infima

Why Infima's deep learning technologies deliver superior prepayment predictions for Agency MBS

Kay Giesecke
Founder and CEO

Accurate predictions of prepayment speeds are critical for security selection, portfolio construction and risk management in Agency MBS markets. Infima's disruptive, patent-pending deep learning technologies set new prediction standards, delivering performance-boosting advantages to MBS market participants including investors and dealers. In this paper we outline Infima's technologies and distinguish them from conventional approaches. Infima's platform is superior on three key attributes that matter to Agency MBS participants: accuracy, robustness and computational speed. A companion paper quantifies and documents the superior prediction performance of the Infima platform.[1]

Key Takeaways

- Infima has built robust systems that ingest all available loan performance data on an ongoing basis, processing billions of loan/month observations that cover all borrower events including delinquencies, modifications, and (partial) prepayments.
- The wealth of data enables Infima to discern borrower behavior patterns at the loan level that are ignored by conventional small-sample loan-level or pool-level analyses.
- The Infima approach utilizes a broad array of loan and borrower-level, macro-economic, financial and other variables, going significantly beyond conventional variables such as credit score, loan-to-value ratio, interest rates, and home prices.
- A deep neural network comprehensively describes the time-dependent likelihoods of borrower transitions between any two states. It completely encodes all relationships that exist between the variables and the behavior of an individual borrower, including any non-linear effects. The a-priori specification of the pertinent relationships required in conventional approaches becomes unnecessary.
- Statistical tools make Infima deep neural networks fully transparent. There are no black boxes.
- Patent-pending computational technology enables unprecedented application speeds.

Data and transitions

Our predictions of prepayment speeds are based on historical borrower behavior data—lots of them. The mortgage market is relatively data-rich: there are granular loan-level data for tens of millions of borrowers across the entire United States, describing their individual monthly behavior over several decades. Rather than taking a small sample of that data and discarding the rest, we have built robust systems that ingest all available data on an ongoing basis. We process *billions* of loan/month observations that cover all borrower events including delinquencies of all forms (1-6 months past due), modification, curtailment, prepayment, and other relevant events.

The data generate a transition matrix (Figure 1) that describes the likelihoods of an average borrower's transitions between any two states (eg. current, 1 month delinquent, prepaid) over a typical month. The matrix shows that on average, the likelihood of a borrower to remain current is 97.6%. The probability of a current borrower falling behind payment is just 0.6% on average, while voluntary prepayment (refinancing, housing turnover) occurs at a rate of 1.7%. We can also discern a momentum effect: a borrower who is already 3 months delinquent is relatively likely (37.9%, to be precise) to miss the next payment again. It gets worse: a 4-month delinquent borrower is even more likely (68.1%) to fall behind even further. The recovery of a 5-month delinquent borrower is entirely unlikely. An understanding of these delinquency effects is important for capturing loan buyouts by the agencies (involuntary prepayments). Moreover, notice

[1] Available at www.infima.io/insights.

that delinquent borrowers prepay at meaningful rates (between 1.4% and 2.7%), so the modeling of delinquencies is also important for capturing voluntary prepayments.

	Current	1M	2M	3M	4M	5M	6M	Other	Prepaid
Current	97.6	0.6	-	-	-	-	-	0.0	1.7
1M	43.0	40.3	14.1	-	-	-	-	0.0	2.7
2M	14.3	16.4	29.2	**37.9**	-	-	-	0.1	2.0
3M	6.3	3.2	6.3	13.8	**68.1**	-	-	0.6	1.7
4M	4.7	0.9	1.0	2.3	7.6	**80.1**	-	1.9	1.5
5M	3.1	0.5	0.3	0.5	1.2	5.3	**82.8**	4.9	1.4

Figure 1: The table represents the borrower state transition matrix, which provides the likelihoods of an average borrower's transitions between any two states over a typical month. The state "Other" includes REO, short sale, and third-party sale.

Going beyond the transition matrix, the data expose granular borrower behavior across the population during multiple economic cycles. These cycles include, for example, the run-up to the financial crisis of 2007-09 as well as the post-crisis period. The wealth of data enables us to discern borrower behavior patterns at the loan level that are ignored by conventional small-sample loan-level or pool-level analyses. The latter only considers pool-level average attributes, completely ignoring the often very significant heterogeneity of loan/borrower attributes within a pool.[2] The former fails to address infrequently observed borrower behavior, for example behavior after natural disasters (eg. during a period of excessive local unemployment following Hurricane Katrina). It also tends to ignore borrower behavior that is confined to certain geographical regions (influenced by jurisdiction or other factors) or relatively smaller pockets of the borrower population which may not be represented well in a small data set.

