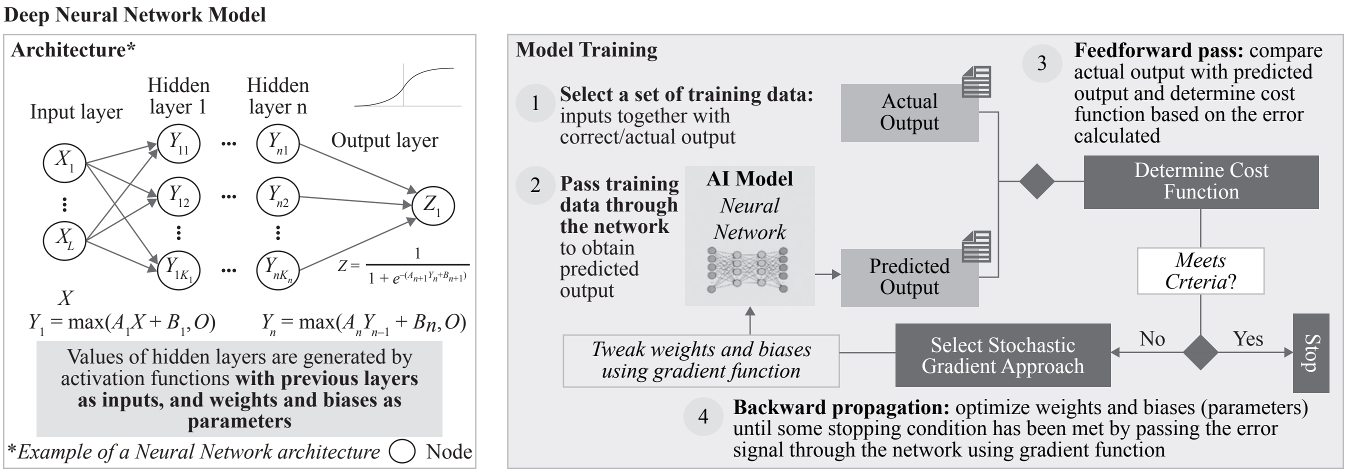
Artificial neural networks are inspired by the biological neural networks of animal brains and are composed of connected artificial neurons, which loosely model the biological brains. Deep learning is a class of methods and techniques that employ artificial neural networks with multiple layers of increasingly rich functionality. Deep learning has been intensively studied and successfully applied to many fields, including computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, board games (for example, AlphaGo), and computer games (LeCun et al. 2015).

Deep neural networks perform very well on image, audio, and text data, and they can be easily fine-tuned with new data using batch propagation. Their architectures (i.e., the number and structure of layers and nodes) can be adapted to many types of problems and can simplify the feature engineering process when sufficient samples are provided.[1] A two-layer neural

[1] A feature is an individual characteristic to represent the target object. Feature engineering is the process to derive relevant features based on available data, leveraging domain knowledge of the data in order to improve the predictive power of the machine learning algorithms.

Exhibit 1

Schema of Deep Neural Networks for Prepayment Modeling



network can represent any bounded degree polynomial, provided the neural networks are sufficiently large under mild assumptions on the activation function (Andoni et al. 2014). Thus, neural networks can learn extremely complex patterns that may prove challenging for other algorithms.

Deep learning algorithms are usually not suitable as general-purpose methods. Although deep learning has the capability of high prediction power, it requires much more data to train than other algorithms because the models have dramatically more parameters to estimate. It also requires much more expertise to tune (i.e., architecture setup and hyperparameter selection, which are intensively time consuming). In addition, deep learning is sometimes outperformed by tree ensembles for machine learning classification problems (Schmidhuber 2015). Besides, the black-box nature of neural networks is a barrier to adoption in applications where interpretability is essential.

The original concept of a neural networks can be traced back more than half a century. Due to the recent development of computation power and data storage capacity, neural networks algorithms can now be efficient for analyzing business problems. Recent analytical tools also provide good transparency on the neural network approach taken, for example, layer-wise relevance propagation (LRP; Bach et al. 2015) and deep learning important features (DeepLIFT; Koh and Liang 2017 and Shrikumar et al. 2017). Therefore, financial institutions have more motivations to pursue deep learning techniques nowadays.

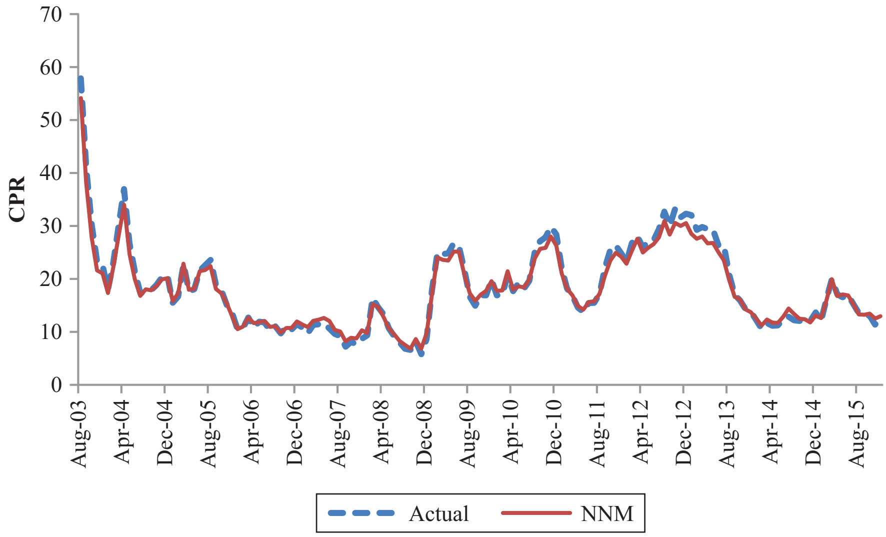
Many of the MBS prepayment modeling difficulties—for example, large datasets, large numbers of risk factors, nonlinear and highly interactive natures between the risk factors—may be suited to the neural networks modeling approach. This article describes an example of applying this new approach to agency MBS pool-level fixed-rate prepayment modeling. The AI modeling project is a collaboration between MSCI's Securitized Products Research group and Ernst & Young LLP's Quantitative Advisory Servises (QAS). We also contrast the AI model results with a production prepayment model used at MSCI Securitized Products Research (the "human" model or "Hmodel" in the exhibits' captions) that was constructed and maintained over a long period using the traditional modeling approach discussed in the previous section (Yu 2018).

Exhibit 1 shows the schema of the neural networks modeling approach for the agency prepayment model. We formulate the prediction of prepayment speed as a regression problem. We apply feedforward neural networks here along with several functional sublayers and development techniques to predict prepayment speeds.

This MBS machine learning work involves an iterative process that includes exploratory data analysis (EDA), feature selection, machine learning modeling

Exhibit 2

NNM In-Time Out-of-Sample Error Tracking for Prepayment Speeds for the Overall Agency 30-Year Universe



design and development, and performance evaluation. It incorporates the following:

- *Exploratory data analysis*. Examine data quality, perform data cleansing (missing values imputation and outlier detection), and conduct data transformation
- *Model feature selection*. Identify risk drivers and construct input variable sets.
- *Neural network model design and development*. Specify the neural networks model parameters, objective functions, and convergence/training methods
- *Performance evaluation*. Evaluate model performance, understand model sensitivities, and identify potential model overfitting

The training for a single neural networks model using a 10% data sample takes about three hours. This modeling efficiency compares extremely favorably with the traditional modeling approach, which often takes months or years.

