Deep Learning for disentangling Liquidity-constrained and Strategic Default

Arka Prava Bandyopadhyay and Yildiray Yildirim *

February 8, 2022
Preliminary Draft

Abstract

We disentangle liquidity-constrained default and the incentives for strategic default using Deep Neural Network (DNN) methodology on a proprietary Trepp data set of commercial mortgages. Our results are robust to the shock of the severe Financial Crisis (2008) and the plausible economic catastrophe ensuing from COVID-19 pandemic (2020-2021). We retrieve the motive of default from observationally equivalent delinquency classes by bivariate analysis of default rate on Net operating income (NOI) and Loan-to-Value (LTV). NOI, appraisal reduction amount, prepayment penalty clause, balloon payment amongst others co-determine the delinquency class in highly nonlinear ways compared to more statistically significant variables such as LTV. Prediction accuracy for defaulted loans is higher when DNN is compared with other models, by increasing flexibility and relaxing the specification structure. These findings have significant implications for investors, rating agencies and policymakers.

Key words: Strategic default, CMBS, Machine learning, Stress Test 2008, COVID-19

*Bandyopadhyay, ArkaPrava.Bandyopadhyay@baruch.cuny.edu, PhD Candidate, William Newman Department of Real Estate, Baruch College, New York, NY 10010 USA; Yildirim, Yildiray.Yildirim@baruch.cuny.edu, Chair, William Newman Department of Real Estate, Baruch College, New York, NY 10010 USA. We thank Johannes Stroebel, Lilia Maliar, Arpit Gupta, Agostino Capponi, Kay Giesecke, Christopher Palmer, Brent Ambrose, Sumit Agarwal, Sasha Indarte, Rawley Heimer, Anthony DeFusco, Peter Carr, Linda Allen, Liuren Wu, Guillaume Haeringer, Stefania Albanesi and Joonsung Won for helpful comments. We also benefited by conference participants at ASSA 2020, FMA 2020, Bloomberg BBQ, BlackRock FMG and Baruch Annual Symposium.

1 Introduction

We reconcile the long-standing debate between two competing theories (negative equity and life events) of mortgage default using Deep Neural Network (DNN), disentangling competing motives that produce observationally equivalent results (see Indarte (2020) for Regression Kink Design approach). Riddiough (1991) claimed that mortgage default in triggered by life events that reduces borrower cash flows. Foster and Order (1984) opine that default is caused by negative equity (when option to sell the home in the future is worth less that her loan obligations) as borrowers treat their homes like a financial asset. We create a framework in which both of the above can be observed, and more importantly strategic default can be disentangled from liquidity-constrained default. The advent of large scale use of DNN and the computational resources for handling big data has enabled us to address this central research question, impossible even ten years back. Identification strategy has been defined in the literature, vis-a-vis uniqueness for *parametric* models in economics, albeit with some assumptions (Kahn and Whited (2017)). DNN can provide a unique non-parametric model which serves the same purpose as canonical identification strategy in a high-dimensional setup without additional assumptions. The variable importance (in terms of marginal contribution) of the economically meaningful variables is consistent even during dire economic conditions, ensuring the stability of the non-parametric model and providing a structure-free pseudo-identification strategy.

Although defaults are observed, one cannot observe whether a default is strategic as strategic defaulters are pooled together with borrowers who cannot afford to pay.[1] But, to the best of our knowledge, none have been able to disentangle who defaulted strategically (which can lead to spatial clustering of default) and the intent to default strategically (which can lead to aggregate contagion).[2] Although, moral hazard is time-invariant, but the incentive of a borrower for moral hazard needs to be *triggered*. This is just one special case among the several elements of strategic behavior we document. We use several key covariates, e.g., Net Operating Income, Appraisal Reduction Amount, Prepayment Penalty Clause, Balloon Payment at Maturity, Non-Recoverability, etc. to identify when moral hazard is triggered vis-a-vis higher order non-linear interactions during severe stress in the 2008 Financial Crisis.

[1] Bajari, Chu, and Park (2008) assess the likelihood of strategic default by estimating a structural model of default that includes both cash flow considerations and negative equity considerations. Similarly, in a survey of a sample of U.S. households, Guiso, Sapienza, and Zingales (2013) ask about a person's willingness to default at different levels of negative equity, keeping the level of wealth and other individual characteristics constant, thereby separating contagion effects from sorting effects, by asking questions about social and moral attitudes and perceptions toward default.

[2] Ganong and Noel (2020) try to identify the reason for default from life-events and adverse cashflow events in the context of Residential Mortgages

Rather than assume a single default trigger based on property value (measured by contemporaneous LTV), our model incorporates a second trigger based on contemporaneous property income (NOI).[3] Net Operating Income (NOI), a key indicator for an investment property's financial standing, is the income generated by an investment property after subtracting the operating expenses and vacancy losses but before principal and interest payments, capital expenditures, depreciation, and amortization. The variation in the *ability to service debt*, measured by Debt Service Coverage Ratio, i.e., $DSCR = \frac{NOI}{ScheduledPayment}$ provides the identification strategy in disentangling liquity-constrained and strategic defaulters. Commercial Real Estate (CRE) borrowers with $DSCR < 1$, do not have the ability to service the debt obligation in a given month as they are *liquidity-constrained*, whereas, CRE borrowers with $DSCR > 1$ have the immediate available liquidity, but may choose to default *strategically.* [4]

We list the possible combinations of LTV and NOI that can *disentangle* Liquidity-constrained Default and the incentives for Strategic Default behavior in Figure 2. For a given Loan-to-Value (LTV) bucket, if default rate monotonically increases in Net Operating Income (NOI), we call those defaulters strategic, as their ability to pay increases with NOI but they still increase their default rate. When $LTV > 1$, the borrower is insolvent and Default Rate increases with NOI in the Bivariate Heatmap in Figure 1a. In fact, DNN algorithm can identify the threshold of NOI^* (**percentile 6** in Trepp data) which disentangles the cases (1) and (2) in Figure 2. When $LTV^*(0.82) < LTV < 1$, the borrower passes on the NOI risk to the lender to maintain equity and there is high default across all borrowers, in anticipation of the abrupt jump to insolvency ($LTV > 1$) in the event of appraisal reduction. When $LTV^{**}(0.6) < LTV < LTV^*(0.82)$, there is no risk of negative equity, hence the borrower negotiates the loan and strategic default starts from low NOI^{**} (**percentile 4.0** in Trepp data) in Figure 1a which disentangles cases (3), (4) in Figure 2. In the LTV bucket $LTV^{***}(0.2) < LTV < LTV^{**}(0.6)$, the borrower default rate is very low as there is neither liquidity-constraint (interest payment is mostly complete) or any incentive for strategic default. When $LTV < LTV^{***}(0.2)$, the loan is close to maturity, hence all borrowers default at a significantly higher rate, due to change in underwriting standard towards maturity or the inability refinance while balloon payment is looming. NOI^{***} can be assumed to be

[3] This is in line with Foote, Gerardi, Goette, and Willen (2009): when equity is negative but above a threshold, default occurs with negative income shock, although our context is commercial real estate (CRE).

[4] Institutional details on Property Type: We assume that property owners manage properties in isolation and do not cross-finance. This is reasonable to assume for investment-type, ring-fenced properties. Some property owners are more constrained for financing than others, e.g., some industries face shock independent of real estate market and suddenly have trouble repaying debt on their buildings. It is reasonable to assume for consumption-type properties held directly be user-firms. In this case, the differences between industries could provide variation in ability.

(**percentile 0**, since all borrowers are strategic defaulters and hence case (6) in Figure 2 is mostly not realized.

DNN can capture the above thresholds of NOI and LTV buckets seamlessly and identify the motives of default, namely liquidity-constrained or strategic. Figure 3a describes the NOI^{***} as percentile 0, NOI^{**} (**percentile 4.0** and NOI^{*} (**percentile 6.0** from the partial dependence plot. It captures the marginal contribution of NOI towards the mean default rate based on NOI, *all possible interactions* with other variables and *all possible highers exponents* of NOI. Similarly, DNN captures the thresholds for LTV to distinguish liquidity-constrained amd strategic defaulters. In Figure 3b, we see the same thresholds for LTV as the bivariate diagram in Figure 2. DNN uses 33 variables and more accurately predicts these thresholds beyond the bivariate diagram using LTV and NOI. No other parametric or non-parametric models can identify these critical thresholds such seamlessly like DNN does.

The importance of these strategic (contractual) variables is captured only in DNN, which cannot be captured in Multinomial Logistic Regression, Lasso, Ridge and even in Distributed Random Forest and Gradient Boosting Machine, as evidenced by the higher ranking of NOI over statistically significant LTV vis-a-vis variable importance tables. [5] We then test the robustness of this higher ranking of NOI over LTV, by leaving out other strategic contractual features (year_month, prepayment penalty, balloon payment, occupancy, appraisal reduction, etc.) one at a time. We further test the robustness of the ranking order of NOI over LTV during the Financial Crisis of 2008 by training the DNN on data from 2000-2006 and testing on 2007-2008 data. We further test this order during COVID-19, the current ongoing pandemic and find the exact same results. This proves beyond any doubt, that NOI (and other contractual features mentioned above) are more important than LTV, even during the dire economic circumstances.

There are often situations in which there are no or few good quasi-natural experiments. [6] We exploit the massive proprietary data set on commercial mortgages from Trepp to disentangle liquidity-constrained default (from lack of Net Operating Income (NOI)) from defaults motivated by strategic behavior, as evidenced by default rate increasing monotonically with increasing NOI, for certain LTV buckets. The context of commercial mortgages borrowers is appropriate to document strategic default as these borrowers are institutions [7]

[5] This could be because: (i) High NOI implies Low Cap Rate for value-added or opportunistic properties compared to core properties; (ii) debt yield from the frothiness of local market; (iii) market cycle channel; (iv) income deficiency related to occupancy.

[6] Differences in Differences require an exogenous treatment and parallel pre-trend, Regression Discontinuity Design requires randomness around one observable characteristic, Instrumental Variables require rigorous explanations on plausibility and ruling out alternative channels, etc.

[7] In an institution, the responsibilities for payment of debt obligation are diffused. The blame for non-payment is not bourne out on one person, but on the institution. Hence there is agency conflict and real

and not households. [8] The commercial borrowers are savvy businessmen and hence their delinquency behavior is possibly much more P&L - oriented [9] based on mortgage *contractual features* (prepayment penalty clause, balloon payment indicator) and *financial constraints*, such as Net Operating Income (NOI), emanating from the unbalance in terms of the amount and time lag between cost of funding and income cash flows, and much less from macroeconomic conditions, supply and demand in the local geography. We contribute an alternative and new DNN approach (see drawbacks and inconsistencies of the other causal mechanisms described in Black, Kim, and Nasev (2012)).

Our paper uses big data with medium frequency to alleviate concerns raised in Manski (2004).[10] We overcome the challenge and observe beliefs [11] and actions in the same data. We implore this novel DNN approach which is much more accurate and robust than Survey measures of (intended) actions and inferring beliefs [12] from actual actions, which assumes that beliefs affect actions. The DNN model can calibrate the thresholds across key variables like NOI, LTV, etc. beyond which there are sharp changes in borrower behavior. The flexibility of not having a pre-specified structure to a model helps us capture univariate and bivariate visualizations of the impact of nonlinearity and higher order interactions.

Goldsmith-Pinkham, Sorkin, and Swift (2018) create of a Bartik instrument (formed by interacting local industry shares and national industry growth rates), where the shares measure differential exposure to common shocks, and identification is based on exogeneity of the shares. Our DNN model can seamlessly exploit the heterogeneous shares of property-type and measure the differential exogenous exposure to the common COVID-19 shock. The variation in cashflow (NOI) from lessee to borrower across property types provides the exogenous heterogeneity across local geography. We no longer need to argue about

possibilities to discern strategic behavior.

[8]Ganong and Noel (2020) find only 3% strategic default for households. Also they define a default as strategic only when the property is under water, i.e., the house value is less than the outstanding loan amount.

[9]Our definition of strategic behavior is not the same as bourne out of strategies in game theory, but is more in line with the mortgage default literature. The Profit and Loss is part of the financial statement that summarizes the revenues, costs, and expenses incurred during a specified period, and hence P&L management becomes important for businesses from a strategic viewpoint.

[10]Beliefs are essential in appreciating the inter-temporal decisions regarding financial choices. The traditional benchmark has been *rational expectations* based on all (publicly) available information, but has scanty evidence in data (Manski (2004)).

[11]The key questions of active research are how beliefs are formed, how different beliefs affect behavior and what are the implications in macroeconomic models and asset pricing.

[12]We contribute to the growing literature of time-inconsistent beliefs or time variation in average beliefs. Giglio, Maggiori, Stroebel, and Utkus (2019) conduct a variance decomposition of beliefs by heterogeneous individual fixed effects, which cannot be explained by observable demographic characteristics. Our paper extends the literature on wealth redistribution between optimists/pessimists (Geanakoplos (2010)), which is a model with constant difference in beliefs.

the plausibility of the identifying assumption, as the flexible nature of DNN provides an alternative identification strategy that is so robust that it weathers 2008 and current crisis from COVID-19 ongoing pandemic.

Beyond the identification strategy from several variables in this big data setting, our DNN methodology also extends the scope of "Frailty Model"[13]. DNN not only captures latent time-fixed macroeconomic effect but also loan specific idiosyncratic effects beyond what has been captured in prior literature in Commercial Mortgages. We include 29 variables from Trepp in our DNN model along with state-level macro variables like unempoyment, GDP growth and 2-Year & 10-Year treasury rates and recently created indices. These 33 variables capture the loan-specific unobserved effects and the macroeconomic variables proxy for unobserved common latent variables. Morover, using the DNN, we can capture the highly non-linear interaction among the covariates. We conduct a horse racing among all the models based on misclassification errors for 7 delinquency classes and conclusively show that DNN has the hightest accuracy of prediction along with Gradient Boosting Machine (GBM) in Table 4. Since GBM is a greedy algorithm, the variable importance is not robust (in Figure 10b) and hence we choose DNN (in Figure 10c) as the best model due to it's interpretability along with the accuracy of prediction.

Our findings yield important new insights into the interplay of borrower behavior, various risk triggers and the macroeconomy. They significantly differ from the findings of Campbell and Dietrich (1983), Cunningham and Capone (1990), Deng (1999), Elul, Souleles, Chomsisengphet, Glennon, and Hunt (2010), Foote, Gerardi, Goette, and Willen (2009), Heimer and Imas (2018), Heimer and Simsek (2019) and others. These prior studies highlight loan level variables such as loan-to-value ratio, loan age, etc., as major predictors of borrower behavior. We test whether by adding macroeconomic variables, we can delve into the realm of omitted variable bias found in all hedonic models. We extend the literature on hedonic models by systematically different macro variables which are exogenous in the hedonic regression model beyond the characteristics (used as covariates) and can explain a lot of the unobserved effects Childs, Ott, and Riddiough (1996). Other than 2Yr Treasury Rate and State Unemployment Rate, the macroeconomic variables do not directly affect the strategic delinquency behavior and timing. National interest rates, e.g., 2-Year and 10-Year Treasury rates impact occupancy of commercial properties directly, as well as through state-level GDP. The unemployment is also captured at the state level. The local State level Unemployment Rate in urban centers and occupancy of lessees in commercial properties are highly

[13] Duffie (2009) created an MCMC methodology that updates the posterior distribution of unobserved risk factors based on Bayes' rule whenever defaults cluster at a given point in time. In the event forecasting literature, such a dynamic unobserved covariate's effect is termed "frailty". Yildirim (2008) also propose a mixture model to disentangle the probability of long term survivorship and the timing of the default event.

correlated, because the lessees are the job-creators locally. We claim that all of the above can be captured by NOI, since both occupancy and unemployment rates affect the NOI from the property, as is bourne out of the variable importable in Figure 10c where NOI is much higher in importance and internalizes the effect of local unemployment and occupancy through non-linear interactions among these variables.

We also add more broadly to the literature on neural networks. Several authors have used shallow neural networks in other areas of financial economics. Bansal and Viswanathan (1993) approximate the pricing kernel using a neural network. Hutchinson, Lo, and Poggio (1994) pioneered the use of neural networks for nonparametric option pricing. Brown, Goetzmann, and Kumar (1998) use neural networks to predict stock markets. Swanson and White (1997) propose the use of neural networks for macroeconomic forecasting. Lee, White, and Granger (1993) construct tests for neglected nonlinearities in time series models using neural networks. Granger (1995) and Kuan and White (1994) study nonlinear or neural network modeling of financial time series. Khandani, Kim, and Lo (2010b) and Butaru, Chen, Clark, Das, Lo, and Siddique (2016) examine other machine learning models of financial default. Recent applications of DNN in financial economics include Klabjan (2007) who model market movements. Heaton, Polson, and Witte (2017) use DNN for portfolio selection. The purpose for a deep learning model is bourne out of the need to have transparency and accountability Albanesi and Vamossy (2019).