The wealth of granular loan-level data also enables us to examine a broad array of loan and borrower-level, macro-economic, financial and other variables, going significantly beyond conventional variables such as credit score, loan-to-value ratio, interest rates, and home prices. These data efforts, combined with a powerful AI system, yield best-in-class predictive performance for prepayment speeds.

Deep learning

Our proprietary deep learning system harnesses the wealth of data we are processing on an ongoing basis. Its design and implementation are based on the results of years of unique research conducted at Stanford University by members of Infima's founding team.[3] At the system's center is a deep neural network that comprehensively describes the borrower- and time-dependent likelihoods of transitions between any two states over a month. The model represents the *conditional* transition matrix for a given borrower at a given point in time. It provides borrower-level predictions for a transition between any two states, for example from 90 days past due to prepayment (from the unconditional transition matrix shown above, the likelihood of such a transition is 1.7% for the average borrower).

Unlike conventional approaches, the deep neural network model completely encodes the complex, typically nonlinear relationships that exist between the loan/borrower-level and macro-economic variables and the behavior of an individual borrower. This powerful AI technology eliminates the need of conventional modeling approaches to specify the pertinent relationships a priori, which is severely limiting since one can only account for the few well-known relationships. Our unique approach shows that there are multiple other, previously unknown, effects that play a significant role for borrower

[2] The significant limitations of pool-level analysis are well known; see, for example, "Comparing loan-level and pool-level mortgage portfolio analysis," by S. Chinchalkar and R. Stein, Moody's Research Labs, 2010.

[3] See "Deep Learning for Mortgage Risk" by Apaar Sadhwani, Kay Giesecke, Justin Sirignano, *Journal of Financial Econometrics*, Volume 19, Issue 2, 2021, Pages 313–368. Available at www.infima.io/insights.

behavior. For example, it reveals that the sensitivity of prepayment speeds to changes in home prices, interest rates and other macro-economic variables strongly depends on multiple loan/borrower attributes such as credit score, loan-to-value ratio, and unpaid principal balance (a variable interaction effect). To illustrate, Figure 2 shows S-curves generated by our fitted model for several credit scores intervals. It implies that the model sensitivity of prepayment speed depends on the current prepay incentive as well as credit score. Borrowers with different credit scores respond differently to changes in interest rates; low-FICO borrowers are the least sensitive. A complete understanding of these sensitivities is a prerequisite for constructing effective hedges of interest rate risk.