We discuss the NNM model results in the following section, leaving details of the data and model specifications to the Appendix.

SAMPLE MODEL RESULTS

In order to test our NNM approach for out-of-sample forecast ability and model overfitting issues, we use a 10% random sample of the data between years 2003 and 2015 to train the model, then compare the model's performance with actual prepayment behavior for years between 2003 and 2015 (in-time out-of-sample tests[2]) and for years between 2016 and 2018 (out-of-time tests[3]).

We show three categories of model results:

1. Model versus actual prepayment speeds for the overall fixed-rate 30-year mortgage universe.
2. Model versus actual prepayment speeds for pool cohorts of various loan/pool attributes under different rates and macroeconomic environments (to test the ability to model complex, nonlinear, and interactive risk factors).
3. Model's prepayment sensitivities to various risk factors, such as loan/pool attributes and rates/macroeconomic variables (to test whether the model's behaviors are consistent with intuitions embedded in the MSCI production model—the Hmodel—that has been developed over the years using the traditional approach; this also tests for potential overfitting issues that are often suspected of the neural networks modeling approach).

[2] An in-time out-of-sample test uses the data from the same time period as the training data, but the sample data are exclusive from the training data.

[3] An out-of-time test uses the data from a different time period from the training data.

EXHIBIT 3

NNM and Hmodel Out-of-Sample Error Tracking for Prepayment Speeds for the Overall 30-Year Universe, 2016 and 2018

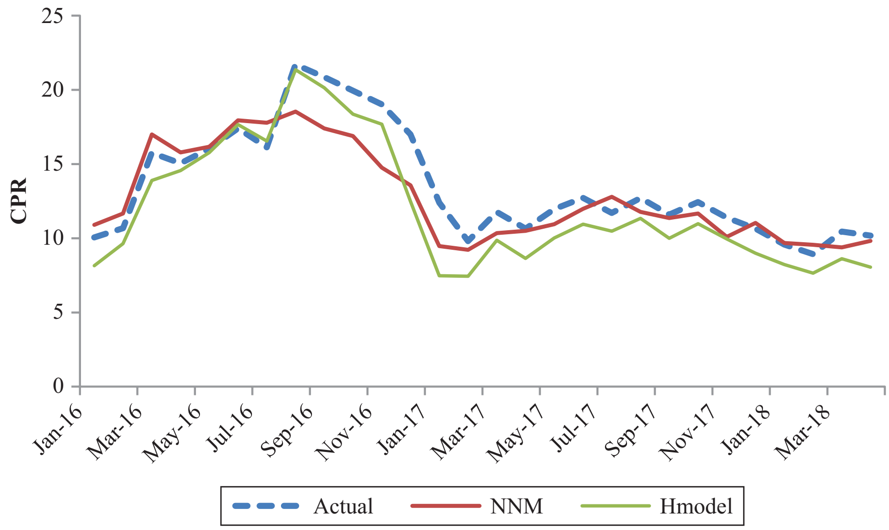


Exhibit 2 shows very good in-time out-of-sample error tracking for the overall 30-year sector prepayment speeds. NNM is able to accurately replicate the overall speeds with training from only 10% of the pool sample data. These forecast errors are generally marginally smaller than those achieved from the Hmodel.

Exhibit 3 shows NNM has good out-of-sample error tracking for the overall 30-year sector prepayment speeds. The model is able to accurately forecast the overall speeds between 2016 and 2018 with training from only 10% of the pool sample data from years prior to 2016.

NNM exhibits modest under-forecasts for the second half of 2016. However, our MSCI human models also show similar under-forecast patterns for this period. We also tested NNM by varying the training periods and the relative weight of the training data. The patterns of under-forecast persist. We conclude that the modestly higher-than-expected prepayment speeds in the second half of 2016 were likely caused by risk factors outside of those embedded in our data. In this case, NNM can serve as indicator of true prepayment surprises.

For the out-of-sample forecast period of 2017, however, NNM forecasts are close to actual prepayment speeds while Hmodel forecasts are generally under-forecasting. The short-term (month-on-month) prepayment forecasts from broker-dealers' MBS research publications during this periods often have forecasting errors of similar amplitudes.

Exhibit 4 shows error tracking results against loan/pool variables of FICO and SATO (spread at the origination), a credit indicator that complements other borrower credit variables such as FICO, OLTV (original loan-to-value ratio), loan size, and CLTV (current loan-to-value ratio). These are constructed by bucketing the prepayment speeds and model forecasts based on the pool variables.

NNM is able to capture accurately the sensitivities to the loan/pool variables, on par with the MSCI Hmodel. Note the contrast between the pre- and post-crisis prepayment behavior versus borrowers' FICO scores, as discussed in the previous section: On average, prepayment speeds were higher with low-FICO borrowers pre-crisis, while the relationship was reversed after the crisis, since mortgage credit has been generally tighter.

In addition, as shown, NNM is able to accurately capture the general relationship between prepayment speeds and SATO, loans sizes, and CLTV:

- Prepayment speeds are generally lower for higher SATO pools because borrowers with poorer credit are often slower to respond to refinance incentives.
- Prepayment speeds are generally higher for large-loan-size pools because the fixed portion of refinance costs often reduces economic benefits of refinance for lower loan sizes.
- Prepayment speeds are generally slower for high CLTV loans due to higher credit risk.