2 Commercial Real Estate Vs Household Finance

The residential real estate bubble from 2004 (emergence) to 2008 (burst) has attracted a lot of scholarly work and policy attention. Surprisingly, the commercial real estate price impact (potential bubble) has been ignored in comparison. Since the inception of securitization as a means of financing commercial real estate (CRE) mortgages from 2000, the sophisticated B-piece investors have been outbid beyond sustainable long-run fundamentals, over time (from 2004) by investors who "originated to securitize", thereby, resulting in decline in underwriting standards in CRE (Levitin and Wachter (2012)). Despite some overlap in multi-family property type, Commercial and Residential Real Estate (RRE) is markedly different markets and hence has attracted dissimilar government (e.g., GSE) intervention. One specific difference we will focus on in this paper is the unobservable strategic default behavior of commercial mortgage borrowers, which is different from the residential counterpart due to the nonlinear relationship between multiple property level financials and the mortgage terms. In particular, our goal is to disentangle liquidity-constrained default and the incentives for strategic default based on the Debt-Service Coverage Ratio, a.k.a., DSCR. The the **ability** to pay is turned on when $DSCR \geq 1$ but the *willingness* to pay and motives for strategic default demonstrate the need for a Deep Neural Network (DNN) methodology.

Non-recourse RRE and CRE have an implicit put-option structure equivalent to the repurchase of the loan with the value of the property as the strike, wherein the borrower can satisfy the debt obligation by surrendering the property to the lender. This is the theoretical reason for the **LTV** being the primary driver of default behavior in previous literature (e.g., Ambrose and Jr. (2012), Ambrose, Capone, and Deng (2001)). Since RRE is both investible and consumable, tax-deduction acts as an incentive, and the foreclosure and recourse laws act as disincentives for strategic default for households. Although the individual CRE loans are much bigger in size compared to their residential counterpart, the partially amortizing structure, defeasance, yield maintenance clauses discourage refinancing and/or curtailment, and hence CRE is exposed to **strategic** default, where the borrowers may choose to stay in 90-120 days delinquency bucket strategically, not being liable to be in foreclosure and not having to be REO. We see a huge surge in these loans in this delinquency bucket after 2008 Financial Crisis in Figure 4.

Commercial mortgages are used to finance income-producing properties. Therefore, a borrower's default decision depends on not only the asset value (i.e., borrower equity) but also the property liquidity (i.e., property income). A rational borrower would not default when property net cash flow is positive and is enough to service the scheduled debt obligation, even if the owner's equity position is negative. To properly reflect a rational borrower's default decision, a model for commercial mortgages needs to include both property value and property income as default triggers. Also, unlike residential mortgages that are typically fully amortizing, most commercial mortgages are partially amortizing (7-12 year term and 25-30 year amortization schedule), i.e., a balloon payment is due when the mortgage matures. Borrowers usually fund the balloon payment by refinancing the current mortgage, which may be complicated, and hence **strategic**, at maturity due to higher interest rates or tighter underwriting standards even for a borrower in good standing.

The lender makes a judgement about the riskiness of the borrower in terms of continued payment towards a loan obligation and underwrites the risk vis-a-vis the mortgage rate in Figure 5. Although, the lender is well aware of the reputation and past loan repayment behavior of the borrower, the type of the borrower is noisy and hence this leads to **adverse selection**. The lender makes an actuarily fair **take-it-or-leave-it** offer, adding some risk premium, to the borrower at origination. Moreover, there is competition among borrowers in the same business of leasing income-producing commercial properties. Due to the search friction for the lessee in Figure 5, there is considerable uncertainty about the NOI coming from the lessee's rent payment. The borrowers have no bargaining power in terms of loan pricing, but they could use the act of strategic delinquency as an **insurance policy** against the premium they had to pay at origination via the mortgage rate. We assume that act of strategic delinquency of some borrowers can be captured by the first-order stochastic dominance of the cumulative default rate (higher default rate for strategic defaulters than liquidity-constrained defaulters) of the bad-type over the good-type, for different buckets of LTV in Figure 1a. There could be moral hazard from the lessee in Figure 5, in terms of continued rent payment and servicers are used by the borrower as a commitment device, since the lease is not negotiation-proof.

The commercial lender (debtholder) and the commercial borrower (equityholder) enter into a **contract** (firm) for the business of leasing/renting out the property to a lessee.[14] The prospect of a favorable debt renegotiation not only increases the expected payoff to shareholders in default, but also induces them to anticipate the timing of default, hence increasing the bargaining power of the commercial borrowers (equityholders in this context). There is a subtle difference between **firm** strategic default and the strategic default of the **CRE borrower**, as the CRE borrower usually have different LLCs for each property and hence their strategic default is more property-centric, specifically in terms of the rent payment from the lessee in the property.

Our findings reveal that the effects of property income along with prepayment penalty clause and balloon risk are significant to assess total credit risk adequately.[15] The estimation of the relevant parameters is itself a nontrivial problem, given the sparsity and the diversity of historical CMBS data. The empirical mortgage literature identified a linear combination of variables for the commercial mortgage credit and prepayment risk including creditworthiness and free cash flow of the entity, current leverage ratio, loan age, interest rates, and CMBS indices (e.g. von Furstenberg and George (1969), Curley and Guttentag (1974), Campbell and Dietrich (1983)). The commercial mortgage performance data, however, tell a different story. The presence of nonlinear effects obviates the need for a more general form but it is difficult to identify all the factors and their mutual interactions. Instead of specifying a functional form for commercial mortgage performance, we include all possible factors and let the data dictate the model, which also allows for highly non-linear interaction terms between factors. Since, our data set is nationally representative, the pooled model computes an estimate of aggregate default risk in the commercial mortgages especially well for 2007-

[14]The prospect of a debt reduction through renegotiation may induce shareholders to strategically default even if the firm is solvent [Favara, Schroth, and Valta (2012)].

[15]The complexity of CMBS modeling is due to the simultaneous inclusion of four significant risks: market, credit, prepayment (Christopoulos, Jarrow, and Yildirim (2008)) and liquidity (Ambrose and Sanders (2003)). The cash flows to the underlying CMBS loan pools, the cash flow allocation rules to the various bond tranches, the prepayment restrictions and the prepayment penalties differ across the different CMBS trusts.

2009. Our estimation result provides a ranking of individual commercial mortgages in terms of their delinquency behavior and can be aggregated to a systemic measure of default risk in the commercial sector.

The remainder of the paper proceeds as follows. We first motivate a toy theory model in Section 3, provide details on the big data in Section 4 and provide descriptive statistics and conduct rigorous exploratory analysis to give an idea of the trends in data. In Section 5, we motivate Ordinal, Multinomial Logistic, Lasso & Ridge (which are logistic models with regularization), in Section 6 Distributed Random Forest, Gradient Boosting Machine models and provide empirical results and list the deficiencies in each of them. In Section 7, we describe the DNN model and motivate how the DNN model can alleviate most of the issues in the earlier models. We also point out the key findings of the paper in this section and how they differ from the earlier literature. We test robustness of NOI and LTV order during the ongoing COVID-19 pandemic in Section 8 and document industry-level heterogeneity. In Section 9, we provide concluding remarks followed by references.

3 A Simple Theoretical Motivation

We provide a simple model framework, in line with Guiso, Sapienza, and Zingales (2009), to motivate that the "Optimists" (or Strategic Defaulters) would prefer to maintain a consistently higher LTV during the good portion of business and economic cycle. They will then have the option to strategically default in the future. Whereas, "Pessimists" (Non-Strategic or Liquidity-constrained Defaulters) would prefer to continually reduce LTV, in anticipation of different forces increasing LTV in the future and also to alleviate the consequences of default in the event they are liquidity-constrained. This differential behavior across the cohort of borrowers will price in their heterogenous beliefs (π) in the expectation of occupancy of the property. The differential behavior of these two types of borrowers across different LTV buckets are explained in Figure 1a, based on NOI and based on other variable interactions in the other subfigures in 1. The motivation for using Net Operating Income (NOI), Debt-Service Coverage Ratio (DSCR), Balloon Payment, Scheduled and Unscheduled Payments, Occupancy is explained with the toy model below.

In residential market, while negative equity (whenever the value of the mortgage exceeds the value of the property), in nonrecourse states, is a necessary condition for strategic default Guiso, Sapienza, and Zingales (2009), it is not sufficient. Even in nonrecourse states, there are frictions that make defaulting less appealing. Consider a borrower who at time t owns a property worth A_t and faces a bequest mortgage balloon payment equal to B_T. From a purely financial point of view the borrower will not default as long as $A_t > B_T$. In the decision whether to default strategically, however, there are considerations other than the financial gain or loss from defaulting. For example, by not defaulting, a borrower enjoys the benefit of defaulting in dire conditions in the future. The intertemporal substitution of default choice is co-determined by timing of Appraisal Reduction, Non-recoverability as to whether the Master Servicer/Special Servicer has ceased advancing (P&I and/or Servicing) for the related mortgage loan, etc. Also, by defaulting she faces higher cost of borrowing in the future due to differential credit-rationing by the lender, since lenders are generally NPV-neutral and default is a deadweight loss for them. Let us define K_t as the net benefit (opportunity cost of cash) of not defaulting at t. Then a rational borrower will not default if $A_t - B_T + K_t > 0$.

If the commercial borrower does not have a bequest balloon payment due in the near future, then her decision of whether to default strategically is more complex, because she must trade off the decision to default today with postponing the decision and possibly defaulting tomorrow. In addition, the option to default tomorrow is conditional on the ability (DSCR) of the borrower to serve her mortgage debt, which is highly correlated with the probability of occupancy and positive cash flow from the lessee in the property. If the property is vacant or if the lessee does not pay up, the borrower is likely to default next period and thus loses the value of the option. Let $V_T = A_T - B_T + K_T$, where T is the day the balloon payment is due. Then the value Bajari, Chu, and Park (2008) of not defaulting at T-1 is:

$$V_{T-1} = a_{T-1} - m_{T-1} - B_T + K_{T-1} + (1 - \pi_{T-1}) E_{max}(V_T, 0) \tag{1}$$

where a is the monetary value of the cashflow and the serviceflow enjoyed between time T - 1 and T, m is the mortgage payment (scheduled and unscheduled) between T-1 and T, π_{T-1} is the probability of vacancy of the property (i.e., not having a lessee, and E is the expectation operator. The value of not defaulting at a generic date t can be deduced from backward induction:

$$V_t = a_t - m_t - B_T + K_t + (1 - \pi_t) E_{max}(V_{t+1}, 0). \tag{2}$$

From (1), the decision to default strategically at a generic time t can be described by the following relationship

$$StrategicDefault = F(A - B, a, m, \pi, K). \tag{3}$$

The functional form of $F(.)$ is extremely difficult, if not impossible to pin down. Even locally, to define $F(.)$ piecewise using Implicit Function Theorem, one would need the partial derivative of $F(.)^{-1}$ with respect to the shortfall $A - B$, the monetary value of the cashflow and the serviceflow a, the scheduled and unscheduled mortgage payments m, belief about the property occupancy π, non-monetary benefit K to be well-behaved. We show in 3, this is not the case. The LTV is a function of the shortfall $A-B$, the NOI is a function of the cashflow and the serviceflow a, Debt-Service Coverage Ratio (DSCR) is a function of $\frac{a}{m}$, occupancy is the expectation of the belief π, non-monetary benefits K are mostly unobservable. This is further clouded by the fact that Recourse Laws are not strictly implemented in most states. Bankruptcy Laws need to be fairly strong in a state to reinforce recourse laws.

We give indication from the data, how non-linear the interactions among the above variables can be in Figure 1 and hence resort to the most flexible yet robust DNN methodology in Section 7.

4 Data

We have monthly proprietary novel data set of 91,767 loans (only US loans) from January 2000 to September 2016 from Trepp, the leading provider of analytics, information, and technology to the global CMBS, commercial mortgage finance, and banking industries. We exclude CRE loans, as our research focused on NOI generated from income-producing properties only found in CMBS loans. Trepp is the largest commercially available database containing detailed information on over 1,800 deals and more than 100,000 loans, which support close to $800 billion in securities. Deal coverage includes North American, European, and Asian CMBS, as well as Commercial Real Estate backed CDOs.

We include the variables used in previous CMBS literature, like An, Deng, and Gabriel (2009), Ambrose and Sanders (2003) and preclude the following key loan-specific variables: log(original balance), LTV, time ofamortization, time to maturity, lockout, lockout expiration, corporate bondcredit spread Titman, Tompaidis, and Tsyplakov (2005), yield curve, mortgage-treasury rate spread, region dummy, seasonal/quarter dummy, among others.

We finally decide to use loan-to-value (ltv), occupancy rate (occ), tranche loan-to-value, (securltv), tranche weighted average cost (securwac), annualized gross rate (actrate), outstanding scheduled principal balance at end of current period (obal), derived most recent net operating income (noi), outstanding legal remaining outstanding principal balance reflecting defeasance of the loan as of the determination date (balact), securitization balance of the loan predged to the trust (face), most recent appraised value else securitization appraised value (appvalue), total amount of principal and interest due (actpmt), regularly scheduled principal to be paid to the trust (curschedprin), principal prepayments and prepayments (full or partial), discounted payoffs, and/or other proceeds resulting from liquidation, condemnation, insurance settlements (curunschedprin), interest basis of an adjustable rate loan (pmtbas), net proceeds received on liquidation of loan (liqproceeds), expenses associated with the liquidation (liqexpense), difference between Net Proceeds (after Liquidation Expenses) and Current Beginning Scheduled Balance (realizedloss), amount received from a borrower as a pay off a loan prior to the maturity or anticipated repayment date (pppenalties) as the loan-specific variables. Age of the property is include as a control in addition to the age of the loan. We add age^2 as as a control variable too to capture the non-linear relationship of aging of the loan with the delinquency classes. We calculate "time to maturity" to extract any strategic default behavior closer to the realized maturity of the loans.

We use the loan vintage (to capture if origination and underwriting standards have an effect on the delinquency class of the loans), 51 states in USA (msa, county, zip have severe missing values, hence the identification comes at a state level), property type (we bucket thousands of property types into 8 unique types), fixed/floating as dummy variables. We use "Number of Properties" (numprop) in a deal as a deal-specific variable. We control for refinance pipeline and/or balloon payment by assigning a dummy if a loan is within 3 months threshold to its original scheduled maturity date. We use MIT Commercial Index, National Council of Real Estate Investment Fiduciaries (NCREIF) regional property value indices. Additionally, we include state-level quarterly GDP (converted to monthly), monthly historical unemployment data by state and historical interest rates of different maturities.

For the classifications models to generate realistic results and capture the marginal contributions of the features in a scale-free way, we convert numerical variables like: $x :\longrightarrow \frac{x-min(x)}{max(x)-min(x)}$. This keeps the distributional characteristics of the numerical variables, but makes them all scale-free so that their marginal contributions towards the output can be uniform. We avoid the other more frequently used *standardization* technique where $x :\longrightarrow \frac{x-Mean(x)}{StdDev(x)}$ as it converts all variables into standard normal. The entire valuable information, e.g., skewness, kurtosis and all distributional characteristics are lost in this imposition of normal distribution across feature space.

The summary statistics for the cleaned data containing 9,617,333 observations of continuous variables is provided in Table 1. "One hot encoding" technique converts categorical variables as binary vectors without any order.

5 Parametric Models and Empirical Results

Our first set of empirical results are based on **parametric** models: Ordinal Logistic, Multinomial Logistic, Lasso & Ridge (logistic with regularization), harnessing the unprecedented size of our sample set and the heterogeneity in the incentives of default and beliefs we investigate. The models calculate the accuracy of prediction for 7 different delinquency states starting from Current/Performing classes **W0_30D** which includes "loans with payments not received but still in grace period or not yet due", Late/Non-Performing classes **W30_60D**, **W60_90D** which includes loans with "Late Payment beyond 30-days but less than 60 days, beyond 60-days but less than 90-days, Default state **W90_120D** ((within 90 to 120 days of delinquency), Liquidation Proceedings & Final Resolution state **B120D** (beyond 120 days of delinquency), combined together as "limbo" loans. We add further states in the **PrfMatBal** (Performing, Mature and Balloon Payment due) and **NPrfMatBal** (Non-Performing, Mature and Balloon Payment due) classes to capture the incentives delay in resolution for foreclosed loans to REO/prepaid. Although **PrfMatBal** is a performing loan, but it can be anywhere between 0-90 days of delinquency. **PrfMatBal** are also close to maturity, rendering itself vulnerable to strategic behavior from changing interest rate environment and underwriting standards. Hence, **PrfMatBal** is treated as a separate delinquency status. The same argument holds even stronger for **NPrfMatBal** loans. We motivate below the reasons why these parametric models misrepresent the risk for the delinquent loans in this context.