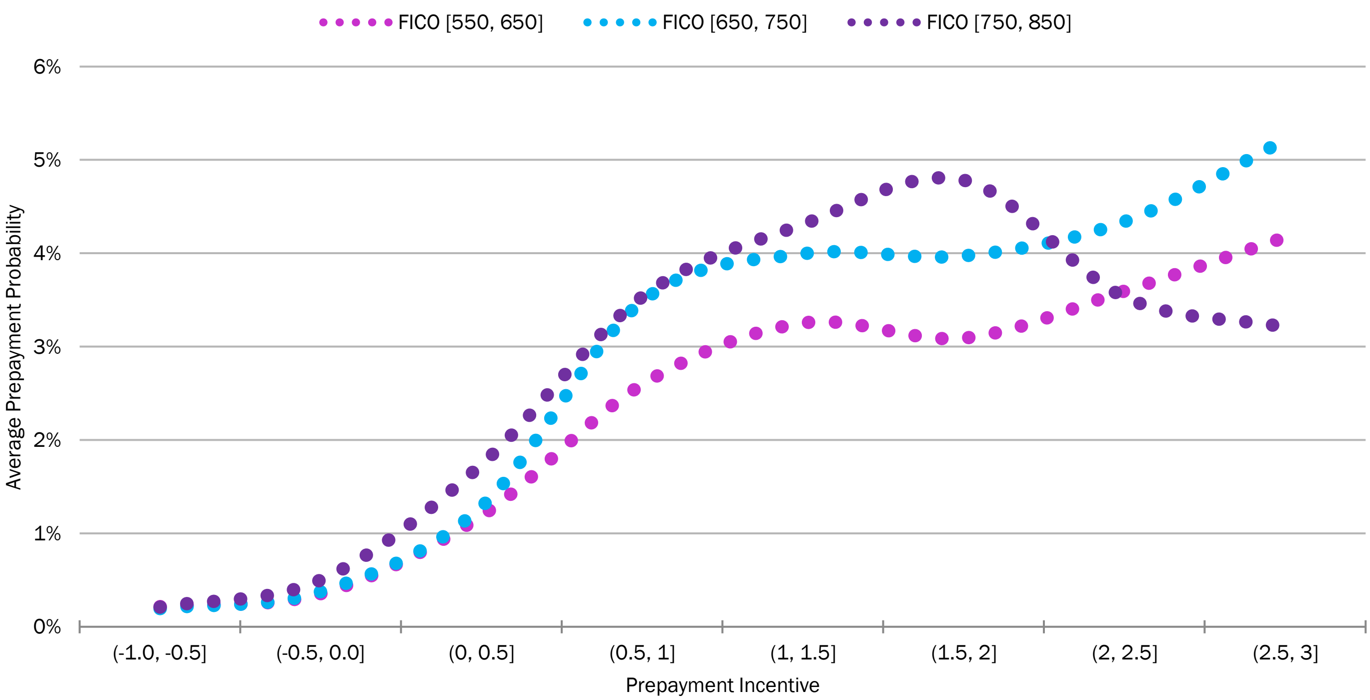

Figure 2: The chart shows S-curves implied by the fitted Infima model for several credit score intervals. Here, the "Prepayment Incentive" (x-axis) is the difference between a mortgage's own interest rate and the market mortgage rate. The "Prepayment Probability" (y-axis) is the model-implied likelihood of prepayment over a month.

Performance

The ability of our deep learning technology to automatically (without manual input) uncover and encode all stable patterns hidden in the data significantly boosts predictive performance for prepayment and delinquency events. Conventional prediction approaches, being based on relatively small portions of the available data, relatively few variables, and strong a priori assumptions on the relationship between variables and borrower behavior (linearity in particular), miss many of these hidden patterns. As a result, their predictive performance suffers—their prepayment forecast errors are often large, particularly during disruptive market environments when analytical guidance is needed most.

To illustrate, Figure 3 below shows the ratio of Actual 1M CPR to Predicted 1M CPR month by month starting in September 2019, for the entire universe of FNCL pools (Fannie Mae and Freddie Mac) with vintages between 2013 and 2021. The ideal is ratio is 1.0. We see that Infima's deep learning predictions hug this ideal line consistently even throughout the tumultuous COVID period. Our predictions are robust and unbiased even in disruptive markets. In contrast, the predictions provided by two widely-used vendors exhibit excessive errors. Both vendors consistently understate prepayment speeds. One vendor's errors are so large during the pandemic as to render predicted speeds unusable.