Exhibit 4

NNM Error Tracking against Pool Variables—FICO, SATO, Loan Size, and CLTV

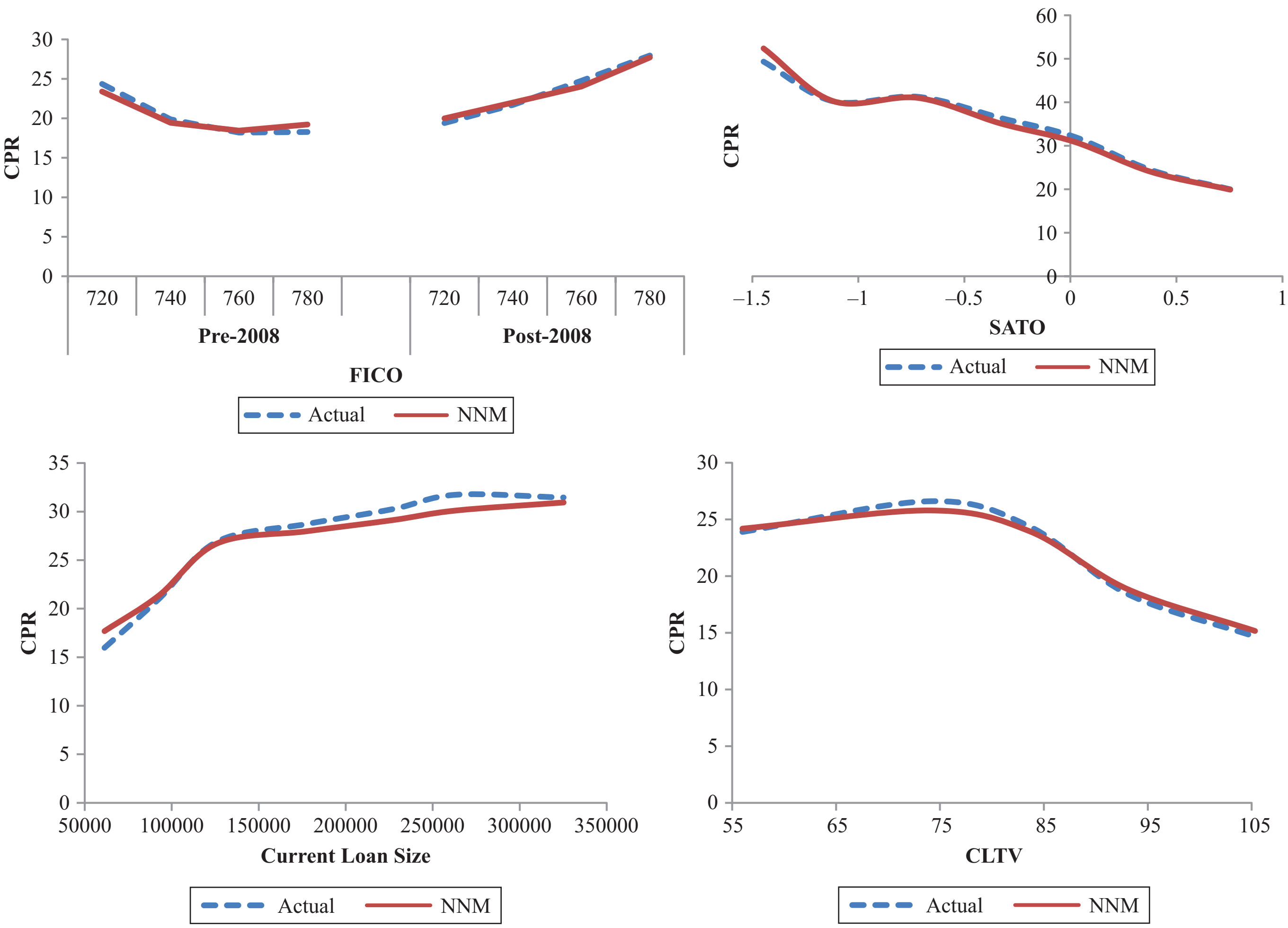


Exhibit 5 shows an example of how NNM accurately captures state-level prepayment behaviors (and generally better the human model). Generally, California loans have higher refinance speeds due to large loan size and more efficient mortgage refinance business practice, while New York loans are generally slower due to mortgage-recording taxes. In this case, NNM performance is generally better than the MSCI Hmodel. Accurately modeling state-level prepayment behavior is often difficult for the traditional cohort building approach to prepayment modeling. State pool cohorts tend to have different distributions of other pool-level variables, for example, loan size and house price appreciation rates. This makes it difficult for model specification and estimation to isolate the pure state-level effect.

Exhibits 4 and 5 measure model performance against a single pool variable. However, MBS pools are measured by about 30 variables. We have been advocating a new ranking-based comprehensive pool-level error tracking methodology (Zhang 2018b).

Exhibit 6 shows an example of this ranking-based error tracking for coupon 4s. All pools for coupon 4s are ranked and sorted based on their *model* prepayment forecasts, from slowest to the fastest. Then the whole cohort is bucketed into 15 equal-sized groups based on this model's forecasted speed-based ranking, from the slowest (group 1) to the fastest (group 15). The number 15 is chosen so that each group is reasonably large and statistical errors are smaller than differences in prepayment speeds between groups. We believe that this

EXHIBIT 5

NNM and Hmodel Error Tracking against State Variables (pools with 90%+ concentration in California and New York)

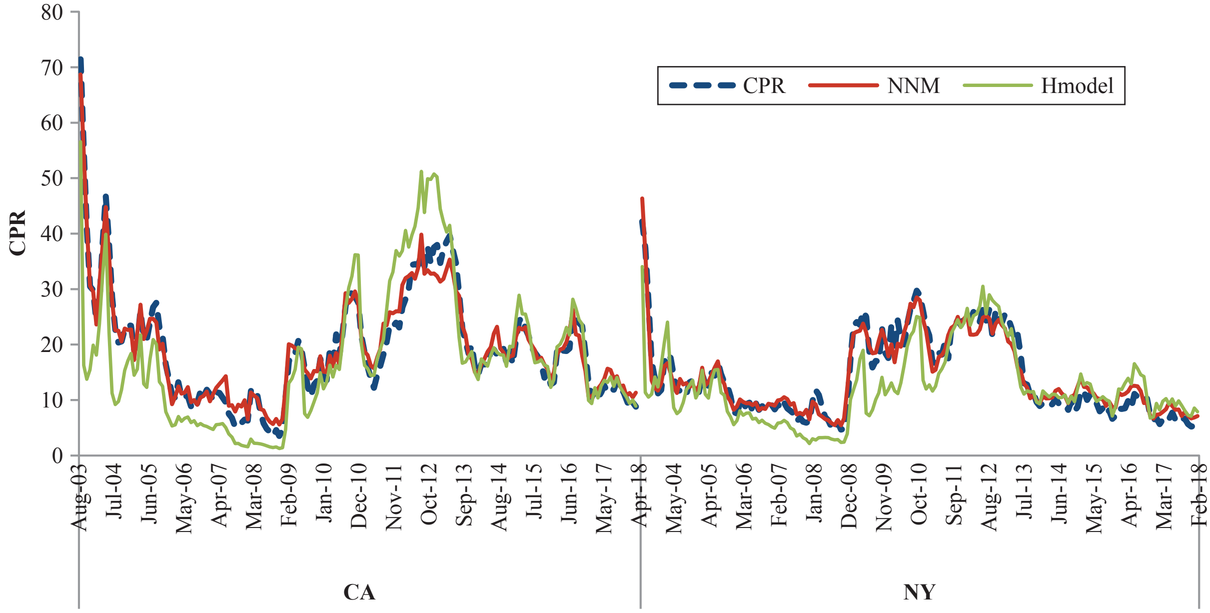
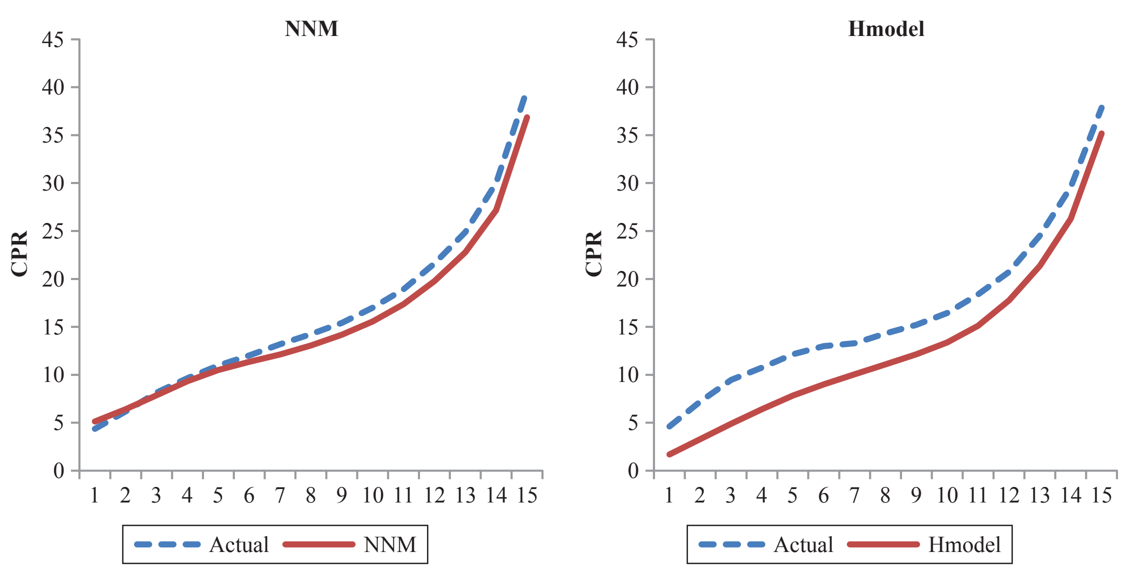


EXHIBIT 6

Ranking-Based Sample Error Tracking for Coupon 4s for NNM and Hmodel



ranking-based error tracking methodology provides a comprehensive measure of model accuracy across all pool variables (Zhang 2018b).