A **Ordinal** classifier estimates the conditional a-posterior probabilities of a categorical variable given independent covariates using the Bayes rule. The assumption of **independence** of the covariates is key to the success of the Ordinal classifier. We see that **W0_30D**, **W30_60D** & **W60_90D** classes have less mis-classification in Table 4 error in NB than other models, since the assumption of independence among the co-variates holds until a loan is in these classes. This analysis is still kept in the paper to motivate why we eventually need DNN as a means of avoiding this strong assumption of independence among the covariates.

Ordered Logistic Model exploits the natural order of delinquency classes and computes transition probabilities in that order. Ordered Logistic Model does not allow all the back transitions from a worse delinquency state to a better delinquency state, which can be shown in a Finite State Automaton. **Multinomial Logistic Model** assumes Independence of Irrelevant Alternatives (IIA)A.2, which is not true in this situation as we will see in the next section. Suppose, hypothetically, there are two choices given to a borrower to be either **within 30 days of delinquency** or **between 90 days and 120 days of delinquency**, which is not true in this situation as we will see in the next section. Suppose, hypothetically, there are two choices given to a borrower to be either **within 30 days of delinquency** or **between 90 days and 120 days of delinquency**. Clearly, the borrower would like to stick with the first choice, as the second choice classifies him/her in the default category and is detrimental for her creditworthiness from a lender's perspective. Now suppose, one more choice for being in **30 days to 60 days of delinquency** is given to the borrower, s/he may choose to rather be in this new state instead of less than 30 days of delinquency and may **strategically** miss one payment if there is a great investment opportunity for him/her in that one month horizon. In fact, none of the models (except Ordinal) can distinguish these three classes (**W0_30D**, **W30_60D** & **W60_90D**) and considers all of them as **Current Loans** in Table 4.

The granularity of delinquency classes brings out the gradual transition of loans into adverse states rather than simply having a cutoff for default which would imply that we are assuming that loans "Within 30 days delinquency", "Between 30 days and 60 days delinquency" and "Between 60 days and 90 days delinquency" have the same default risk. If all the loans which are less then 90 days delinquent had the same default risk, a borrower would only pay off just before 90 days delinquency in order to avoid default and facing derogatory consequences. The fact that the above three buckets represent different default risk categories imply that the borrower's default behavior will change when she/she is between 30 days and 60 days of delinquency compared to the situation when all the above three categories are bucketed together as "Non-Default".

In Table 4 the row labels are the predicted classes and the column labels are the actual classes. As is evident from the Sensitivity and Error , the Multinomial Logistic Model can correctly classify the Current or "**W0_30D**" really well, but the Specificity is really low, i.e., the model cannot classify the loans that are **not** in "**W0_30D**" correctly vis-a-vis the "**W0_30D**" class. Also the error rates for the classes "**W30_60D**", "**W60_90D**" are 100% which means the model cannot identify any those classes correctly. Similarly, the classes "**W90_120D**" and "**B120D**" are also identified very poorly the Multinomial Logistic Model. In fact, some of the risks (Current Note Rate, LTV, Unemployment Rate, etc.) are misrepresented in Multinomial Logistic Model, e.g., if local Unemployment increases, the *Current* Commercial Loan Default should increase (Table 2). **Lasso** and **Ridge** do not improve the performance of Multinomial Logistic Model in Table 4A.2.

There has been attempts in literature on recursive application of the logistic regression model. The non-linearity and the hierarchical nature of variables in terms of their marginal contributions have been documented in Yildirim (2008). This method uses frailty variables, whose lifetimes are independent conditionally on some common latent factor. The can be thought of as the random effect, used to overcome the unobserved heterogeneity. We also want to point out the reasons for 100% misclassification error rates for the delinquency classes **W030_60D**, **W60_90D** and **B120D**. We use a big data setting in the classification of delinquency classes, and our goal is to accurately classify the severely adverse delinquency states. But this comes at a cost. For **W30_60D** and **W60_90D** mortgage loans, the variation from important variables do not change materially until the loan crosses the 90 day delinquency mark. Hence, with 33 variables used in classification, none of the models are able to accurately classify loans in terms of whether they missed 1 month or 2 months or 3 months of payments. Technically, these loans are not in default and they tend to cure or stay in the same state for a while. These loans do not alarm the lenders. Only Ordinal Logistic Regression is somewhat able to classify these loans with some accuracy because of the inherent worsening of states built into the model, i.e., this method treats **W0_30D** loans as good loans and structures **W30_60D** and **W60_90D** as worse off, rather than treating them as independent alternatives for the loans to be in. The **B120D** loans are very difficult to classify due to the long renegotiation process between the property owners and the lenders in that severe a delinquency status. This has been documented as "limbo" loans in residential mortgages in Allen, Peristiani, and Tang (2015).

6 Vanilla Machine Learning Models & Results

In the current section, we parallelize Random Forest and implement adaptive gradient boosting after bagging. We finally implement DNN in Section 7 and compare the prediction on different mortgage states on the holdout sample.

6.1 Distributed Random Forest

The confusion matrices of the delinquency classes for in-sample/training set are calculated for the entire data in Table 3 and also subsample in Table 3 until the December, 2006 for stress testing the robustness for Out-of Sample Prediction during the Financial crisis in Figure 10. As is evident from the Error in Table 4, the **Distributed Random Forest** Model can correctly classify the Current or "**W0_30D**" **completely** in Table 4. Also the error rates for the classes "**W30_60D**", "**W60_90D**" are 98% which means the model cannot identify any those classes correctly but better than Multinomial Logistic Model. Similarly, the classes "**W90_120D**" and "**B120D**" are also identified very poorly but better than the Multinomial Logistic Model in Appendix A.3.

As is evident from the **Out-of-Sample** Errors in Table 4, the Distributed Random Forest Model can correctly classify the Current or "**W0_30D**" **completely**. here the column labels are the predicted classes and the row labels are the actual classes.Also the error rates for the classes "**W30_60D**", "**W60_90D**" are 100% which means the model cannot identify any those classes any better than Multinomial Logistic Model. Similarly, the classes "**W90_120D**" and "**B120D**" are also identified very poorly but better than the Multinomial Logistic Model **Out-of-Sample**. The Out-of-sample predictions worsen during the Financial

Crisis. [16]

6.2 Gradient Boosting Machine (GBM)

As is evident from the **In-Sample** Errors in Table 4, the **Gradient Boosting Machine** can correctly classify the Current or "**W0_30D**" **completely**. Also the error rates for the classes "**W30_60D**", "**W60_90D**" are almost 100% which means the model cannot identify those classes any better than Multinomial Logistic Model. Similarly, the classes "**W90_120D**" and "**B120D**" are also identified very poorly but better than the Multinomial Logistic Model **In-Sample** in Figure 3. We also attach the Variable Importance for GBM during the using data before Financial Crisis in Figure 10.

Here the column labels are the predicted classes and the row labels are the realized delinquency classes. The out-of-sample predictions for GBM perform as good as DNN in our preliminary analysis. This method uses the same approach as a single tree, but sums the importances over each boosting iteration (see the gbm package vignette)A.4.

7 DNN for disentangling Default Incentives

DNN is a form of machine learning with multiple layers that learns multiple levels of representations for different levels of abstraction Sirignano, Sadhwani, and Giesecke (2016). It captures associations and discovers regularities within sets of patterns; it is suited for high volume, high dimensional data. It performs well when the relationships are dynamic or non-linear in Figure 1, when the standard regression models perform very poorly. No assumptions on normality, linearity, variable independence are needed.

We use a multi-layer feedforward DNN, trained with stochastic gradient descent using back-propagation. Each compute node trains a copy of the global model parameters on its local data with asynchronous multi-threading and contributes periodically to the global

[16] Along with training a model that classifies accurately in a hold-out sample, one needs to be able to interpret the model results. Feature importance is the most useful interpretation tool (such as the coefficients of linear models), to identify important features. Most random Forest (RF) implementations also provide measures of feature importance via permutation importance. Permutation importance is obtained by observing the effect on model accuracy of randomly shuffling each predictor variable. This technique is broadly-applicable because it doesn't rely on internal model parameters even while using Lasso or Ridge regularization in the presence of highly correlated features.

For each tree, the prediction accuracy on the out-of-bag portion of the data is recorded. Then the same is done after permuting each predictor variable. The difference between the two accuracies are then averaged over all trees, and normalized by the standard error. If the standard error is equal to 0 for a variable, the division is not done. here is the Variable Importance table 10a for the Random Forest Model Khandani, Kim, and Lo (2010a). The Variable Importance for Out-of-Sample predictions during the Financial Crisis in Figure 10 give similar results.

model via model averaging across the DNN. We tune both and Optimizer and Model-specific Hyperparameters (described in Appendix A.6). We use SMOTE technique to reduce class imbalance. According to Appendix A.7 [17], we use Variable Importance to compare the most significant marginal contributions of the features (described in Appendix ??).

7.1 Model Results

We motivate the highly strategic delinquency behavior of the savvy commercial borrowers/ business-owners from two different angles. We provide evidence from the Trepp data in Figure 6a that from 2012, the number of loans have remained flat but the outstanding balance of loans have steadily increased until 2016. This could have serious implications. There can only be two possibilities: if the same loans stay and there is no origination at all, and further if the outstanding balance is increasing, it means there is serious delinquency in the loans and the servicers are unable to secure the payment from the borrower and all these loans could potentially become limbo loans. Figure 6b furthers the narrative. From mid-2014, the age of the loans is decreasing and the time-to-maturity is increasing. This could mean that from mid-2014, there are an equal number of originations to the number of maturing loans. But the fact that the Outstanding Balance is increasing in this entire period could only mean that the same loans are getting rolled over to new contracts, when balloon payments are missed during maturity.

Figure 6c clearly shows that LTV (widely used in previous literature and used by most banks/asset managers for credit risk calculations) is flat throughout the data horizon. The interest rate is decreasing almost monotonically in the data and there seems to be no sensitivity of LTV to interest rate. This means LTV is probably not the right way to think about credit risk. It could also be that the commercial borrowers **target** LTV. They strategically make payments towards their obligation so that the ratio of "Book Value of Loan" and the "Value of the Property" remains relatively stable over time. It would make sense for them to do this as banks/asset managers use LTV at origination as the primary determinant of creditworthiness of the borrowers. Further, the Contemporaneous LTV (CLTV) is used to calculate LGD (Loss given Default or 1-Recovery Rate). So, CLTV could also be targeted and there is no evidence of voluntary deleveraging from the borrower inspite of widely changing macro-economic conditions, e.g., interest rate.

Figure 6d corroborates that the NOI monotonically increases in the data and the occupancy is almost 100% in the entire data. So, there may be strategic saving of internal cash

[17] As is clear from the similar counts of the loans of different categories in the in-sample confusion matrix in Table 3, we have **undersampled** the W0_30D class/Current Loans to alleviate the class imbalance problem. The Out-of-Sample predictions across different delinquency classes are as good as GBM in Table 4.

flow from income producing properties. Because of the strictly increasing NOI level, the strategic dominance of NOI over other factors can have disastrous aggregate macroeconomic consequences. To capture this, we try different methodologies like vanilla models (Ordinal Logistic) and machine learning models (Distributed Random Forest, Gradient Boosting) and finally Deep Neural Network (DNN) and find that DNN is best positioned to address the above issue and does capture NOI as the most significant strategic variable from the Variable Importance (VI) tables of the models. This difference does not stem from sample bias. This is the core reason for our choice of big data for training all the models. Also, Trepp is the largest provider of CMBS data and hence the sample is representative of the entire market and does not have any selection bias.

We normalize Net Operating Income (NOI) as a percentage in the pooled data for loans and create histograms of relative frquency of the number of loans in different delinquency classes with respect to the different percentiles of NOI. We see a sheavy support for the relative frequency across all the delinquency classes at the NOI percentages 5%-7%. We call them **dominant** NOI buckets. We show the distribution of different delinquency classes with all the NOI buckets including the dominant ones (see figure 7a). The significant heterogeneity across the delinquency classes and the highly non-linear effect of NOI towards the strategic choice of the borrower to be in a specific delinquency class is not bourne out of this diagram. Beyond the above dominant buckets, we see highly non-linear **strategic** behavior for commercial mortgage borrowers to choose different delinquency classes for different buckets of NOI. To visualize this, we zoom in and remove the dominant buckets and form the **rescaled** (without dominant NOI bucket masses) relative frequency histogram across all delinquency classes.

Figure 7b highlights the complex relationship that exists between the percentage of loans across the different delinquency classes "Within 30 Days"(**W0_30D**), "30 Days to 60 Days"(**W30_60D**), "60 Days to 90 Days"(**W60_90D**), "90 Days to 120 Days"(**W90_120D**), "Beyond 120 Days" (**B120D**) and the buckets of net operating income (NOI) excluding the dominant NOI buckets, which can be incentivized by the macro-economy. The sensitivity varies significantly in a highly non-linear way in both magnitude and sign. There is a **U-shaped** choice between NOI buckets 37%-45%for the borrowers in the delinquency class W90_120D. This means that when a borrower is already beyond the default threshold of 90 days, but less than the cutoff of 120 days, they are incentivized to stay there for a while and time their future payments based on cash flow. Since these NOI bucktes are higher, the borrowers make some profit from the income generated from the property, but they still stay at the same delinquency classes and do not pay-off the earlier missed payments to come back to the Current State (less than 30 days of delinquency). Similarly, the borrowers in delinquency class B120D choose to be in lower NOI buckets in a non-linear way. This is because of the lack of net cash flow income for them to be able to pay off the earlier missed payments. They end up in a vicious cycle of making less money from the property and becoming worse off in terms of their creditworthiness. We call them "limbo" loans as these loans stay in this state for a while before they are resolved. The sensitivity estimates generated by vanilla models can misrepresent the influence of risk factors because of naive choice of linear specification. This can make it difficult to make economic conclusions from the borrower behavior. In our approach, the relationship is entirely dictated by data, thereby minimizing model misspecification and bias of the variable estimates.

The accuracy of predictions change dramatically, if NOI is taken out. We also conduct a robustness check by leaving out each of the *strategic* variables from the DNN model. When year and month fixed effects are taken out in Figure 9, NOI loses its importance significantly! This clearly indicates that NOI is not a statistically significant variable by itself. It is used strategically by borrowers when clustering of macro-economic events happen and when NOI is taken out, the constraint variable like prepayment penalty clause and voluntary prepayment variable like current unscheduled pricinpal payment show up higher in the variable importance in Table 4 than LTV. Similarly, when Prepayment Penalties are taken out of the list of variables. When Balloon Payment constraints are taken out of the list of variables.

For neural networks, two popular methods for constructing Variable Importance (VI) scores are the Garson algorithm, later modified by Goh (1995), and the Olden algorithm Olden, Joy, and Death (2004). For both algorithms, the basis of these importance scores is the network's connection weights. The Garson algorithm determines VI by identifying all weighted connections between the nodes of interest. Olden's algorithm, on the other hand, uses the product of the raw connection weights between each input and output neuron and sums the product across all hidden neurons. This has been shown to outperform the Garson method in various simulations. For DNNs, a similar method due to Gedeon (1997) considers the weights connecting the input features to the first two hidden layers (for simplicity and speed); but this method can be slow for large networks. For Deep Learning, there is no impact of scaling, because the numbers were already scaled. hence, the relative importance is the same as the absolute importance in Figure 9.

Compared to model-specific approaches, model-agnostic interpretation via VI methods are more flexible (since they can be applied to any supervised learning algorithm). We intend to further investigate model-agnostic methods for quantifying global feature importance using three different approaches: 1) PDPs, 2) ICE curves, and 3) permutation Greenwell, McCarthy, Boehmke, and Liu (2018).

As is clear from the preliminary analysis, the Net Operating Income (NOI), the Prepayment Penalty clause and the Balloon Payment triger are significantly high in the variable importance table 4. NOI is even higher than LTV as found in the VI tables for other previous models. This provides evidence on how these three less statistically significant features contribute much more towards the classification, via highly non-trivial and non-linear interactions with more statistically significant variables.

Our DNN Variable Importance table in Figure 9 shows that NOI is the key endogenous feature for understanding strategic delinquency behavior of the commercial mortgage borrowers. We intend to further investigate how prepayment penalty clause and indicator for balloon payments co-determine the strategic delinquency behavior along with the NOI using Shapley values by capturing the marginal contributions.