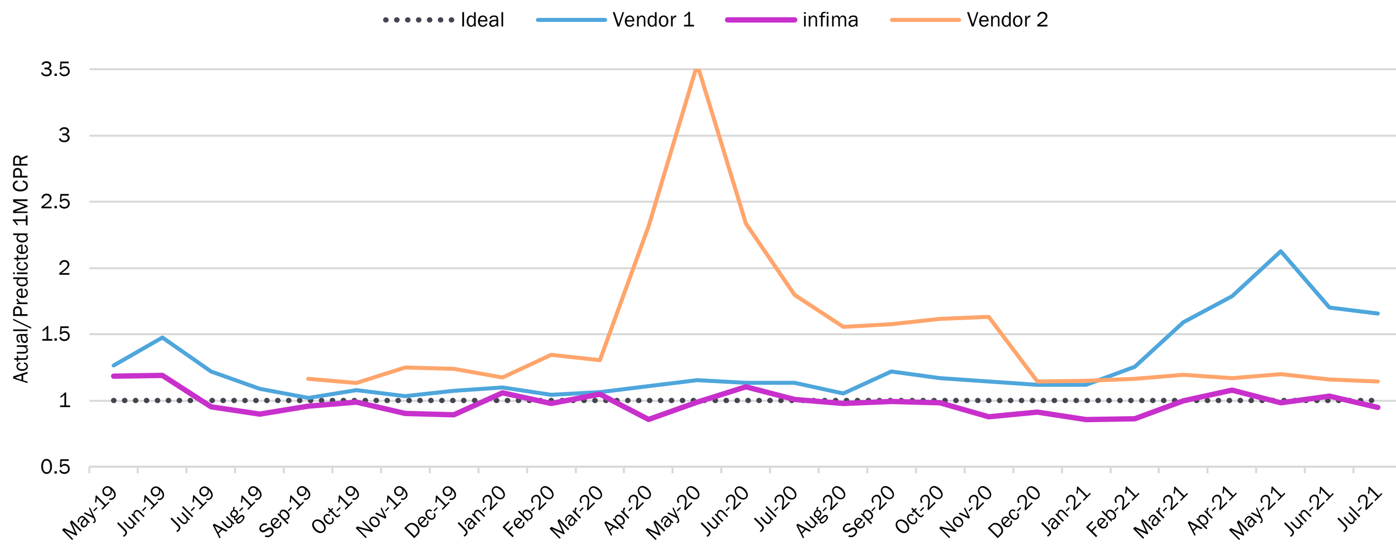

Figure 3: The chart shows the ratio of Actual 1M CPR to Predicted 1M CPR month by month between May 2019 and July 2021, for the universe of FNCL pools with vintages between 2013 and 2021. The ideal is ratio is 1.0. The chart demonstrates the superior performance of Infima's deep learning CPR predictions.

We highlight that the predictions we are examining here are *out-of-sample* predictions. Specifically, the Infima model producing the predictions for May 2019 to the present shown in Figure 3 was fitted using data available until December 2018. While the model has not seen any data relating to the pandemic, it performs remarkably well at all times: pre, during, and post pandemic. Please see our companion paper[4] for a range of additional analyses of predictive performance.

The monthly predictions of the deep neural network can be combined to obtain predictions at the loan-level over any time horizon (eg. 360 months). This is facilitated by a proprietary simulation engine that takes account of all possible paths a borrower can take to reach a given target state (eg. prepayment or 6-month delinquency) over the horizon of interest. The monthly loan-level predictions can also be aggregated efficiently across a pool to obtain predictions of CPRs at CUSIP level. The aggregation is driven by powerful proprietary numerical algorithms which can achieve speedups of more than 10,000x vs. conventional simulation approaches.[5] This patent-pending computational technology enables unprecedented applications speeds. For example, we will be able to offer intra-day prediction updates for the entire pool universe, allowing users to respond to significant rate moves in real time.

Explainability

Many powerful AI prediction models such as neural networks are perceived to be "black boxes" that permit little insight into how the different input variables influence predictions. This perception is no longer justified, however. Recent research[6] has developed statistical methods to rigorously assess the significance of the variables going into a neural network. These tools effectively render deep neural networks as transparent as conventional (linear) models. They allow

[4] Available at www.infima.io/insights.

[5] See "Risk Analysis for Large Pools of Loans" by Justin Sirignano and Kay Giesecke, *Management Science*, Volume 65, Issue 1, 2019, Pages 107-121. Available at www.infima.io/insights.

[6] See "Computationally Efficient Feature Significance and Importance for Machine Learning Models" by Enguerrand Horel and Kay Giesecke, and "Significance Tests for Neural Networks" by Enguerrand Horel and Kay Giesecke, *Journal of Machine Learning Research*, Volume 21, Issue 227, 2020, Pages 1–29. Both papers are available at www.infima.io/insights.

us to distinguish influential variables from non-influential ones and rank-order variables according to their importance for the outcomes of interest (eg. prepayment, delinquency). While many of the conventional loan/borrower-level variables are found to be influential, our analysis also points to several other variables that strongly influence borrower behavior. Examples of these include variables that track borrowers' behavior in the recent past. Going beyond individual variables, we also systematically test for variable interactions, and our tests emphasize the importance of the interaction effects discussed above.

Conclusion

Deep neural networks underpin many of the best-performing AI systems, including speech recognition on smartphones or Google's latest automatic translator. The tremendous success of these applications has spurred the interest in applying neural networks in a variety of other fields. Our pioneering work developing deep neural network technologies for prepayment speed prediction highlights their strong potential for challenging forecasting problems in fixed-income and credit markets. It shows that these innovative technologies set new prediction standards—significantly outperforming the conventional prepayment prediction technologies which are widely used today, delivering superior security selection capabilities to market participants.

Questions? Interested in a demo? Please reach out to Kay Giesecke at kay@infima.io