Exhibit 6 shows that NNM is accurate across all pool variables for coupons 4s and performed better than the Hmodel.

Exhibit 7

Sample Ranking-Based Error Tracking across Various Coupons and Historical Time Periods for NNM and Hmodel

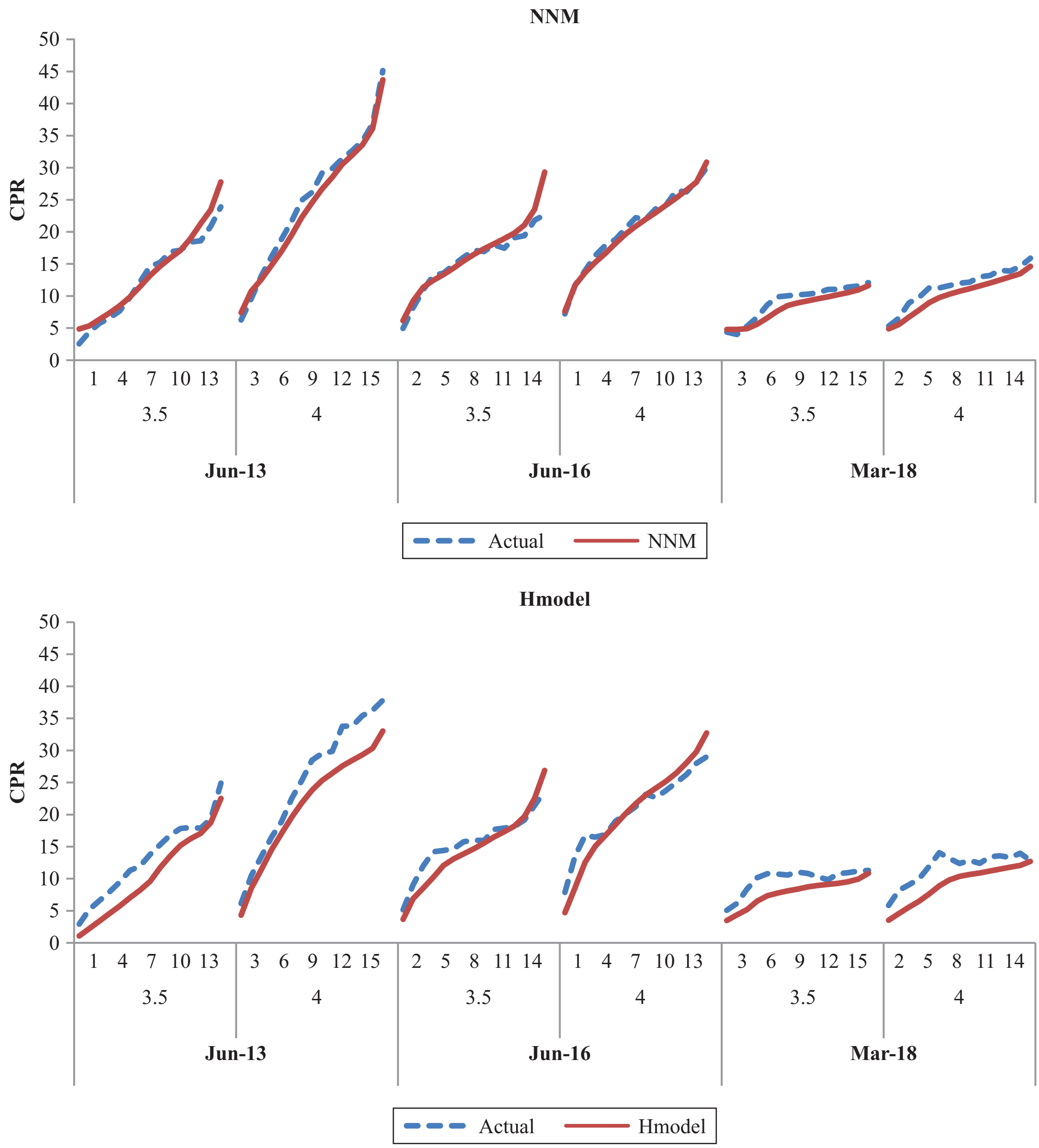
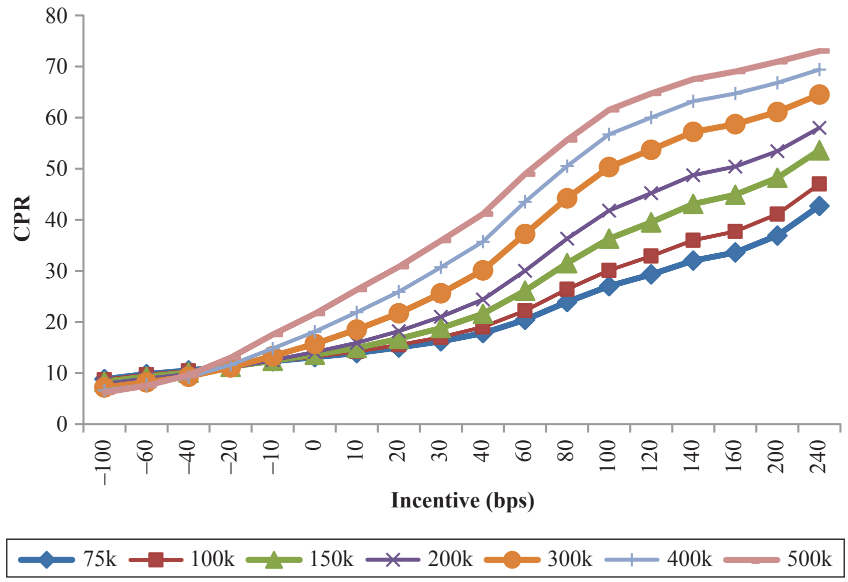


Exhibit 7 shows sample ranking-based error tracking across various coupons and historical time periods. NNM is able to accurately differentiate prepayment behavior across all pool variables, often better than the MSCI human model. This is likely due to the ability of neural networks to model highly nonlinear and interactive risk factors.

The flip side of neural networks models' ability to model highly nonlinear and interactive risk factors is lack of transparency. Given the multiple hidden layers and large numbers of nodes involved (see Exhibit 1 and the later discussion on neural networks development in the Appendix), the relationship between prepayment and input variables is not transparent and can

Exhibit 8

NNM S-Curve as Function of Loan Size: Model Prepayment Sensitivities to Loan Sizes and Refinance Incentives



potentially be overly noisy. In the context of the neural networks modeling methodology, this is often referred to as the overfitting problem. In the neural networks development discussion in the Appendix, we discuss the modeling techniques we employed to avoid the overfitting issue.

In order to enhance the transparency of the neural networks model, we test its sensitivities to risk factors/input variables to see whether these behaviors are consistent with economic intuition, which are often obtained through the traditional modeling approach. We pick representative loan/pool cohorts and compute how the prepayment model forecasts may change with varying pools variables and macroeconomic variables.