To test the robustness and stability of our DNN, we present the Variable Importance Plots of Predicted Default Rate from June 2006 to December 2008 with several features in Distributed Random Forest (DRF) in Figure 10a, Gradient Boosting Machine (GBM) in Figure 10b and DNN in Figure 10c, trained on data before June 2006 and motivate why we need a highly non-linear model and also why we allow for high-dimensional interaction among the borrower-specific, macroeconomic, spatial, vintage effects in the features. Time-to-Maturity, Geographical cross-correlation, NOI, Appraisal Reduction, Bankruptcy Flag, Property Type, Non-Recoverability, Appraised Value supercede Securitized LTV in the Variable Importance chart for DNN in Figure 10c. Morover, Balloon Payment supercedes LTV, corroborating the robustness of our DNN model. DRF captures non-lineary of the covariates but still ranks LTV much above NOI and other strategic variables in Figure 10a, even after tuning and bagging. GBM is a greedy algorithm and hence finds more occurances of local minima for LTV and hence ranks LTV higher than NOI in Figure 10b, even after boosting.

8 COVID-19 Results without internalizing 2008 Crisis

After cleaning the data, we have 1,315,421 observations from Jan 2017 - Sep 2020. We first try training the DNN model with data from Jan 2017 - Feb 2020 and confirm that NOI is more important than LTV leading upto the COVID-19 pandemic in Table 5. However, consistent with our conjecture that Commercial Mortgage delinquency leads a financial (in this case, induced by a global health crisis), the out-of sample predictions are inaccurate. Then we train the DNN model from Jan 2017 - Nov 2019, as the first cases of COVID-19 were identified in Wuhan, China during December, 2019. We still find that NOI is more important than LTV in Table 6 and also provide evidence about other variables which become important in co-determining the default behavior. We find similar accuracy in predictions in Table 7 compared to Table 4. This solidifies our previous and we can indeed claim that DNN extracts the inherent structural relationship among the covariates and can robustly predict even during several financial crises.

8.1 Results by Industry during COVID-19

We track the evolution of the number of loans in different delinquency classes across time (Dec 2019 - Sep 2020) and across industries (Assisted, Healthcare, Hotel, Industrial, Multi-family, Office, Special) in Figure 11. We clearly see the hotel industry being massively displaced. We see enormous number of commercial mortgage loans in the hotel industry degrade from *W0_30D* to *W30_60D* (orange bar) from May, 2020. There is some degradation from *W30_60D* (orange bar) to *W60_90D* (ash-color bar). These loans become limbo from August 2020 as evident from the rise in dark-blue bar of *B120D* loans. The office space has also seen massive cashflow shortages from the lack of business activities and inability of payments therefrom for tenants in big cities. This is evidenced by constant green bar (Non-Performing Mature Balloon Loans) which are trying to roll-over the loan contract and some oare successful from the lax underwriting standards due to historical levels of low interest rate monetary policy.

8.2 Determinants of Each Delinquency Class

We zoom in to individual delinquency classes and find the determinants of each adverse delinquency class using DNN. In Table 9, we still find NOI higher than LTV for DNN trained on Jan 2017 - Nov 2019 and misclassification error is almost 0%. For Real-Estate-owned (REO) loans, the misclassification error is again 0% in Table 10, but NOI is no longer higher than LTV in Table 11 as there is no NOI when the lender takes back the property. For foreclosure, again the misclassification error is again 0% in Table 12 and similarly NOI is no longer higher than LTV in Table 13 as foreclosure proceedings are lengthy processes which start after Bankruptcy Chapter 13 and there is no renegotiation to be done and either the loan gets resolved in court or leads to the terminal state of Bankruptcy Chapter 7 (liquidation). As shown in Subsection 8.1, the behavior of W90_120D and B120D loans are determined by industry heterogeneity. There is slightly higher error (in Tables 15 and 17) for these two uncertain delinquency states as the COVID-19 is an ongoing pandemic and still unfolding.

9 Conclusion

Using DNN, non-linearities of dependence of the response and interactions among features can be captured, without specifying the relationships apriori. DNN provides an alternative identification strategy, specially when there are no available quasi-natural experiments. Net Operating Income, Prepayment Penalty Clause, Appraisal Reduction, Non-Recoverability, Bankruptcy Flag, Liquidity Proceeds, Liquidity Expense and Balloon Payment Indicator co-determine the strategic delinquency behavior of a commercial mortgage borrower. Loan-to-Value is unable to capture this Strategic behavior as obviated by the Variable Importance charts since statistical significance cannot capture the non-linear effect during Financial Crisis. Hyperparameter Tuning during the implementation of DNN is still an art and not a science. The classification of critical delinquency states of systems when the agent decisions are endogenous while the data is highly unbalanced across states can only be captured through DNN.

References

Albanesi, Stefania, and Domonkos Vamossy, 2019, Predicting Consumer Default: A Deep Learning Approach, p. 70.

Allen, Linda, Stavros Peristiani, and Yi Tang, 2015, Bank Delays in the Resolution of Delinquent Mortgages: The Problem of Limbo Loans, *Journal of Real Estate Research* 37, 65–116.

Ambrose, Brent W., Jr. Charles A. Capone, and Yongheng Deng, 2001, Optimal Put Exercise: An Empirical Examination of Conditions for Mortgage Foreclosure, *The Journal of Real Estate Finance and Economics* 23, 213–234.

Ambrose, Brent W., and Richard J. Buttimer Jr., 2012, The Adjustable Balance Mortgage: Reducing the Value of the Put, *Real Estate Economics* 40, 536–565.

Ambrose, Brent W., and Anthony B. Sanders, 2003, Commercial Mortgage-Backed Securities: Prepayment and Default, *The Journal of Real Estate Finance and Economics* 26, 179–196.

An, Xudong, Yongheng Deng, and Stuart A. Gabriel, 2009, Value Creation through Securitization: Evidence from the CMBS Market, *Journal of Real Estate Finance and Economics* 38, 302–326.

Bajari, Patrick, Chenghuan Sean Chu, and Minjung Park, 2008, An Empirical Model of Subprime Mortgage Default From 2000 to 2007, NBER Working Papers 14625 National Bureau of Economic Research, Inc.

Bansal, and Viswanathan, 1993, No Arbitrage and Arbitrage Pricing: A New Approach, *Journal of Finance* 48, 1231–1262.

Black, Bernard S., Woochan Kim, and Julia Nasev, 2012, The Effect of Board Structure on Firm Disclosure and Behavior: A Case Study of Korea and a Comparison of Research Designs, *ECGI - Finance Working Paper* p. 52.

Brown, Stephen, William Goetzmann, and Alok Kumar, 1998, The Dow Theory: William Peter Hamilton's Track Record Reconsidered, *Journal of Finance* 53, 1311–1333.

Butaru, Florentin, Qingqing Chen, Brian Clark, Sanmay Das, Andrew W. Lo, and Akhtar Siddique, 2016, Risk and risk management in the credit card industry, *Journal of Banking & Finance* 72, 218–239.

Campbell, Tim S, and J Kimball Dietrich, 1983, The Determinants of Default on Insured Conventional Residential Mortgage Loans, *Journal of Finance* 38, 1569–81.

Chawla, Nitesh V., Kevin W Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer, 2002, SMOTE: Synthetic Minority Over-sampling Technique, *J. Artif. Intell. Res.* 16, 321–357.

Chen, Hugh, Scott Lundberg, and Su-In Lee, 2018, Hybrid Gradient Boosting Trees and Neural Networks for Forecasting Operating Room Data, *CoRR* abs/1801.07384.

Chen, Tianqi, and Carlos Guestrin, 2016, XGBoost: A Scalable Tree Boosting System, *CoRR* abs/1603.02754.

Childs, Paul, Steven H. Ott, and Timothy J. Riddiough, 1996, The Pricing of Multiclass Commercial Mortgage-Backed Securities, *The Journal of Financial and Quantitative Analysis* 31, 581–603.

Christopoulos, Andreas D., Robert A. Jarrow, and Yildiray Yildirim, 2008, Commercial Mortgage-Backed Securities (CMBS) and Market Efficiency with Respect to Costly Information, *Real Estate Economics* 36, 441–498.

Cunningham, Donald, and Charles Capone, 1990, The Relative Termination Experience of Adjustable to Fixed-Rate Mortgages, *Journal of Finance* 45, 1687–1703.

Curley, Anthony J., and Jack M. Guttentag, 1974, The Yield on Insured Residential Mortgages, in *Explorations in Economic Research, Volume 1, Number 1* (National Bureau of Economic Research, Inc,).

Daelemans, Walter, Bart Goethals, and Katharina Morik (eds.), 2008, *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML/PKDD 2008, Antwerp, Belgium, September 15-19, 2008, Proceedings, Part I* vol. 5211 of *Lecture Notes in Computer Science*. Springer.

Deng, Yong, 1999, Network Power: Japan and Asia. Edited by Katzenstein Peter and Shiraishi Takashi. Ithaca, NY: Cornell University Press, 1997. 399p. $55.00 cloth, $22.50 paper, *American Political Science Review* 93, 226–227.

Duffie, Darell, 2009, Frailty Correlated Default, *Journal of Finance* LXIV, 34–52.

Elul, Ronel, Nicholas Souleles, Souphala Chomsisengphet, Dennis Glennon, and Robert Hunt, 2010, What "Triggers" Mortgage Default?, *American Economic Review* 100, 490–94.

Favara, Giovanni, Enrique Schroth, and Philip Valta, 2012, Strategic Default and Equity Risk Across Countries, *Journal of Finance* 67, 2051–2095.

Foote, Christopher, Kristopher Gerardi, Lorenz Goette, and Paul Willen, 2009, Reducing Foreclosures: No Easy Answers, NBER Working Papers 15063 National Bureau of Economic Research, Inc.

Foster, Chester, and Robert Van Order, 1984, An Option-Based Model of Mortgage Default, *Housing Finance Review* pp. 351–372.

Friedman, Jerome H., 2000, Greedy Function Approximation: A Gradient Boosting Machine, *Annals of Statistics* 29, 1189–1232.

Ganong, Peter, and Pascal Noel, 2020, Why Do Borrowers Default on Mortgages?, p. 43.

Geanakoplos, John, 2010, The Leverage Cycle, *NBER Macroeconomics Annual* 24, 1–66.

Gedeon, Tamás D., 1997, Data Mining of Inputs: Analysing Magnitude and Functional Measures, *Int. J. Neural Syst.* 8, 209–218.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, 2019, Five Facts about Beliefs and Portfolios, NBER Working Papers 25744 National Bureau of Economic Research, Inc.

Goh, A. T. C., 1995, Back-propagation neural networks for modeling complex systems, *AI in Engineering* 9, 143–151.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, 2018, Bartik Instruments: What, When, Why, and How, NBER Working Papers 24408 National Bureau of Economic Research, Inc.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville, 2016, *Deep Learning*. The MIT Press.

Granger, Clive, 1995, Modelling Nonlinear Relationships between Extended-Memory Variables, *Econometrica* 63, 265–79.

Greenwell, Brandon M., Andrew J. McCarthy, Bradley C. Boehmke, and Dungang Liu, 2018, Residuals and Diagnostics for Binary and Ordinal Regression Models: An Introduction to the sure Package, *The R Journal* 10, 381–394.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2009, Moral and Social Constraints to Strategic Default on Mortgages, Working Paper 15145 National Bureau of Economic Research.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2013, The Determinants of Attitudes toward Strategic Default on Mortgages, *Journal of Finance* LXVIII.

Heaton, J. B., N. G. Polson, and J. H. Witte, 2017, Deep learning for finance: deep portfolios, *Applied Stochastic Models in Business and Industry* 33, 3–12.

Heimer, Rawley, and Alex Imas, 2018, Biased By Choice: How Financial Constraints Can Reduce Financial Mistakes, *9th Miami Behavioral Finance Conference 2018* pp. 1–83.

Heimer, Rawley, and Alp Simsek, 2019, Should retail investorsâ leverage be limited?, *Journal of Financial Economics* 132, 1–21.

Hornik, Kurt, 1991, Approximation capabilities of multilayer feedforward networks, *Neural Networks* 4, 251 – 257.

Hutchinson, James M, Andrew Lo, and Tomaso Poggio, 1994, A Nonparametric Approach to Pricing and Hedging Derivative Securities via Learning Networks, *Journal of Finance* 49, 851–89.

Indarte, Sasha, 2020, Moral Hazard versus Liquidity in Household Bankruptcy, *Working Paper* p. 77.

Kahn, R, and Toni M Whited, 2017, Identification Is Not Causality, and Vice Versa, *The Review of Corporate Finance Studies* 7, 1–21.

Khandani, Amir, Adlar Kim, and Andrew Lo, 2010a, Consumer Credit Risk Models Via Machine-Learning Algorithms, *SSRN Electronic Journal*.

Khandani, Amir E., Adlar J. Kim, and Andrew Lo, 2010b, Consumer credit-risk models via machine-learning algorithms, *Journal of Banking and Finance* 34, 2767–2787.

King, Gary, and Langche Zeng, 2001, Logistic Regression in Rare Events Data, *Political Analysis* 9, 137–163.

Klabjan, Diego, 2007, Subadditive approaches in integer programming, *European Journal of Operational Research* 183, 525–545.

Kuan, Chung-Ming, and Halbert White, 1994, Adaptive Learning with Nonlinear Dynamics Driven by Dependent Processes, *Econometrica* 62, 1087–1114.

Kuncheva, Ludmila I.and Whitaker, Christopher J., 2003, Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy, *Machine Learning* 51, 181–207.

Kvamme, Håvard, Nikolai Sellereite, Kjersti Aas, and Steffen Sjursen, 2018, Predicting mortgage default using convolutional neural networks, *Expert Syst. Appl.* 102, 207–217.

Lee, Tae Hwy, Halbert White, and Clive Granger, 1993, Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests, *Journal of Econometrics* 56, 269–290.

Levitin, Adam, and Susan Wachter, 2012, Explaining the Housing Bubble, p. 36.

Manski, Charles F., 2004, Measuring Expectations, *Econometrica* 72, 1329–1376.

Montufar, Guido F, Razvan Pascanu, Kyunghyun Cho, and Yoshua Bengio, 2014, On the Number of Linear Regions of Deep Neural Networks, in Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds.: *Advances in Neural Information Processing Systems 27* (Curran Associates, Inc.,).

Olden, Julian D., Michael Kevin Joy, and Russell George Death, 2004, An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data, .

Riddiough, Timothy, 1991, Equilibrium mortgage default pricing with non-optimal borrower behavior, *PhD Dissertation, Univseristy of Wisconsin*.

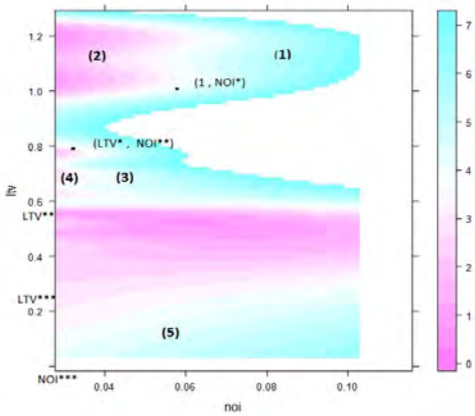
Sirignano, Justin, Apaar Sadhwani, and Kay Giesecke, 2016, Deep Learning for Mortgage Risk, *SSRN Electronic Journal*.

Swanson, Norman, and Halbert White, 1997, A Model Selection Approach To Real-Time Macroeconomic Forecasting Using Linear Models And Artificial Neural Networks, *The Review of Economics and Statistics* 79, 540–550.

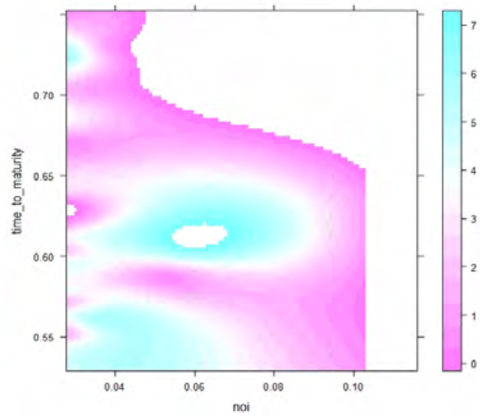
Titman, Sheridan, Stathis Tompaidis, and Sergey Tsyplakov, 2005, Determinants of Credit Spreads in Commercial Mortgages, *Real Estate Economics* 33, 711–738.

von Furstenberg, and George, 1969, Default Risk on FHA-Insured Home Mortgages as a Function of the Terms of Financing: A Quantitative Analysis, *Journal of Finance* 24, 459–77.

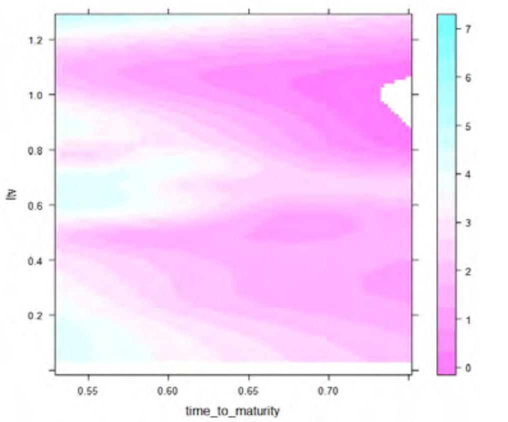
Yildirim, Yildiray, 2008, Estimating Default Probabilities of CMBS Loans with Clustering and Heavy Censoring, *The Journal of Real Estate Finance and Economics* 37, 93–111.