Exhibit 8 shows an example for loan size. NNM s-curves (how prepayment responds to mortgage rate incentives) are steeper for loan/pools with larger loan sizes. In addition, the discount speeds (prepayment speeds for loans with coupons below prevailing mortgage rates) are generally higher for low-loan-size pools.

Exhibit 9 shows the reversed loan size sensitivities for rate refinance (positive incentive) and housing turnover (negative incentive). NNM is able to capture these behaviors, and is generally consistent with economic intuitions and with the MSCI human model specification.

We now examine NNM's ability to model several more complex prepayment behaviors: burnout, "media effect," and HARP.

In burnout, loans or pools that saw refinance incentives in the past are generally slower when exposed to subsequent refinance opportunities. By not responding to earlier refinance opportunities, loans revealed their hidden attributes that are not conducive for refinance, and these hidden attributes are often more indicative than original explicit loan attributes in terms of forecasting future refinance intensity.

We examine NNM burnout effect by error tracking against the past refinance incentive. Exhibit 10 shows that pools that have higher past refinance incentives are generally slower in refinance speeds, hence NNM burnout effect is reasonably accurate, on par with the MSCI Hmodel.

Exhibit 11 shows an example of the so-called media effect. When mortgage rates are at or close to historical lows, as in 2012 and 2016 (Exhibit 12), the whole mortgage universe becomes refinance-able. The large refinance volumes strain the resources of originators/servicers, which often optimize their workforce to focus on a certain segment of refinance applications, for example on newer and often lower coupons.

EXHIBIT 9

The Loan Size Effects in NNM Are Correctly Modeled: The Loan Size Effects Are Reversed for Rate Refinance (positive incentive) and Housing Turnover (negative incentive)

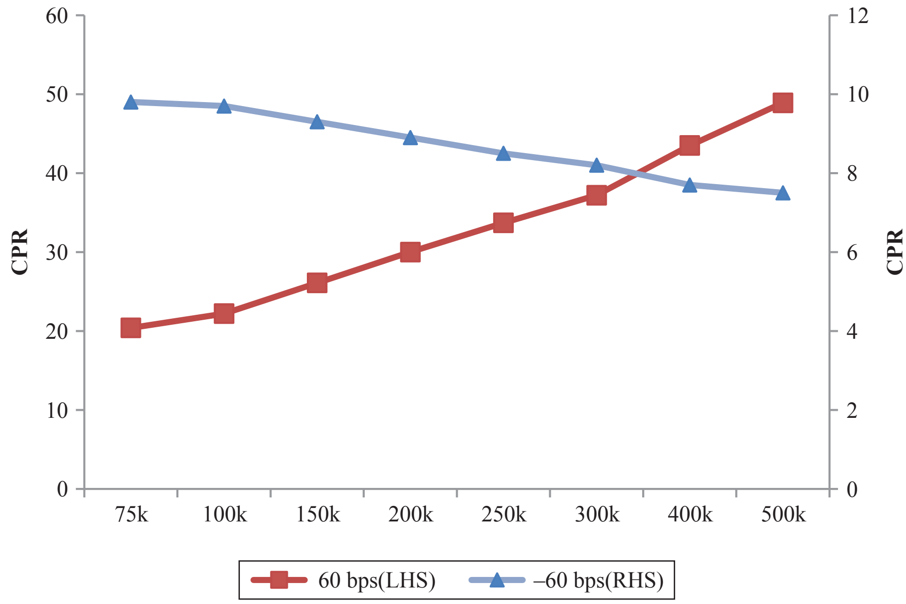
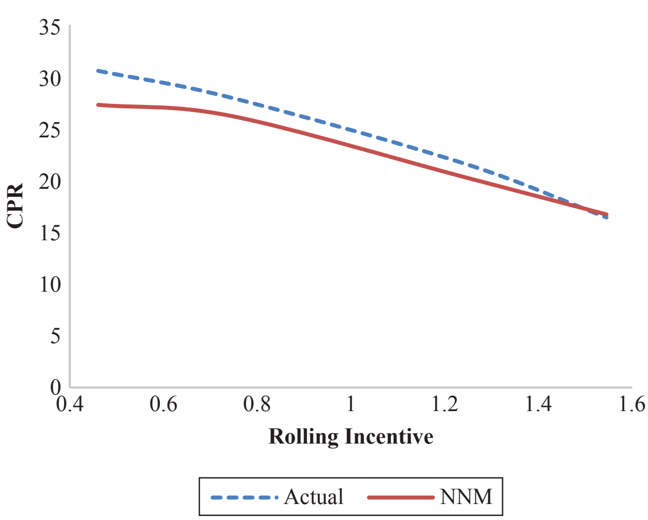


EXHIBIT 10

Error Tracking NNM Burnout Effect: Model and Actual Prepayment Speeds against Average Incentive in Prior 20 Months



This sometimes leads to faster prepayment speeds for lower coupons and hence an inverted s-curve.

Exhibit 11 shows FH 2011 3.5 versus 2010 4.0 prepayment speeds comparisons for July 2012–December 2012 and November 2011–February 2012, across TPO/Retail and Refi/Purchase combinations. While the loan attributes are similar and refinance incentives are similar, the 3.5s are much faster than 4s.

Exhibit 13 shows that NNM is able to capture this phenomenon for the 2012 refinance wave. For example, 2012 (and many other vintages) 3.5s ramp up prepayment speeds much faster than 4s and 4.5s, and reach much higher peak speeds, despite having fewer rate incentives.

Exhibit 14 shows that NNM is able to capture this phenomenon for the 2016 refinance wave, which was in the out-of-sample period. The 2015 (and many other vintages) 3.5s and 4s ramp up prepayment speeds much faster than 4.5s and reach much higher peak speeds, despite having fewer rate incentives.

For both the 2012 refinance wave (in-time out-of-sample) and 2016 refinance wave (out-of-time), NNM is able to model the relative prepayment strength between 3.5/4/4.5 coupons—and often better than the MSCI Hmodel.

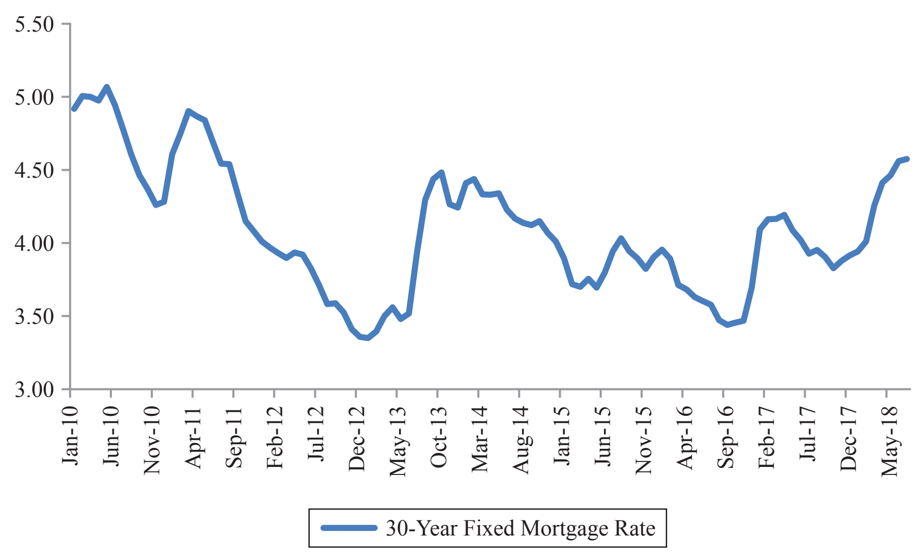
As discussed in the introduction, HARP was a federal refinance program focused on refinancing high-CLTV agency loans caused by the large house price depreciation in the aftermath of the Great Recession, to

EXHIBIT 11
FH 2011 3.5 vs. 2010 4.0 Comparisons

Cohort	Observation Range	CPR	WALA	SATO	CLTV	CLNSZ	Incentive	FICO	AvgUPB (billion)
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.12–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

Notes: 3.5s are faster than 4s, across TPO/Retail and Refi/Purchase combinations for July 2012–December 2012 and November 2011–February 2012. Refinance incentive and pool attributes (purchase/refinance, age, SATO, CLTV, FICO, loan size) are compared across origination channels (Retail/TPO).

EXHIBIT 12
History of Agency 30-Year Fixed Mortgage Rate



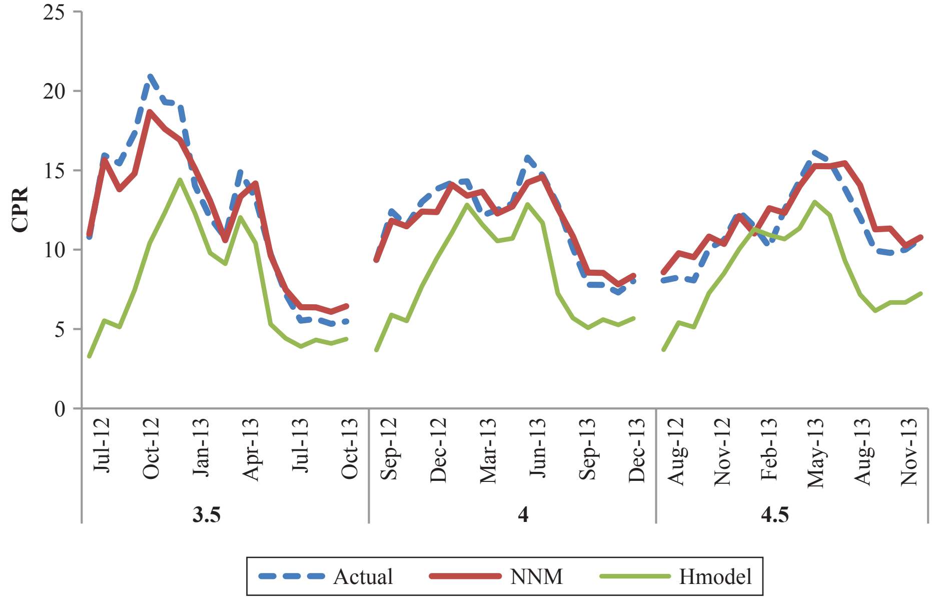
Note: Mortgage rates were at or close to historical lows in 2012 and 2016.

help homeowners reduce mortgage payments and avoid default. The effectiveness of the HARP and the subsequent HARP2 programs evolved in a complex pattern as the mortgage industry adjusted its policies and procedures to this unprecedented federal intervention.

Exhibit 15 shows that the neural networks model follows the general trend in the effectiveness of the HARP program but misses the complexity of its evolution. However, we are not aware of any industry models that were able to model these trends in detail during these periods.

EXHIBIT 13

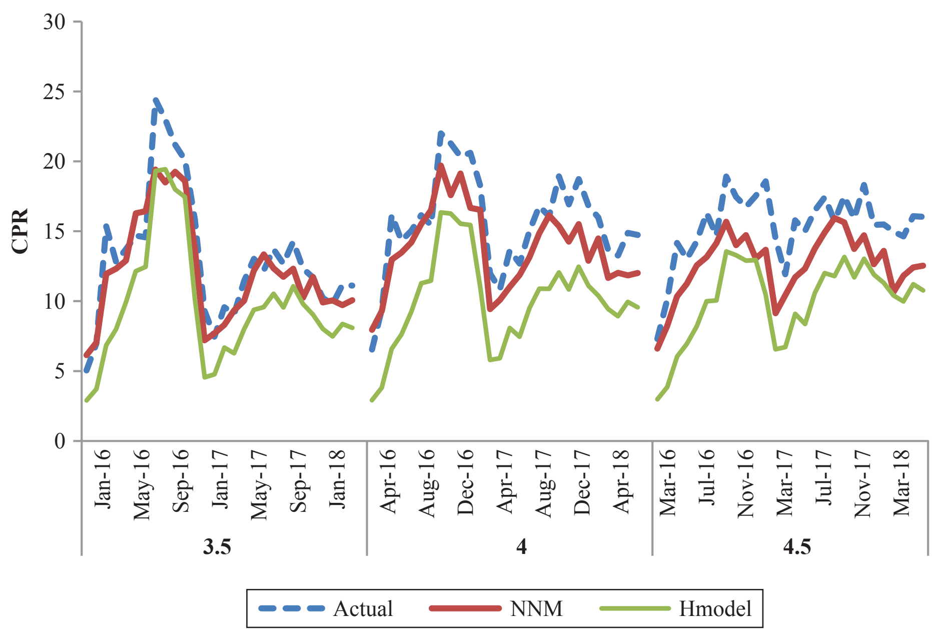
NNM and Hmodel Media Effect: 2012 Vintage 3.5/4/4.5s Prepayment Experience in 2012 Refinance Wave



Note: Lower coupons ramp up much faster in response to rate drops and higher peak speeds.

EXHIBIT 14

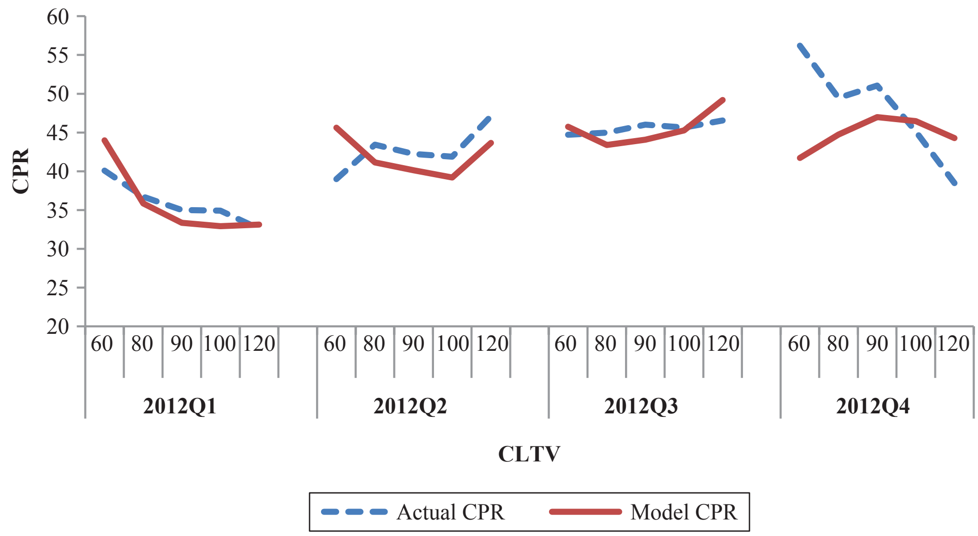
NNM and Hmodel Media Effect: 2015 Vintage 3.5/4/4.5s Prepayment Experience in 2016 Refinance Wave



Note: Lower coupons ramp up much faster in response to rate drops and higher peak speeds.