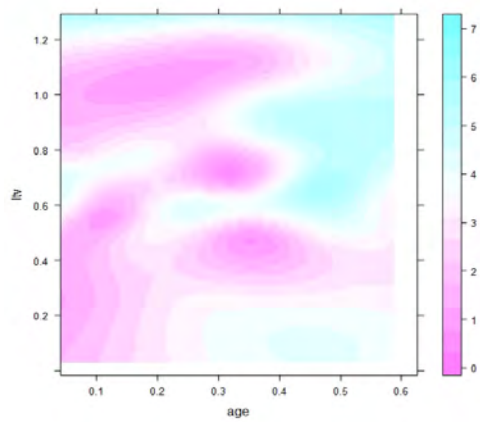
(a) Default Rate Vs NOI & LTV



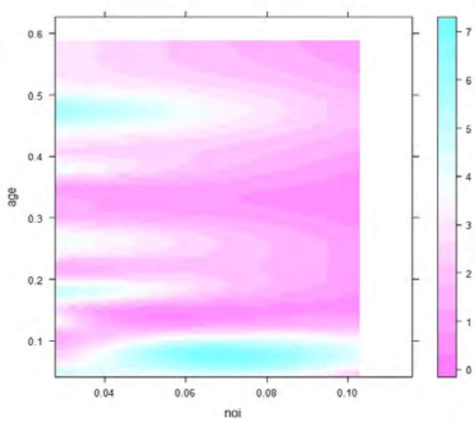
(b) Default Rate Vs NOI & Time to Maturity



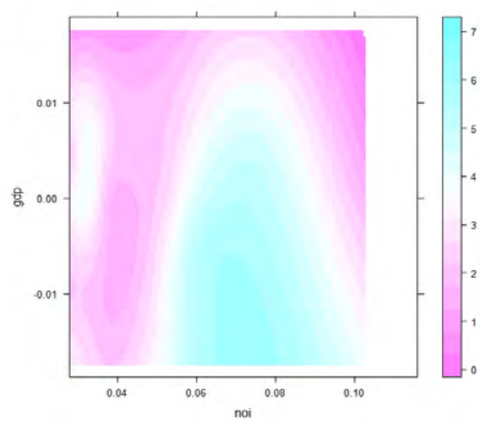
(c) Default Rate Vs Time to Maturity & LTV



(d) Default Rate Vs Age & LTV



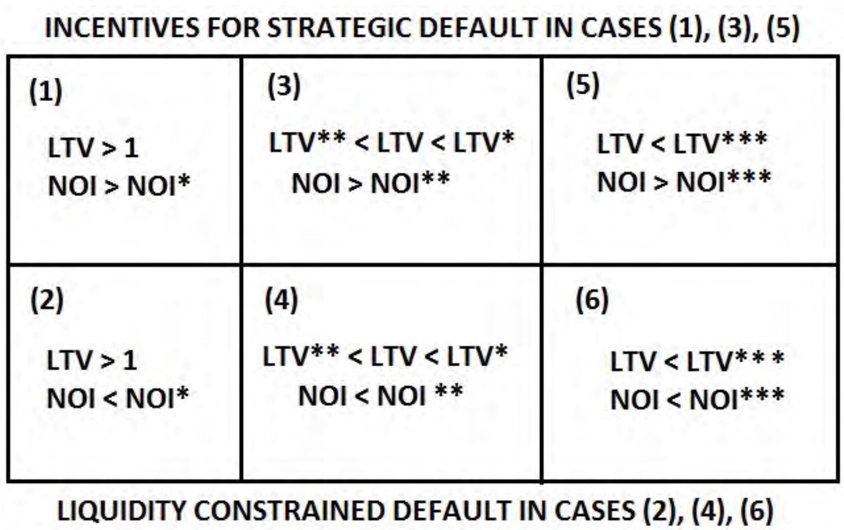
(e) Default Rate Vs NOI & Age

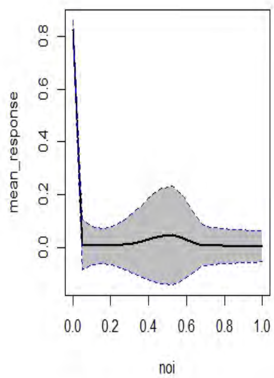


(f) Default Rate Vs NOI & GDP

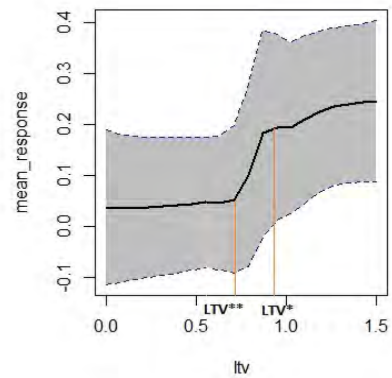
Figure 1: Bivariate Heatmaps are 2D projections of Default Rate surface over two co-vaoriates. Blue color signifies high default rate and pink color low default. **DSCR** = Debt-Service Coverage Ratio & **NOI** = Net Operating Income.

Figure 2: We list the possible combinations of LTV and NOI that can *disentangle* Liquidity-constrained Default and the incentives for Strategic Default behavior in this figure. We use DNN and show that case (2) is liquidity-constrained default in the Partial Dependence Plot (PDP) in Figure 3a since there is no spread in terms of predictability of NOI. The effect is verified by ensuring high LTV values (under-water properties with $LTV > 1$) in the Bivariate Heatmap in Figure 1a. In fact, DNN algorithm can identify the threshold of NOI^* (**percentile 6** in Trepp data) in Figure 3a. Case (1) is more interesting since there is some spread in default predictability w.r.t NOI. We claim from Figure 3b, this is where strategic defaulters cannot be identified from non-strategic defaulters after $NOI > NOI^*$ (**percentile 6** in Trepp data). But, most likely a strategic defaulter would not default in Case (2) to have the option to default in Case (1). This gives us a mechanism to identify the strategic defaulters from non-strategic ones once the threshold is identified. Cases (3), (4) in Figure 2, described in Figure 1a, behave in a similar way since LTV is still very high (LTV higher than **0.82** but less than 1, i.e., property is not under-water). In fact, DNN shelps us identify LTV^* (**0.82**), NOI^{**} (**percentile 4.0** in Trepp data) first in subfigures 3a for NOI and 3b for LTV and then in Figure 1a based on the interaction between LTV and NOI. For LTV bucket between 0.6 and 0.82, financial friction acts as a liquidity-constraint for non-strategic defaulters. Strategic Defaulters are also in this cohort and NOI^{**} (**percentile 4.0** in Trepp data) determines (in Figure 1a) the cutoff beyond which again the behavior of Strategic Defaulters from the Non-Strategic defaulters. The heterogeneity occurs because of constraint on time or limited attention for Non-Strategic Defaulters. Cases (5), (6) (in Figure 2) are much more interesting and there is a whole host of factors that we conider in DNN to identify the strategic defaulters and their incentives. Again, DNN indicates a *possible* LTV^{***} (**0.2**) but NOI^{***} can be assumed to (**percentile 0**) since these are strategic defaulters due to bequest balloon payment due, evident from the lower portion of Figure 1a.

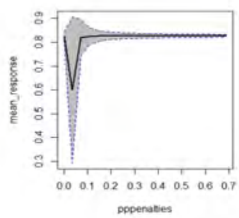




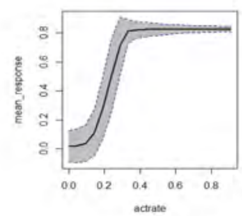
(a) Default Rate Vs Net Operating Income



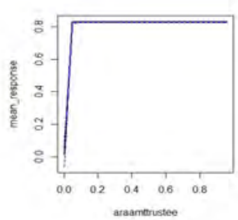
(b) Default Rate Vs Loan-to-Value



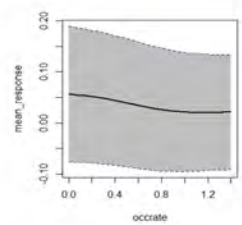
(c) Default Rate Vs Prepayment Penalties



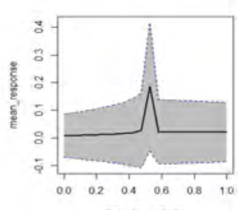
(d) Default Rate Vs Current Note Rate



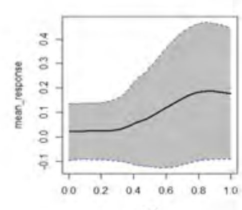
(e) Default Rate Vs Appraisal Reduction Amount



(f) Default Rate Vs Occupancy Rate



(g) Default Rate Vs Time to Maturity



(h) Default Rate Vs Age of Loan

Figure 3: Partial Dependence Plots for Predicted Default Rate

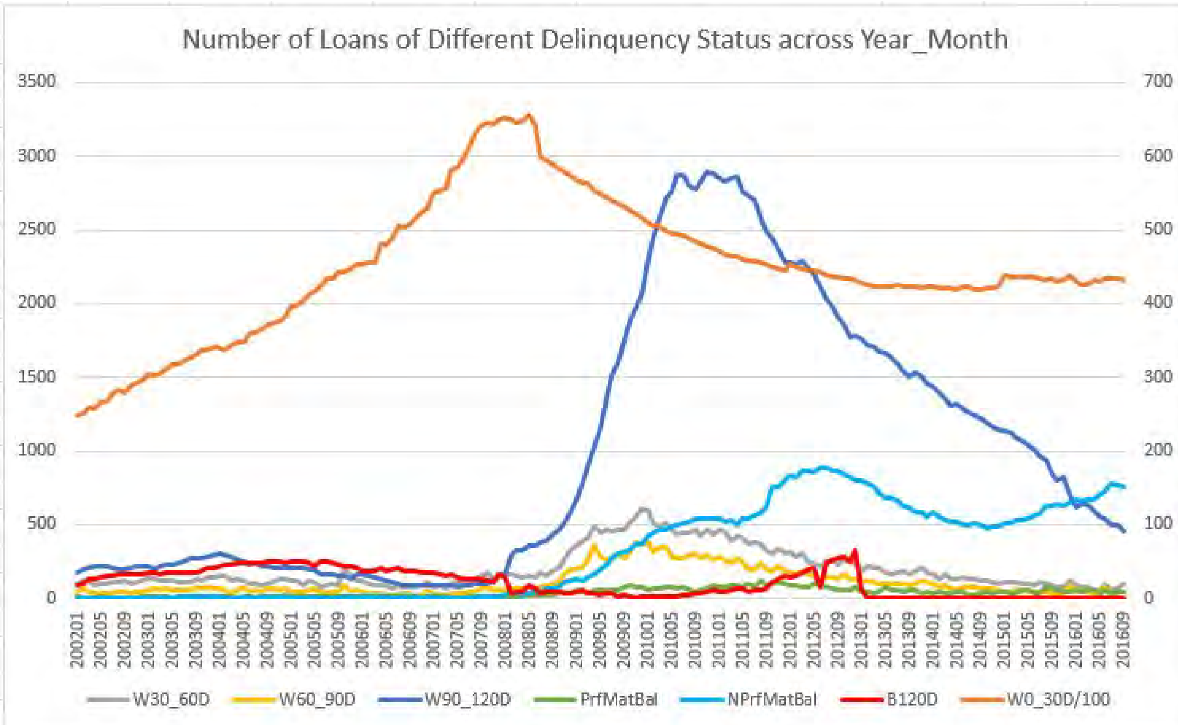


Figure 4: This diagram provides the evolution of delinquency buckets by year. The delinquency states are Current/Performing class **W0_30D** which includes "loans with payments not received but still in grace period or not yet due" (marked on the primary axis, as current loans constitute most of the sample), Late/Non-Performing classes **W30_60D**, **W60_90D** which includes loans with "Late Payment beyond 30-days but less than 60 days, beyond 60-days but less than 90-days, Default class **W90_120D** ((within 90 to 120 days of delinquency), Liquidation Proceedings & Final Resolution state **B120D** (beyond 120 days of delinquency), combined together as "limbo" loans. We try further states in the **PrfMatBal** (Performing, Mature and Balloon Payment due) and **NPrfMatBal** (Non-Performing, Mature and Balloon Payment due) classes to capture the incentives delay in resolution for foreclosed loans to REO/prepaid. The number of loans for all delinquency buckets other than **W0_30D** are on the secondary axes.

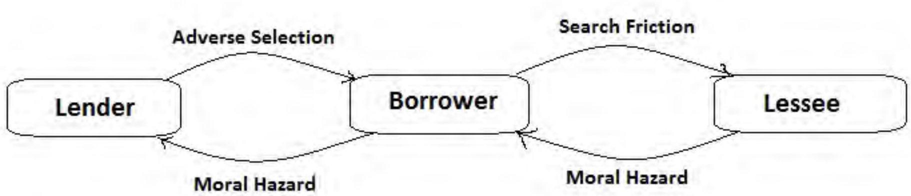
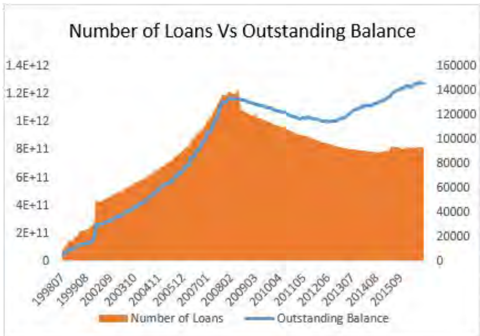
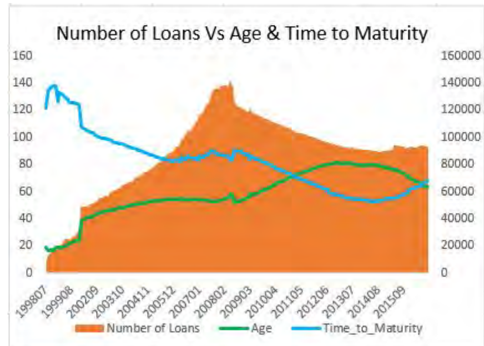


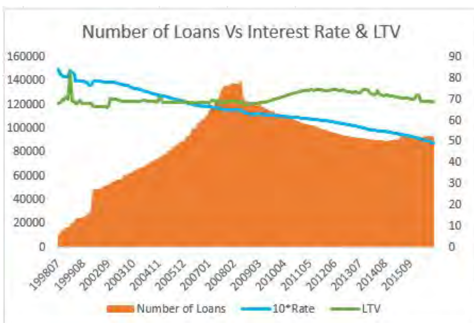
Figure 5: Adverse Selection and Moral Hazard



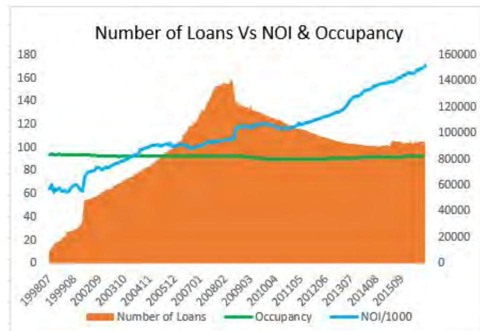
(a) Number of Loans vs. Outstanding Loan Balance



(b) Number of Loans vs. Age and Time to Maturity



(c) Number of Loans vs. Interest Rate and LTV

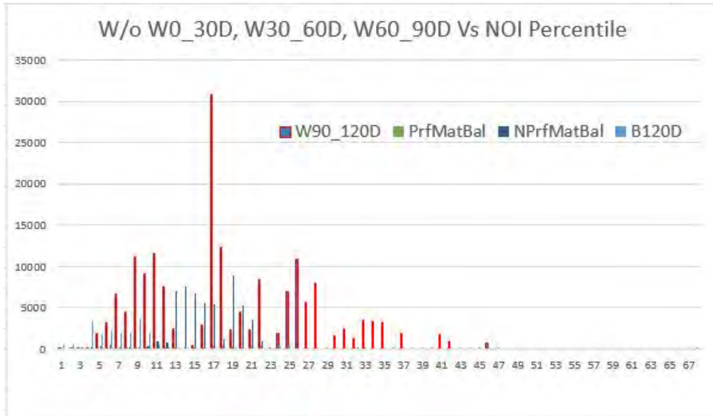


(d) Number of Loans vs. NOI and Occupancy

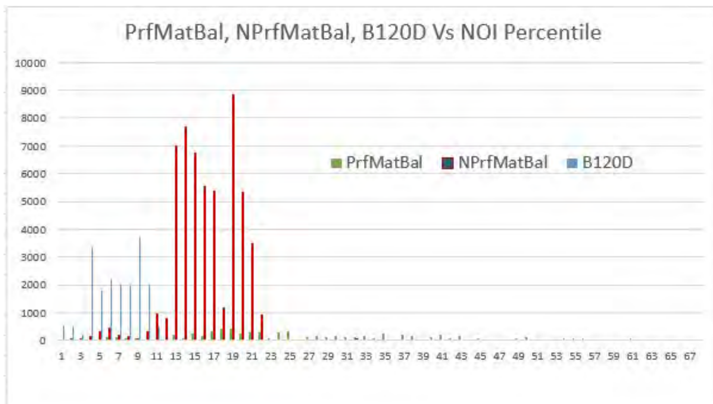
Figure 6: We provide evidence from the Trepp data in Figure 6a that from 2012, the number of loans have remained flat but the outstanding balance of loans have steadily increased until 2016. Figure 6b furthers the narrative. From mid-2014, the age of the loans is decreasing and the time-to-maturity is increasing. This could mean that from mid-2014, there are an equal number of originations to the number of maturing loans. But the fact that the Outstanding Balance is increasing in this entire period could only mean that the same loans are getting rolled over to new contracts, when balloon payments are missed during maturity. Figure 6c clearly shows that LTV (widely used in previous literature and used by most banks/asset managers for credit risk calculations) is flat throughout the data horizon. The interest rate is decreasing almost monotonically in the data and there seems to be no sensitivity of LTV to interest rate. Figure 6d corroborates that the NOI monotonically increases in the data and the occupancy is almost 100% in the entire data. So, there may be strategic saving of internal cash flow from income producing properties.