EXHIBIT 15

NNM Error Tracking against HARP Effectiveness across CLTV Cohorts



CONCLUSION

In this article, we show the promise of a new neural networks/machine learning approach to agency MBS prepayment modeling. The model results compare favorably against an industry production model that was constructed and maintained over a long period, using the traditional modeling approach. The neural networks approach is able to accurately model the highly nonlinear and interactive risk factor drivers and produce generally accurate prepayment forecasts at the pool level. Model transparency and overfitting issues can be overcome and managed by the latest neural networks modeling techniques. The gains in modeling efficiency gains—in the order of a hundred-fold—are potentially revolutionary.

APPENDIX

DATA AND MODELING DETAILS

Model Data

We purchased mortgage data from eMBS, a leading provider of MBS disclosure data for Fannie Mae and Freddie Mac. The eMBS agency MBS pool level prepayment data include all fixed rate 30-year TBA eligible pool attributes and prepayment rates in each month from 2000 to 2018. The total eMBS data is about 25 GB in size and has 30 raw data attributes. In addition, the macroeconomic data include FHFA weekly primary mortgage rates and state-level HPIs (House Price Indices) and unemployment rates.

Exploratory Data Analysis (EDA)

We first examine the data quality and data statistics. From Exhibit A1, we observe that the overall data quality from 2003 to 2018, in terms of missing data, is good. In detail, the ratio of missing data for the last four years is low for most important attributes, while the TPO and Refinance columns have many missing values. The data quality of some attributes is a little worse for the period from June 2008 to November 2012. For example, about 38% of refinance data is missing for June 2008 to November 2012 and only 1% missing for January 2014–April 2018; 60% of TPO data is missing for June 2008–November 2012 versus 40% missing for January 2014–April 2018. In addition, the number of samples per month increases tenfold from 2003 to 2018. Furthermore, FICO, OLTV, and AOLS are missing before June 2003. These issues were due to Fannie Mae and Freddie Mac data disclosure practices. Thus, we use mortgage data after June 2003 for model training and testing.

Data cleansing is another important step, which includes imputing missing values and detecting outliers. Due to good data quality in recent years and the large volume of eMBS data, we focus on cleansing the univariate data in our work. Most of the attributes are time varying, and the value

Exhibit A1

Overview of Input Data Statistics and Data Quality (percentage of missing data)

	Mean	Std	Min	Max	Missing_Percentage
Refi%	55.12	33.08	0	100	11.67%
Prepayment speed	0.02	0.05	0.00	1.00	0%
SecHome%	4.02	10.25	0	100	0.13%
TPO%	29.50	37.44	0	100.01	46.58%
Vintage					0%
WALA	82.57	53.71	–1	355	0%
WAM	265.33	61.52	1	361	0%
wAOLS_Q1MAX	126,515.05	60,665.39	6,000	802,000	0.13%
wAOLS_Q2MAX	155,337.76	77,546.31	6,000	900,000	0.13%
wAOLS_Q3MAX	186,005.53	99,808.08	6,000	930,000	0.13%
wAOLS_Q4MAX	220,241.01	142,062.02	6,000	1,470,000	0.13%
AOLS	142,902.62	68,380.20	6,000	802,000	0.13%
CBAL	7.31	54.71	0.01	10,238.42	0%
Coupon					0%
FICO	709.73	67.58	0	950	0.11%
GrossWAC	6.31	1.14	2.47	11.74	0%
Investor%	10.12	20.99	0	100	0.12%
Issuer					0%
Multifamily	4.95	13.32	0	100.01	0.16%
NLOANS	43.64	241.44	1	52,742	0%
OLTV	75.81	10.89	0	107	0.01%
wCLNSZ	125,650.88	66,518.76	1,806	786,453	0%

in the current month is correlated to the previous month's. To impute missing values for these time-varying attributes, we first group the records by pool ID and the number of loans to impute missing values via interpolation. If there are missing values at the beginning or the end for a particular pool ID and the number of loans, we impute the missing values by using the first value in the following or previous months, which are referred as forward-fill or backward-fill methods, respectively. Missing values are approximated relatively accurately in this scenario. Then, we fill the rest of the missing values via grouping the records by pool ID and imputing missing values using interpolation, forward-fill, and backward-fill methods. Because TPO has a high level of missing data and its range is from 0 to 100, we impute all missing values as −99. For those attributes that are constant for a pair of pool and number of loans, we either use the average or the most frequent value to impute the missing values, for example, for "Vintage," "OLTV," and "FICO." After these steps, all missing values for each pool are imputed, as long as there is a record in at least one month for the pool. As to outliers, we eliminate the unrealistic records based on our experience and restrict the extreme value of some variables to avoid negative effects.

In addition, we conduct data transformation. Based on our domain knowledge, we artificially create spread-at-origination (SATO), incentive, HARP indicators, one indicator for mortgage credit environment and indicators for the HARP program and eligible loans/pools.

Model Feature Selection

Feature selection is a critical step for machine learning projects and is an active research topic in academia. Filter, wrapper, and embedded methods are widely used for feature selection for machine learning modeling in the industry (Kumar 2014). Usually, a neural network does not require complicated feature selection methods, because it can choose the proper nonlinear forms of attributes and interactions based on the intrinsic features of the data, when there are sufficient hidden nodes and data. In practice, however, feature selection is sometimes still needed due to data/hardware limitations and training time tolerances. In our project, we follow four steps for feature selection.

First, we choose variables based on our domain knowledge. From previous experience, we know some variables have a significant impact on the target variable. Second, because we expect good sensitivity analysis using the black box method, we eliminate multicollinearity effects by taking advantage of the covariance matrix.[4] Third, we apply statistical methods to choose the most relevant variables. In particular, we calculate the information value of every single variable and the product of every two variables and then select all variables with information values larger than 0.02;[5] these are considered to be useful predictors. Last, we apply the embedded methods for feature selection. In particular, we build a decision tree model with small mean square error and generate the importance scores for all attributes based on this model. The variable with higher importance score are more important to predict the target variable in the decision tree model. Through these four steps, we achieve a set of input variables containing high correlation with and significant amount of information about the target variable and having limited multicollinearity effects. The input variables for our model are listed in Exhibit A2.

Neural Networks Development

Our neural networks model uses an individual pool's monthly data record as model input and its prepayment speed in the next month as the model output. The well-tuned neural networks have five hidden layers and one output layer. The input layer has around 40 numerical and dummy variables;[6] the first and second hidden layers have twice as many nodes as the second and third hidden layers, respectively; the third and fourth layers have four times as many nodes as the fourth and fifth layers, respectively; the output layer has one node.

In each hidden layer, we adopt four functional sublayers. The first sublayer is a dropout layer, which is used to address overfitting (Srivastava et al. 2014); the second sublayer is a fully-connected layer, which includes L2 and max-norm regularization setting for overfitting mitigation (Lee et al. 2010); the third sublayer is for batch normalization, which normalizes the output of activation function to address the distribution shift issue and could expedite the convergence speed (Ioffe and Szegedy 2015); the last sublayer is the activation function for nonlinear transformation.[7]

In our design, we use a grid search method for the main hyperparameters of batch size, number of nodes and layers, learning rate, max-norm constraint, L2-norm lambda, and dropout rate, etc.[8] We determine the optimal set of hyperparameters by comparing the tenfold cross-validation performance of different tuples of hyperparameters.