Table 1: The summary statistics for the cleaned data containing 9,617,333 observations of continuous variables is provided in this table. **M** stands for Million and **B** stands for Billion. "One hot encoding" technique converts categorical variables as binary vectors without any order. Because of the different scales of different variables and large skewness of several variables, they are later mapped to $[0,1]$ since the models output is a probability in $[0,1]$

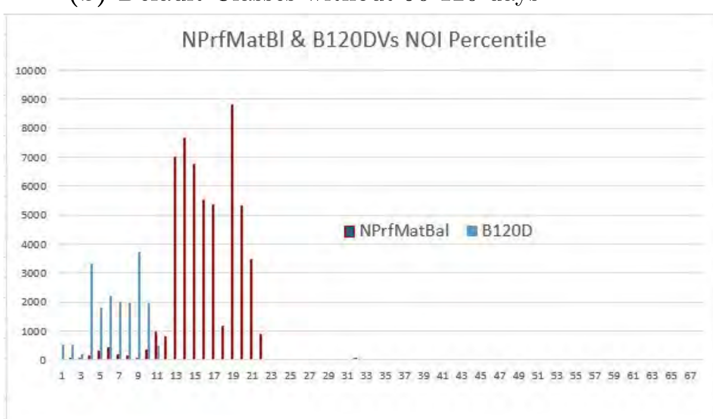
Statistic	N	Min	Pctl(25)	Median	Mean	Pctl(75)	Max
beginbal	9,617,333	0	1,838,146	4,036,697	7,882,727	9,100,000	99,990,043
orig_bal	9,617,333	0	1.92M	4.10M	9.31M	9.14M	1.68B
rate	9,617,333	0	6	6	6	7	9
Sched_princip	9,617,333	0	1,060	4,235	15,665	9,838	430M
Unsched_prin	9,617,333	0	0	0	50,300	0	1.50B
balance_act	9,617,333	0	1,774,511	3,980,981	7,797,502	9,009,824	99,999,000
payment	9,617,333	0	15,000	28,600	65,900	57,800	675M
pppenalties	9,617,333	0	0	0	420	0	29,477,125
liqproceeds	9,617,333	0	0	0	17,900	0	2.56B
realizedloss	9,617,333	0	0	0	5,260	0	204M
liqexpense	9,617,333	0	0	0	3,200	0	1.06B
numprop	9,617,333	1	1	1	1	1	225
Appraisal_Reduc	9,617,333	0	0	0	132,000	0	391M
SecurLTV	9,617,333	0	63	71	67	76	150
Face	9,617,333	0	6	6	6	7	9
NOI	9,617,333	0	268,000	527,000	1.22M	1.10M	1.09B
LTV	9,617,333	0	63	71	69	77	150
AppValue	9,617,333	1,620	3.46M	6.80M	17.4M	14.5M	48.1B
OccRate	9,617,333	0.0	0.89	0.96	0.92	1.0	1.4
Basis	9,617,333	0.0	2.0	2.0	1.7	2.0	4.0
Unemp	9,617,333	0.019	0.049	0.061	0.066	0.080	0.154
GDP	9,617,333	0	008	016	013	048	0.0174
2YrTr	9,617,333	02	06	0.010	0.019	0.031	0.061
10YrTr	9,617,333	0.015	0.024	0.036	0.034	0.043	0.063
NAREIT	9,617,333	-0.45	-0.05	0.01	-0.01	0.04	0.38
MIT.Liq	9,617,333	0.017	0.022	0.036	0.025	0.105	0.184
time_to_maturity	9,617,333	0	35	66	80	100	765
age	9,617,333	0	31	60	63	90	409
age2	9,617,333	0	961	3600	5393	8100	167281



(a) All Default Classes



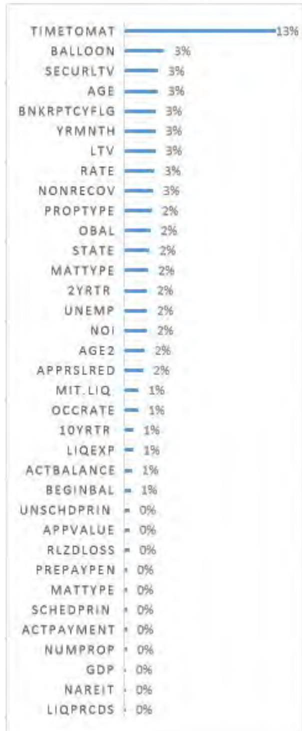
(b) Default Classes without 90-120 days



(c) Default Classes without 90-120 days & Performing

Figure 7: The delinquency states are Current/Performing class **W0_30D** which includes "loans with payments not received but still in grace period or not yet due", Late/Non-Performing classes **W30_60D**, **W60_90D** which includes loans with "Late Payment beyond 30-days but less than 60 days, beyond 60-days but less than 90-days, Default class **W90_120D** ((within 90 to 120 days of delinquency), Liquidation Proceedings & Final Resolution state **B120D** (beyond 120 days of delinquency), combined together as "limbo" loans. We try further states in the **PrfMatBal** (Performing, Mature and Balloon Payment due) and **NPrfMatBal** (Non-Performing, Mature and Balloon Payment due) classes to capture the incentives delay in resolution for foreclosed loans to REO/prepaid.

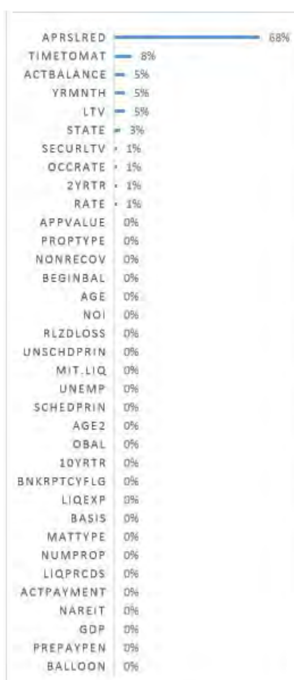
(a) Variable Importance: Lasso



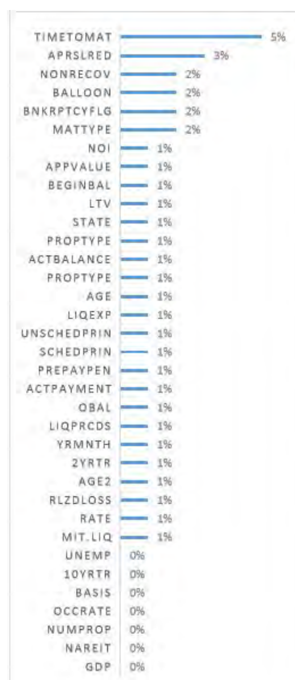
(b) Variable Importance: Ridge

(c) Variable Importance: Ordinal

(d) Variable Importance: DRF

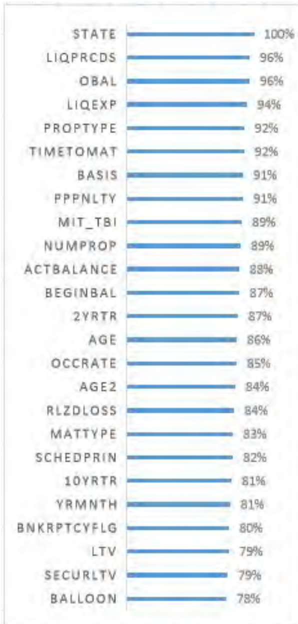


(e) Variable Importance: GBM

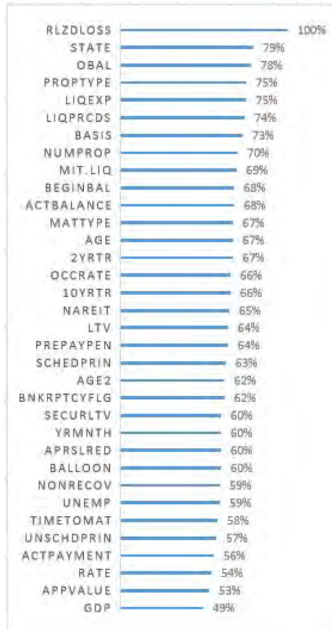


(f) Variable Importance: DNN

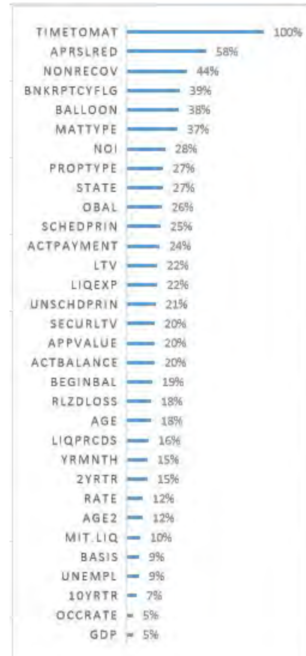
Figure 8: Variable Importance for 6 models, namely, Lasso, Ridge, Ordinal, Distributed Random Forest (DRF), Gradient Boosting Machine (GBM), Deep Neural Network (DNN). Only the variables having more than 0.5% marginal contribution are included in the variable importance visualizations for brevity.



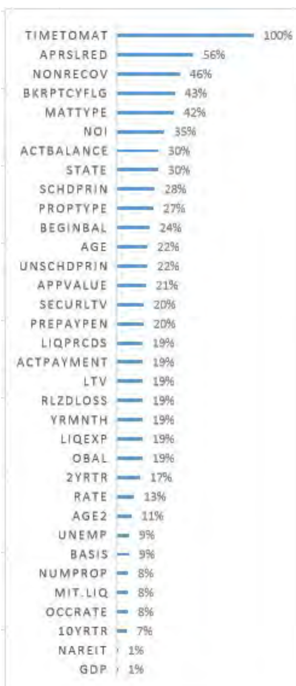
(a) VI without NOI



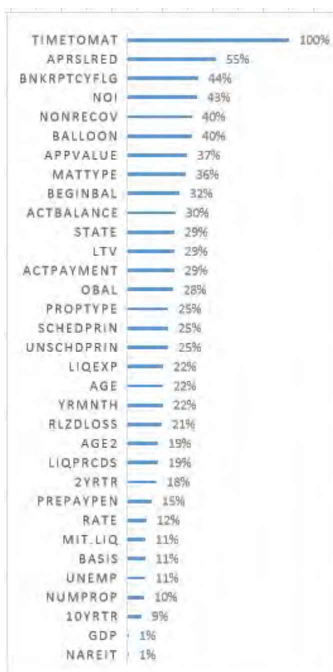
(b) VI without Year_Month



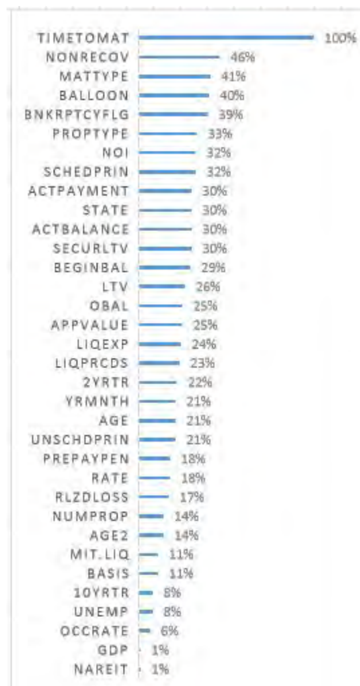
(c) VI without PrePayPen



(d) VI without Balloon Payment

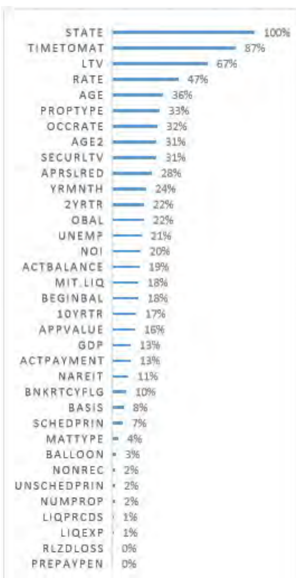


(e) VI without Occupancy

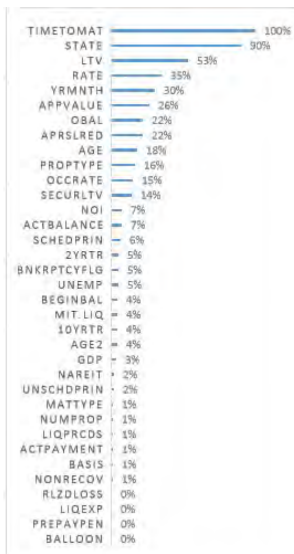


(f) VI without Appraisal Reduc

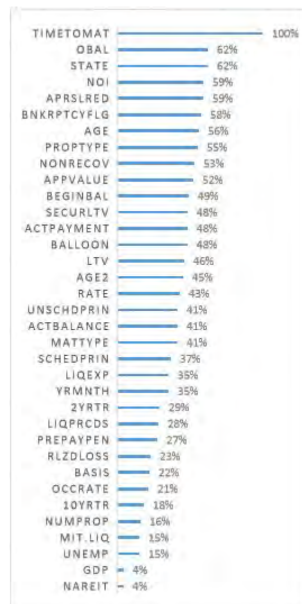
Figure 9: These are Variable Importance (VI) charts, leaving one out. We test the order of variable importance using Gedeon's method for DNN and exclude the key variables one at a time to test the strategic importance of NOI over LTV. NOI supercedes LTV in all of these VI charts, corroborating that NOI provides higher incentive than LTV during strategic default. Only the variables having more than 0.5% marginal contribution are included in the variable importance visualizations for brevity.



(a) VI in 2008: DRF



(b) VI in 2008: GBM



(c) VI in 2008: DNN

Figure 10: Variable Importance is stress-tested during Financial Crisis across all non-parametric models. The ranking order of variables is maintained almost exactly for DNN even in the subsample leading upto the financial crisis. The other models for DRF and GBM are very sensitive to this catastrophic shock in 2008. This corroborates that the Variable Importance from Gedeon's method is robust and provides the order of incentives across variables while borrower chooses to default strategically. Only the variables having more than 0.5% marginal contribution are included in the variable importance visualizations for brevity.