An adaptive moment estimation (ADAM) algorithm is applied to optimize the network weights, to compute adaptive learning rates for each parameter (Kingma and Ba 2015). We also adopt the idea of model ensemble in our design.[9] In particular, we construct parallel neural networks using random data samples and choose the average of the outputs of these parallel neural networks as our final output. This is similar to the bagging idea of random forest and can mitigate the effects of bad local minima and overfitting.[10]

There are tens of hyperparameters in our design. Algorithm-related parameters include learning rate, momentum rate, batch size, learning rate decay, and so on. Structure-related parameters include number of nodes, number of layers, activation function, and so on. Functional parameters include dropout rate, max-out, L2-norm lambda, and so on. Model link weights and hyperparameters are tuned by using data from June 2003 to December 2015 for training and cross validation. Some key hyperparameters are determined by grid search, as follows:

- Dropout rate is 0.4 for hidden nodes
- Batch size is 1,024
- Max-norm has the regularization constraint of 0.5
- Activation function is ReLu
- There are 512 hidden nodes in the first hidden layer, 256 in the second, 128 in the third, 32 in the fourth, and 8 in the fifth, etc.

Evaluation Package

We implemented three different analyses to evaluate our model. First, error tracking is used to measure the prediction power of our model as in Exhibits 2–5. Second, sensitivity analysis is used to interpret the model behavior

[4] Multicollinearity is a phenomenon that significant correlations are present among the input variables. It causes instability when estimating the model coefficients, and hence the accuracy of statistical inference and sensitivity analysis might be compromised.

[5] Information value is a measure based on information theory pioneered by Claude Shannon and is a commonly used metric to measure the information carried by input variables.

[6] The dummy variable takes either 0 or 1 as its value to indicate whether some categorical effect on the target variable is present or not.

[7] In artificial neural network, the activation function transforms the inputs to an output of the node, which is used as input for the next layer. Common activation functions include sigmoid, ReLU, TanH, and so on.

[8] Grid search is a traditional way of performing hyperparameter search and optimization in an exhaustive and brute force manner through a manually specified subset of the hyperparameter space.

[9] Model ensemble is a method to obtain better predictive performance by leveraging multiple learning algorithms/models. Common types of ensembles include bagging, boosting and stacking.

[10] Bagging is a type of ensemble method, which trains each model using a randomly drawn subset of training data.

Exhibit A2
Model Input Variables and Their Descriptions

Name	Description	Data Type
Independent Variables		
WALA	Weighted Average Loan Age	Numerical
WAC	Weighted Average Coupon	Numerical
CLNSZ	Current Average Loan Size	Numerical
OLTV	Original Loan to Value	Numerical
Refi%	Percentage of Refinanced Loans by UPB	Numerical
SecHome%	Percentage of Second Home Loans by UPB	Numerical
MultiFamily%	Percentage of Multi Family Loans by UPB	Numerical
Investor%	Percentage of Refinanced Loans by UPB	Numerical
TPO%	Percentage of Third Party Origination by UPB	Numerical
AOL	Original Average Loan Size	Numerical
LNSZ_Q4	Max Original Loan Size	Numerical
LNSZ_Q3	Max Original Loan Size-3rd Quartile	Numerical
LNSZ_Q1	Max Original Loan Size-1st Quartile	Numerical
Geo_CA%	Percentage of California Loans by UPB	Numerical
Geo_FL%	Percentage of Florida Loans by UPB	Numerical
Geo_TX%	Percentage of Texas Loans by UPB	Numerical
Geo_NY%	Percentage of New York Loans by UPB	Numerical
Geo_NE%	Percentage of New England Region Loans by UPB	Numerical
Geo_NO%	Percentage of North Region Loans by UPB	Numerical
Geo_SO%	Percentage of South Region Loans by UPB	Numerical
Geo_PC%	Percentage of Pacific Region Loans by UPB	Numerical
Geo_AT%	Percentage of Atlantic Region Loans by UPB	Numerical
Geo_NONUS%	Percentage of non-US Region Loans by UPB	Numerical
Seasonality	Calendar month	Categorical
Derived Variables		
Incentive	WAC-Mortgage Rate(t)	Numerical
Rolling Incentive	Average Incentive	Numerical
Loan size dispersion	(LNSZ_Q3-LNSZ_Q1)/AOL	Numerical
SATO	Spread at origination = WAC-Mortgage Rate(0)	Numerical
HPA	House Price Appreciation (HPI(t)/HPI(0)-1)	Numerical
HARP-able	1: IssueMonth <= Jun. 2009 and Factor Date between Mar. 2009 and Dec. 2011	Categorical
	2: IssueMonth <= Jun. 2009 and Factor Date > Dec. 2011	
HARP-ed	Refi% = 100 and OLTV >80 and IssueMonth > Jun. 2009	Categorical
Underwriting standard	0: Before 2008, 1: After 2008	Categorical
Weight		
cBal	Current Balance	Numerical
Dependent Variable		
Prepayment speed	Prepayment Speed in SMM	Numerical

along with a few attribute changes as in Exhibits 8–10. Third, ranking analysis checks the fitting of all attributes as in Exhibits 6 and 7.

Model Computations

We use workstations with 128GB memory and NVIDIA GeForce GTX 1080 Ti GPU. With this hardware configuration, the training for a single neural networks model using 10% data takes three hours. Because we apply grid search to determine the optimal hyperparameters, the total training time is around 150 hours.

REFERENCES

Andoni, A., R. Panigrahy, G. Valiant, and L. Zhang. 2014. "Learning Polynomials with Neural Networks." Proceedings of ICML.

Bach, S., A. Binder, G. Montavon, F. Klauschen, K. R. Müller, and W. Samek. 2015. "On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation." *PLOS ONE* 10 (7): e0130140.

Dunsky, R. 2014. "FHFA Mortgage Analytics Platform." FHFA Research White paper.

Ioffe, S., and C. Szegedy. 2015. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." ICML.

Kingma, D., and J. Ba. 2015. "Adam: A Method for Stochastic Optimization." International Conference on Learning Representations.

Koh, P., and P. Liang. 2017. "Understanding Black-Box Predictions via Influence Functions." Proceedings of ICML.

Kumar, V. 2014. "Feature Selection: A Literature Review." *Smart Computing Review* 4 (3).

LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep Learning." *Nature* 521: 436–444.

Lee, J., B. Recht, N. Srebro, R. R. Salakhutdinov, and J. A. Tropp. 2010. "Practical Large-Scale Optimization for Max-Norm Regularization." *Advances in Neural Information Processing Systems* 1297–1305.

Schmidhuber, J. 2015. "Deep Learning in Neural Networks: An Overview." *Neural Networks* 61: 85–117.

Shrikumar, A., P. Greenside, and A. Kundaje. 2017. "Learning Important Features through Propagating Activation Differences." arXiv preprint, 1704.02685.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." *Journal of Machine Learning Research* 15: 1929–1958.

Yu, Y. 2018. "MSCI Agency Fixed Rate Refinance Prepayment Model." Research paper, MSCI.

Zhang, J. 2018a. "Is MBS Refinance Risk Increasing?" Research paper, MSCI.

Zhang, J. 2018b. "Model Agency Pool Level Variables." Research paper, MSCI.

Zhang, J., and H. Tang. 2014. "US Prime Mortgage Credit Risk—Insight from the GSE Loan Level Credit Data Release." *Real Estate Finance* 30 (3): 78–106.

Zhang, J., and H. Tang. 2017. "Did the US Housing Crisis Start in 2003? The Impact of Borrower Subsequent Debt." *The Journal of Structured Finance* (Fall): 100–107.

Disclaimer
Opinions expressed here are the authors' and do not necessarily represent those of their employers.