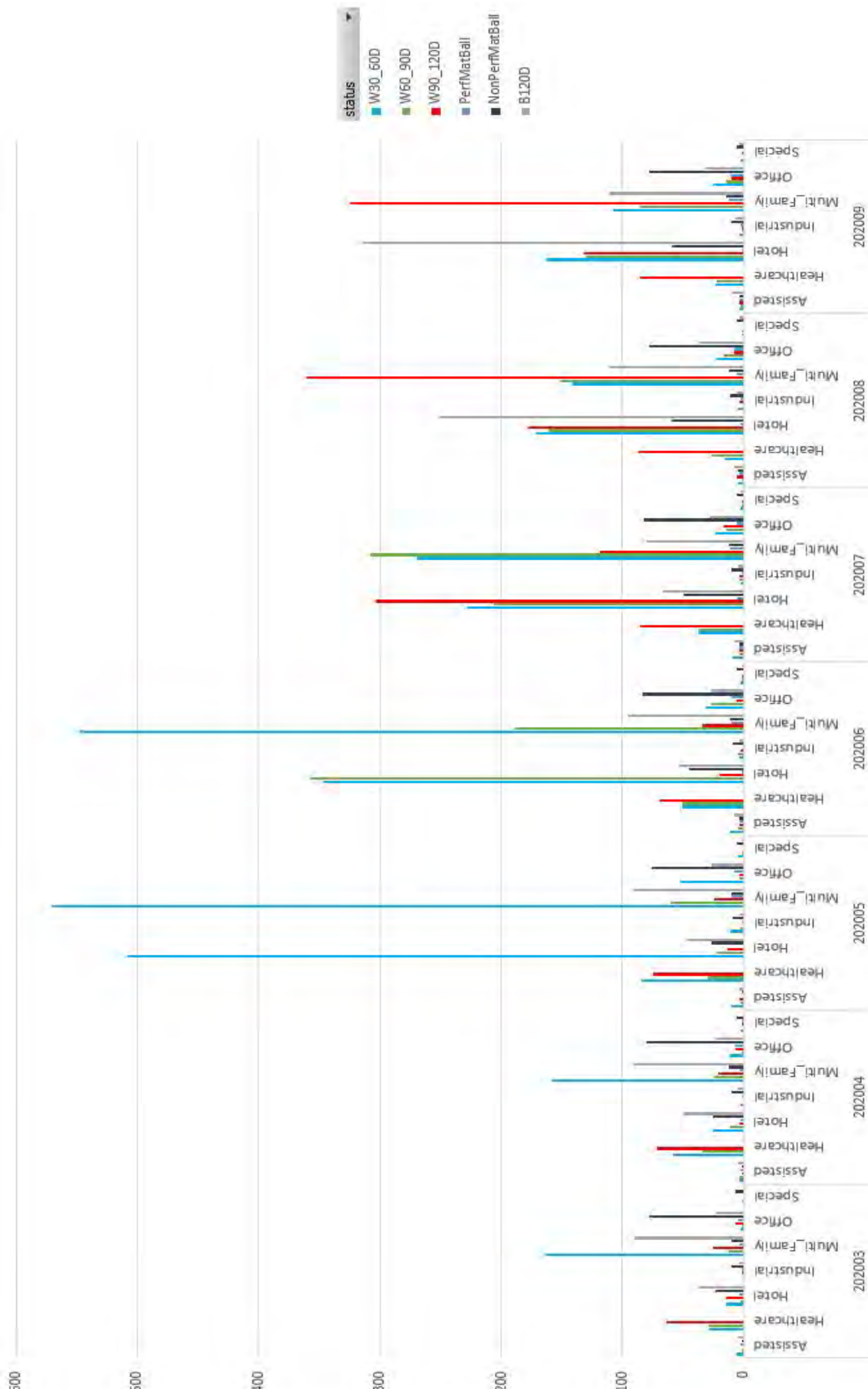


Figure 11: Impact of COVID-19 across industries

Table 2: Coefficients of Multinomial Logistic Regression: The marginal effect of features towards classifying the response set does not have a clear interpretation in terms of sensitivity and directionality. We list the co-efficients of Multinomial Logistic Model for the sake of completeness and only list the standard errors (SE) of the 3 statistically significant variables (Rate, LTV, Unemployment) that we use to demonstrate the counterintuitive interpretation in terms of default. (Table 2)

	names	W0_30D	W30_60D	W60_90D	W90_120D	PrfMatBal	NPfMtBl	B120D
1	Intercept	-24.83	-230.36	-223.45	-319.34	591.88	-170.20	272.02
2	State	-0	0.69	0.30	0.99	-2.12	-0.76	-0.08
52	PropType	-0.39	0.65	-0.12	0.61	0.55	0.62	0.49
60	MatType	0.23	-4.82	-2.87	-3.79	7.32	-57.90	-6.62
61	Balloon	0.33	-5.39	-3.48	-4.05	5.79	-53.02	-5.83
62	NonRecov	-1.59	-0.94	-0.75	0.69	-3.13	0.53	0.43
63	BnkrptcyFlg	-0.15	0.02	0.19	1.10	-0.78	-1.88	-0.09
64	BeginBal	-3.39	-2.13	-3.27	1.56	-0.67	-8.03	5.99
65	Obal	-1.18	21.62	19.06	-41.82	18.39	-7.16	-115.35
66	**Rate**	**-17.31*****	29.77	29.74	34	**-27.34*****	8.05	32.65
	SE	**0.21**	0.35	0.35	0.40	**0.33**	0.10	0.39
67	SchedPrin	134.71	153.33	54.63	89.44	156.85	173.07	41.50
68	Payment	44.44	-361.30	-6.64	-37.43	24.03	-1.95	57.25
69	UnschdPrin	45.46	-133.66	70.57	-47.70	57.08	-71.93	-172.29
70	PrePayPen	11.65	-253.05	-103.59	-61.27	-44.45	-370.52	16.91
71	ActBalance	2.25	0.94	2.73	-2.03	-0.28	8.07	-4.28
72	Liqprcds	-75.90	76.39	35.06	131.46	8.39	179.17	165.38
73	Liqexp	448.71	-1139.31	-2653.88	-518.47	-124.37	-1035.34	72.08
74	RlzdLoss	69.41	26.56	-48.44	-42.75	39.97	-43.45	-58.66
75	NumProp	-2.03	3.72	4.96	16.78	5.49	4.55	-23.82
76	AprslRed	-101.36	8.64	11.74	66.68	-20.31	49.31	87.47
77	SecurLTV	5.65	5.77	5.39	3.49	3.20	3.68	-4.19
78	NOI	159.73	-7.38	-39.41	27.03	124.60	-77.96	-149.62
79	**LTV**	**-6.55*****	-3.10	-2.39	0.03	-3.56	-2.21	0.77
	SE	**1.11**	0.53	0.41	0.01	0.61	0.38	0.13
80	AppValue	-2.27	3.02	4.01	7.02	-201.11	11.14	9.66
81	OccRate	0.42	-3.10	-3.55	-3.51	-1.27	-2.28	-2.84
82	Basis	-7.14	-2.13	-1.89	0.28	19.61	48.04	5.77
84	Age	-6.55	-5.95	-4.29	-4.77	-2.70	6.02	8.03
85	Age2	16.16	0.62	-2.87	-1.52	12.32	-10.49	-8.52
86	**Unemp**	**-1.93*****	17.51	21.77	18.87	25.90	21.55	**-12.23*****
	SE	**0.05**	0.43	0.53	0.46	0.64	0.53	**0.30**
87	GDP	-0.84	0.75	0.27	3.30	8.21	2.35	-6.66
88	2YrTr	4.89	-7.05	-9.52	-34.94	-5.31	-50.57	-3.23
89	10YrTr	10.94	18.14	18.82	37.29	-20.60	7.96	15.26
90	NAREIT	0.10	0.13	0.32	0.20	0.02	0.41	0.17
91	MIT.Liq	0.29	-0.12	-0.35	-0.31	0.27	-0.34	3.71
92	TimeToMat	4.69	4.63	4.43	4.57	-101.20	-140.81	2.91

Table 3: We report the Cross-Validation Training Errors of Models across Delinquency Classes with sample ranging from Jan 2000 - June 2012. The different models Ordinal, Mult (Multinomial Logistic), Lasso, Ridge, DRF (Distributed Random Forest), GBM (Gradient Boosting Machine), DNN (Deep Neural Network) and the different delinquency classes are W0_30D (Within 30 days), W30_60D (Within 30-60 days), W60_90D (Within 60-90 days), W90_120D (Within 90-120 days), PrfMatBal (Performing Mature Balloon), NPrfMatBal (Non-Performing Mature Balloon), B120D (Beyond 120 days). From W90_120D and worse delinquency classes, the misclassification error rates decrease with increasing model flexibility and DNN performs best overall.

	Ordinal	**Mult**	**Lasso**	**Ridge**	**DRF**	**GBM**	**DNN**
W0_30D	1%	1%	1%	1%	1%	1%	0%
W30_60D	**61%**	100%	100%	100%	**98%**	100%	100%
W60_90D	**63%**	100%	100%	100%	**98%**	100%	100%
W90_120D	36%	48%	48%	47%	**21%**	**27%**	**21%**
PrfMatBal	100%	100%	100%	100%	**76%**	**84%**	**78%**
NPrfMatBal	100%	67%	70%	74%	**16%**	**12%**	**12%**
B120D	100%	80%	87%	83%	**22%**	**21%**	**20%**
Totals	3%	2%	2%	2%	1%	1%	1%

Table 4: We report the Out-of-Sample Test Errors of Models across Delinquency Classes with sample ranging from Jul 2012 - June 2016. The different models Ordinal, Mult (Multinomial Logistic), Lasso. Ridge, DRF (Distributed Random Forest), GBM (Gradient Boosting Machine), DNN (Deep Neural Network) and the different delinquency classes are W0_30D (Within 30 days), W30_60D (Within 30-60 days), W60_90D (Within 60-90 days), W90_120D (Within 90-120 days), PrfMatBal (Performing Mature Balloon), NPrfMatBal (Non-Performing Mature Balloon), B120D (Beyond 120 days). From W90_120D and worse delinquency classes, the misclassification error rates (in decimals) decrease with increasing model flexibility and DNN performs best overall.

	Ordinal	**Multi**	**Lasso**	**Ridge**	**DRF**	**GBM**	**DNN**
W0_30D	1%	1%	1%	1%	1%	1%	1%
W30_60D	**89%**	100%	100%	100%	100%	100%	100%
W60_90D	**59%**	100%	100%	100%	100%	100%	100%
W90_120D	20%	37%	37%	37%	**14%**	**15%**	**17%**
PrfMatBal	98%	100%	100%	100%	**97%**	**96%**	**95%**
NPrfMatBal	100%	42%	42%	47%	**19%**	**13%**	**13%**
B120D	100%	100%	100%	100%	**7%**	**8%**	**7%**
Totals	98%	3%	3%	3%	3%	3%	3%

Table 5: We report the Variable Importance table for DNN (Deep Neural Network) across Delinquency Classes W0_30D (Within 30 days), W30_60D (Within 30-60 days), W60_90D (Within 60-90 days), W90_120D (Within 90-120 days), PrfMatBal (Performing Mature Balloon), NPrfMatBal (Non-Performing Mature Balloon), B120D (Beyond 120 days) with sample ranging from Jan 2017 - Feb 2020. NOI is higher than LTV, but accuracy reduces as the effect of COVID-19 was priced in by Feb 2020.

variable	relative_importance
state	1
appvalue	0.99
nonrecover	0.94
bankruptcyflag	0.90
securltv	0.90
proptype	0.89
2Yr_Tr	0.87
Unemployment	0.83
pmtbas	0.83
hasballoon	0.82
maturitytype	0.82
age	0.80
face	0.79
year_month	0.78
gdp	0.76
10Yr_Tr	0.73
actrate	0.73
noi	0.72
age2	0.70
occrate	0.68
securwac	0.68
actpmt	0.68
ltv	0.67
time_to_maturity	0.65

Table 6: We report the Variable Importance table for DNN (Deep Neural Network) across Delinquency Classes W0_30D (Within 30 days), W30_60D (Within 30-60 days), W60_90D (Within 60-90 days), W90_120D (Within 90-120 days), PrfMatBal (Performing Mature Balloon), NPrfMatBal (Non-Performing Mature Balloon), B120D (Beyond 120 days) with sample ranging from Jan 2017 - Nov 2019. NOI is higher than LTV and accuracy increases as the effect of COVID-19 was not priced in by Nov 2019.

variable	relative_importance
age2	1
age	0.93
state	0.90
time_to_maturity	0.86
proptype	0.78
face	0.68
maturitytype	0.64
nonrecover	0.60
hasballoon	0.60
actrate	0.60
securwac	0.59
actpmt	0.57
bankruptcyflag	0.56
pmtbas	0.56
appvalue	0.56
securltv	0.55
year_month	0.52
noi	0.51
2Yr_Tr	0.50
occrate	0.45
ltv	0.41
10Yr_Tr	0.39
Unemployment	0.27
gdp	0.07

Table 7: We report the out of sample Confusion Matrix for DNN (Deep Neural Network) across Delinquency Classes W0_30D (Within 30 days), W30_60D (Within 30-60 days), W60_90D (Within 60-90 days), W90_120D (Within 90-120 days), PrfMatBal (Performing Mature Balloon), NPrfMatBal (Non-Performing Mature Balloon), B120D (Beyond 120 days) with training sample ranging from Jan 2017 - Nov 2019 and test sample from Dec 2019 - Sep 2020. Accuracy becomes comparable to Table 4.

	W0_30D	**W30_60D**	**W60_90D**	**W90_120D**	**PrfMatBal**	**NPrfMatBal**	**B120D**	**Error**
W0_30D	565681	10	81	370	0	3588	1290	1%
W30_60D	3174	0	5	4	0	31	35	100%
W60_90D	1684	0	0	4	0	16	20	100%
W90_120D	1396	0	0	11	0	47	39	99%
PrfMatBal	493	0	0	4	0	148	6	100%
NPrfMatBal	999	0	0	136	0	1566	43	43%
B120D	2495	0	3	16	0	97	277	90%
Totals	575922	10	89	545	0	5493	1710	3%

Table 8: Out of sample Confusion Matrix Misclassification Error for Bankruptcy is close to 0% during COVID-19

	Non-BK	**BK**	Error
Non-BK	581697	13	0%
BK	22	2037	1%
Totals	581719	2050	0%

Table 9: Variable Importance for Bankruptcy using DNN still have NOI higher than LTV

variable	relative_importance
bankruptcyflag	1
state	0.64
nonrecover	0.56
prop	0.54
hasballoon	0.51
maturitytype	0.51
appvalue	0.49
age	0.44
2Yr_Tr	0.43
securltv	0.42
noi	0.40
actrate	0.40
year_month	0.39
face	0.39
10Yr_Tr	0.39
Unemployment	0.38
actpmt	0.38
pmtbas	0.38
gdp	0.37
securwac	0.36
occrate	0.35
age2	0.35
ltv	0.33
time_to_maturity	0.29

Table 10: Out of sample Confusion Matrix Misclassification Error for REO is close to 0% during COVID-19

	Non-REO	REO	Error
Non-REO	364993	1084	0%
REO	782	365042	0%
Totals	365775	366126	0%

Table 11: Variable Importance for REO using DNN no longer have NOI higher than LTV

variable	relative_importance
age	1
state	0.97
prop	0.89
age2	0.85
actpmt	0.79
ltv	0.76
securltv	0.75
securwac	0.73
hasballoon	0.72
time_to_maturity	0.72
bankruptcyflag	0.71
face	0.67
pmtbas	0.67
nonrecover	0.66
maturitytype	0.66
actrate	0.60
noi	0.59
appvalue	0.58
occrate	0.52
year_month	0.44
2Yr_Tr	0.40
Unemployment	0.36
10Yr_Tr	0.26
gdp	0.08

Table 12: Out of sample Confusion Matrix Misclassification Error for Foreclosure is close to 0% during COVID-19

	Non-FCL	FCL	Error
Non-FCL	364993	1084	0%
FCL	782	365042	0%
Totals	365775	366126	0%

Table 13: Variable Importance for Foreclosure using DNN no longer have NOI higher than LTV

variable	relative_importance
age	1
state	0.97
prop	0.89
age2	0.85
actpmt	0.79
ltv	0.76
securltv	0.75
securwac	0.73
hasballoon	0.72
time_to_maturity	0.72
bankruptcyflag	0.71
face	0.67
pmtbas	0.67
nonrecover	0.66
maturitytype	0.66
actrate	0.60
noi	0.59
appvalue	0.58
nonrecover.N	0.54
occrate	0.52
year_month	0.44
2Yr_Tr	0.40
Unemployment	0.36
10Yr_Tr	0.26
gdp	0.08

Table 14: Variable Importance for W90_120D loans using DNN no longer have NOI higher than LTV

variable	relative_importance
state	1
age	0.86
proptype	0.75
bankruptcyflag	0.75
actrate	0.75
securltv	0.74
hasballoon	0.71
ltv	0.71
actpmt	0.69
pmtbas	0.68
securwac	0.68
face	0.64
nonrecover	0.64
noi	0.62
age2	0.61
appvalue	0.61
2Yr_Tr	0.60
maturitytype.	0.59
time_to_maturity	0.58
occrate	0.54
year_month	0.48
10Yr_Tr	0.33
Unemployment	0.31
gdp	0.16

Table 15: Out of sample Confusion Matrix Misclassification Error for W90_120D is high during COVID-19

	Non-W90_120D	W90_120D)	Error
Non-W90_120D	577640	4636	1%
W90_120D	1365	128	91%
Totals	579005	4764	1%

Table 16: Variable Importance for B120D loans using DNN no longer have NOI higher than LTV

variable	relative_importance
age	1
state	0.97
prop	0.89
age2	0.85
actpmt	0.79
ltv	0.76
securltv	0.75
securwac	0.73
hasballoon	0.72
time_to_maturity	0.72
bankruptcyflag	0.71
face	0.67
pmtbas	0.67
nonrecover	0.66
maturitytype	0.66
actrate	0.60
noi	0.59
appvalue	0.58
occrate	0.52
year_month	0.44
2Yr_Tr	0.40
Unemployment	0.36
10Yr_Tr	0.26
gdp	0.08

Table 17: Misclassification Error is high for B120D loans

	Non-B120D	B120D	Error
Non-B120D	579326	1555	0%
B120D	2382	506	82%
Totals	581708	2061	1%

A Appendix

A.1 Loan to Value and Net Operating Income

We take a deeper dive and investigate LTV in the following way:

$$LTV_t = \frac{AOB_{t-1} + DS_t + B_T}{MV_t - AR_t} \tag{4}$$

where AOB_t is the Outstanding Balance at time t-1 that is amortized , DS_t is the scheduled payment due for servicing the debt obligation at time t, B_T is the Balloon Payment due at maturity, MV_t is the Market value of the property/properties at time t (which varies significantly with respect to macroeconomic conditions and spatial/location context) for which the mortgage has been issued, AR_t is the Appraisal Reduction at time t.

AOB remains consistent, since, prepaymant penalty clauses discourage voluntary curtailment/full prepayment. SP obligations are not met both when the borrower is cash-constrained and also when the borrower chooses to strategically default. Proximity to balloon payment at maturity further complicates the endogenous behavior of the commercial borrowers towards maturity of the loan. The market value of a property is a function of the macro-economic factors like state GDP, Unemployment Rate, geographical location, 2 Year and 10 Year Treasury Rates. Until valuation is obtained, Appraisal Reduction Amount (ARA) may be calculated based on the scheduled principal balance or some other formula as defined in the servicing agreement.

NOI calculation involves the following key variables. Potential Rental Income assumes zero vacancy or could be based on a rental market analysis. Vacancy losses are realized when tenants vacate the property and/or tenants default on their lease obligations. Total Operating Expenses on an Investment Property could include "Property Taxes, Rental Property Insurance, Property Management Fees, Maintenance and Repairs, Miscellaneous Expenses, etc. Debt service, depreciation, leasing commissions, tenant improvements, repairs to wear and tear, income taxes, and mortgage interest expenses are not included in the calculation of net operating income". This is because NOI is property-specific devoid of other investor or borrower-specific expenses. NOI helps calculate Cap Rate (property's potential rate of return), ROI, Debt Coverage Ratio, Cash Return on Investment. NOI provides an estimate a property's ongoing operating revenue. NOI analysis can be manipulated from the choice to accelerate or defer certain expenses. The NOI of a property can change depending on the property management. Because other expenses are not considered in NOI, the real cash flow from a property may differ net other expenses. Further projected rents cannot be used to calculate NOI when rents differ from market rents.

A.2 Multinomial Logistic Model

In a Multinomial Logistic Model, log-odds of each delinquency state with respect to the "Current" state assumes a linear specification. The odds that a loan has a delinquency classes j as opposed to the baseline, depending only on individual loan-specific covariates is defined as:

$$\frac{Pr(Y_i = j|Z_i = z)}{Pr(Y_i = 0|Z_i = z)} = exp(Z^{'}\gamma_j) \tag{5}$$

the choice Y_i takes on non-negative, un-ordered integer values $Y_i \in \{0, 1, ..., J\}$. Multinomial logistic regression does not assume normality, linearity, or homoskedasticity; it has a well-behaved likelihood function, a special case of conditional logit. A more powerful alternative to multinomial logistic regression is discriminant function analysis which requires these assumptions are met. Multinomial logistic regression also assumes non-perfect separation.

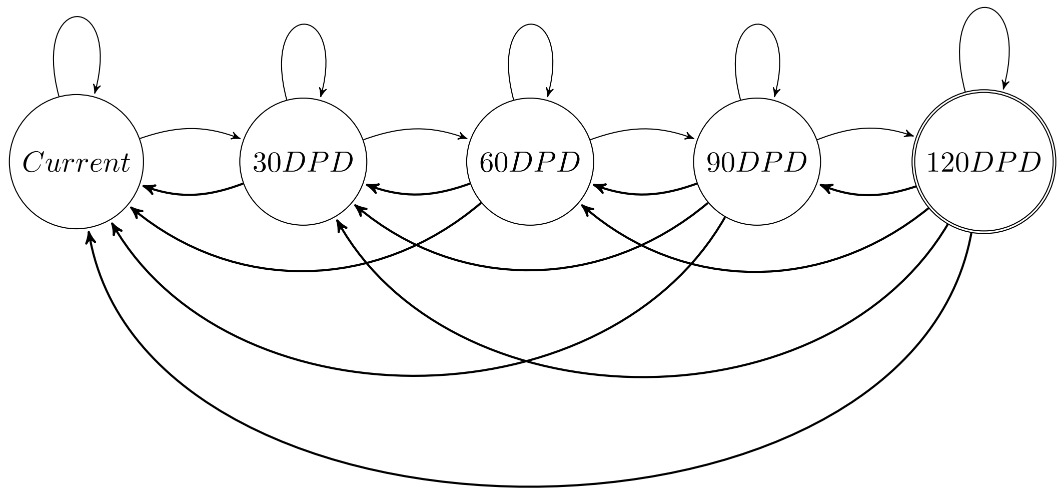
The Independence of Irrelevant Alternatives (IIA) assumption inherent in Multinomial Logistic Model implies that adding or deleting alternative outcome categories does not affect the odds among the remaining outcomes.

$$Pr(Y_i = j|Y_i \in \{j, l\}) = \frac{Pr(Y_i = j)}{Pr(Y_i = j) + Pr(Y_i = l)} = \frac{exp(X_{ij}^{'}\gamma)}{exp(X_{ij}^{'}\gamma) + exp(X_{il}^{'}\gamma)} \tag{6}$$

This can be tested by the Hausman-McFadden test. There are alternative modeling methods, such as alternative-specific multinomial probit model, or nested logit model to relax the IIA assumption.

A.2.1 Independence of Irrelevant Alternatives

Multinomial Logistic Model assumes Independence of Irrelevant Alternatives (IIA). The following Finite State Automaton details all possible transitions so that the above arguments can be visualized.



Clearly, the borrower would like to stick with the first choice, as the second choice classifies him/her in the default category and is detrimental for her creditworthiness from a lender's perspective. Now suppose, one more choice for being in **30 days to 60 days of delinquency** is given to the borrower, s/he may choose to rather be in this new state instead of less than 30 days of delinquency and may **strategically** miss one payment if there is a great investment opportunity for him/her in that one month horizon. In fact, none of the models (except Ordinal Bayes) can distinguish these three classes (**W0_30D**, **W30_60D** & **W60_90D**) and considers all of them as **Current Loans** in Table 4.

The borrower can undertake this decision as she/she is already some days in delinquency and she/she wouldn't mind going to the next bucket until she/she falls in the bucket for **90 days to 120 days of delinquency**. In this situation, the borrower's creditworthiness doesn't change that much from a lender's perspective. hence, the odds for being in the **”less than 30 days delinquency”** to being in the classes of **90 days to 120 days of delinquency** will change drastically in the presence of this new choice of being in **30 days to 60 days of delinquency**. hence the IIA assumption is clearly violated.

Also the marginal effect of features towards classifying the response set does not have a clear interpretation in terms of sensitivity and directionality. We list the co-efficients of Multinomial Logistic Model for the sake of completeness. (Table 2)

A.3 Distributed Random Forest

Recursive partitioning, a critical data mining tool, shelps in exploring the stucture of a data set. This section provides a brief overview of CART modeling, conditional inference trees, and random forests.

Random Forests are developed by aggregating decision trees and can be used for both classification and regression. Each tree is a weak learner created from bootstrapping from subset of rows and columns. More trees will reduce the variance.It alleviates the issue of overfitting, can handle a large number of features. It shelps with feature selection based on importance. It is user-friendly with two parameters: number of trees (default 500) and variables randomly selected as candidates at each split, $\sqrt{ntree}$ for classification and $ntree/3$ for regression. "Out Of Bag Error" is estimated for each bootstrap iteration and related tree.

R's randomForest splits based on the Gini criterion and H2O trees are split based on reduction in Squared Error (even for classification). H2O also uses histograms for splitting and can handle splitting on categorical variables without dummy (or one-hot) encoding. Also, R's randomForest builds really deep trees, resulting in pure leaf nodes, leading to constant increments in prediction and ties and hence relatively lower AUC.The trees in H2O's random forest aren't quite as deep and therefore aren't as pure, allowing for predictions that have some more granularity to them and that can be better sorted for a better AUC score.

CART models an outcome y_i for an instance i as:

$$y_i = f(x_i) = \sum_{m=1}^{M} c_m I x_i \in R_m \tag{7}$$

where each observation x_i belongs to exactly one subset R_m, c_m is the mean of all training observations in R_m.

A.4 Gradient Boosting Machine

GBM Friedman (2000) creates an ensemble Kuncheva (2003) of weak prediction models in stages and utilizes a differentiable loss function. Boosting trees does increase accuracy, but at the cost of speed and meaningful interpretability.

At each step m, $1 \leq m \leq M$ of gradient boosting, an estimator h_m is computed from the residuals of the previous model predictions. Friedman (2001) proposed regularization by shrinkage:

$$F_m(x) = F_{m-1}(x) + \nu\gamma_m h_m(x) \tag{8}$$

where $h_m(x)$ represents a weak learner of fixed depth, γ_m is the step length and ν is the learning rate or the shrinkage factor. XGBoost Chen and Guestrin (2016) is a faster and more accurate implementation of the Gradient Boosting algorithm Chen, Lundberg, and Lee (2018).

A.5 Deep Neural Network

A.5.1 Deep Neural Network for CMBS

The purpose for a Deep Neural Network (DNN) is bourne out of the need to have transparency and accountability Albanesi and Vamossy (2019). By the very nature of the DNN, we do not have to add

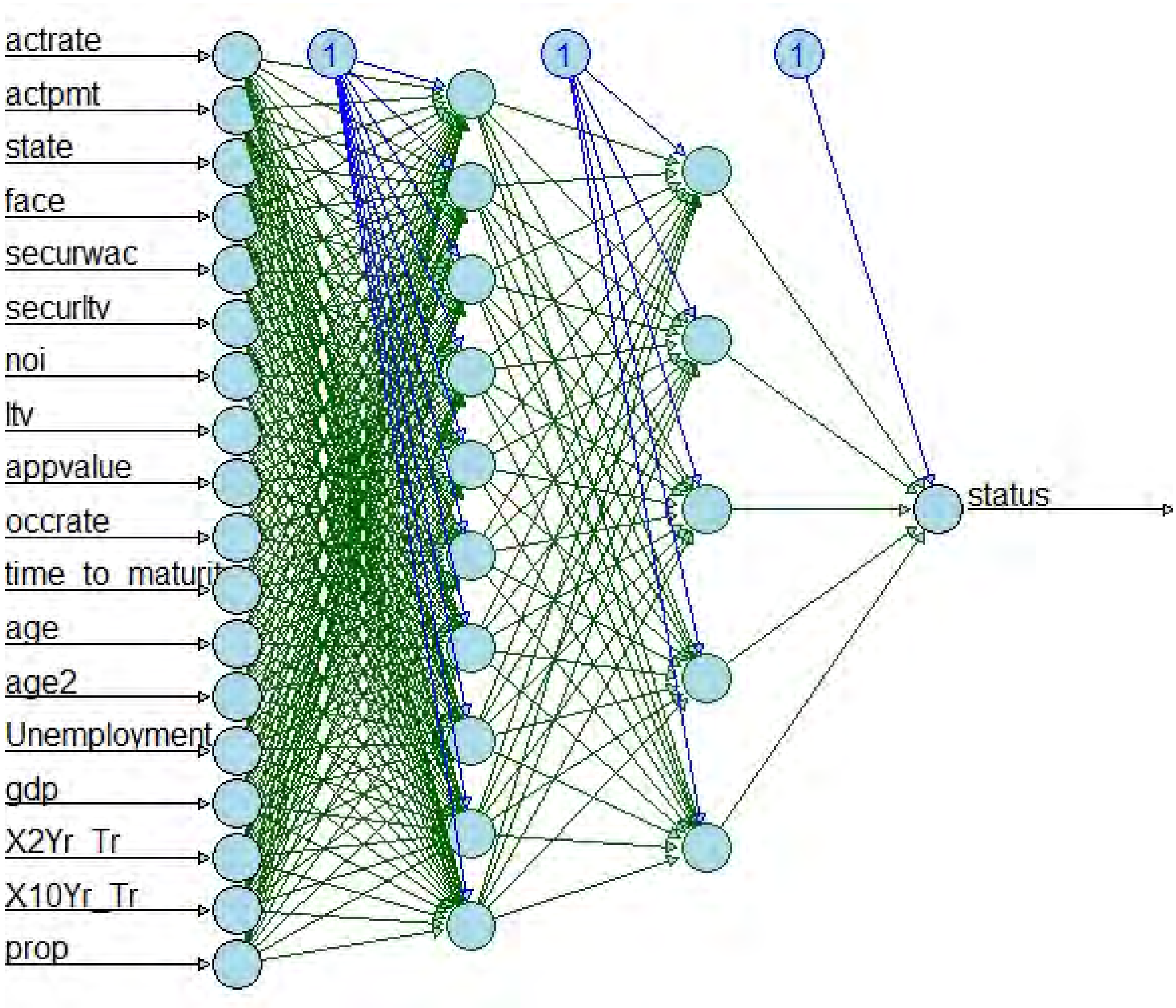


Figure 12: A few variables out of 33 are droppped in this visualization due to several missing values. However, they have been included in the actual implementation of the Feedforward DNN from h2o package. For visualization, the package neuralnet throws an error for variables with several missing values. But this diagram clearly shows how DNN works with hidden layers and how variables have weights during linear combination in each layer and then fed forward using non-linear activation functions and finally to the outcome variable during classification.

interaction terms in the specification of the model, especially in the case of high dimensional data. The sequential layers embody highly non-linear and non-trivial interaction among the variables and capture several latent fundamental features in the process. The causal interpretation of the covariates both in default Kvamme, Sellereite, Aas, and Sjursen (2018) and prepayment calculations have not been explored in details. The broader impact could be traced out by improved allocation of credit and aid in policy design (macroprudential, bankruptcy, foreclosure, etc.).

With the provision of enough hidden units, a neural network can mimic continuous functions on closed and bounded sets really well Hornik (1991), vis-a-vis the product and division of relevant features and their interactions. More layers, and not more units in each layer, learns atures of greater complexity. Deep neural networks 12, with three or more hidden layers, require exponentially fewer units than shallow networks or logistic regressions with basis functions; see Montufar, Pascanu, Cho, and Bengio (2014) and Goodfellow, Bengio, and Courville (2016).

Deep Neural Networks (DNN) **??** have an extensive set of current applications like: System identification and control (e.g., vehicle control, trajectory prediction, etc.), Game-playing and decision making (e.g., chess, poker, etc.), Pattern and sequence recognition (e.g., radar systems, face identification, signal classification, speech/image recognition, etc.), Medial diagnosis and finance (e.g., automated trading systems, cancer diagnosis, etc.).

A.6 Hyperparameter Tuning and Grid Search

Hyper-parameter tuning with Random Grid Search (RGS) tests different combinations of hyperparameters to find the optimal choice based on accuracy, without overfitting.

Hyperparameters can be divided into 2 categories:

- **Optimizer hyperparameters**
- **Model Specific hyperparameters**

Our model hyperparameters are: *score_training_samples* = 6125796, *epoch* = 60000, *hidden* = c(30,20,10,7), *hidden_dropout_ratios* = c(0.01, 0.01, 0.01,0.01), *momentum_start* = 0.5, *momentum_ramp* = 100, *momentum_stable* = 0.99, *missing_values_handling* = "Skip", *initial_weight_distribution* ="Uniform", *nesterov_accelerated_gradient* = TRUE, *activation* = "RectifierWithDropout", *nfolds* = 10, *fold_assignment* = "Stratified", *keep_cross_validation_predictions* = FALSE, *variable_importances* = TRUE, *adaptive_rate* = FALSE, *l1* =1e-5, *l2* =1e-5, *export_weights_and_biases* = FALSE, *mini_batch_size* = 128, *loss* = "CrossEntropy", *distribution* = "AUTO", *balance_classes* = T, *max_after_balance_size* = 1, *rate* = 05, *rate_annealing* = 1e-06, *rate_decay* = 1, *stopping_metric* = "MSE", *seed* = 1122.

A.7 Class Imbalance Problem

Most classifiers are unable to distinguish minor classes Kuncheva (2003) and are sheavily influenced by major classes, e.g., the conditional probability of minor classes are underestimated in a logistic regression King and Zeng (2001), Tree based classifiers, and KNN yield high recall but low sensitivity when the data set is extremely unbalanced Daelemans, Goethals, and Morik (2008). There are a plethora of techniques to balance the data, e.g., oversampling, under-sampling and Synthetic Minority Oversampling Technique (SMOTE) proposed by Chawla, Bowyer, Hall, and Kegelmeyer (2002).