Predicting Consumer Default with Deep Learning*

Stefania Albanesi, University of Pittsburgh, NBER and CEPR
Domonkos F. Vamossy, University of Pittsburgh

April 5, 2019

Summary

Our project develops a model to predict consumer default based on deep learning. We show that the model consistently outperforms standard credit scoring models, even though it uses the same data. Our model is interpretable and is able to provide a score to a larger class of borrowers relative to standard credit scoring models. Additionally, it accurately tracks variations in systemic risk and is able to distinguish between strategic and distressed default. We argue that these properties can provide valuable insights for the design of policies targeted to reducing consumer default and alleviating its burden on borrowers and lenders, as well as macroprudential regulation. We also deploy recent methodological advances that enable to impose fairness constraints on our model to examine how such constrained learning can reduce disparities in credit access for historically disadvantaged socio-economic groups. We contribute to the literature on ethical machine learning in computer science by developing economic rather than algorithmic measures of fairness and by considering general equilibrium feedback loops that are typically ignored by that literature.

*Proposal prepared for NBER Call for Proposals: Big Data and High-Performance Computing for Financial Economics. Slides presenting updated results on this project are available here. Correspondence to: stefania.albanesi@gmail.com.

Project Description

The dramatic growth in household borrowing since the early 1980s has increased the impact of consumer default on household and financial sector balances sheets. Total consumer credit balances in millions of 2018 USD have grown from just under 1.5 billions in 1985 to close to 4 billions in 2017, according to Federal Reserve Board data. The delinquency rate on consumer loans has fluctuated between 2 and 5% over this period, but with the growth of consumer credit, variations in the delinquency rate have an ever larger impact on household and financial sector balances sheets. Credit scoring models are proprietary algorithms designed to rank borrowers based on their probability of future default using information on their credit report and constitute one of the most important factors in the allocation of consumer credit in the United States. Approximately 200 million consumers have a credit score, which is also used to screen job applicants and for insurance and rental agreements. Credit scores can influence whether a consumer is approved for credit and they are also crucial in determining interest rates on loans. Despite their ubiquitous use in the financial industry, there is very little information on these models, and there is emerging evidence suggesting that credit scores as currently formulated have severe limitations.

This project proposes a novel approach to predicting consumer default based on deep learning using the same data as standard credit scoring models. We rely on deep learning as this methodology is specifically designed for prediction in environments with high dimensional data and non-linear high order interactions among factors affecting the outcome of interest, for which standard regression approaches perform poorly.

The first part of our research is devoted to developing a deep learning model of consumer default (see Albanesi and Vamossy (2019)). We focus on predicting 90+ day past due delinquencies on any consumer loan, the outcome targeted by existing credit scoring models. Our model substantially improves the accuracy of default predictions over conventional credit scores and is able to score most consumers. Using the most recent interpretability techniques, we can isolate the most important factors driving default for individuals and classes of borrowers, which increases transparency and accountability. We also show that our deep learning model displays a finer variation in predicted default probabilities for different current delinquency states relative to conventional credit scores, and substantially higher true positive rate and lower false positive rate.

Our approach can also predict variations in aggregate default risk and we show that the model can capture both the magnitude and the timing of the rise in consumer defaults leading up to the housing crisis and recession in 2007-2009, as well as the subsequent decline in defaults. Capturing such sharp inversion points is very difficult using standard econometric techniques, and particularly notable as we do not use any aggregate variables for our prediction, so the variation is completely driven by changes in individual borrowers' debt positions. Most professional and academic economists failed to anticipate the rise in consumer defaults that was a key determinant of the housing and mortgage crisis, whereas our model would have been able to predict this phenomenon with a high level of accuracy. Finally, we use our model to predict strategic and distressed default. A default is considered strategic when the borrower could have continued to make payments on their loans and distressed otherwise. From a policy standpoint, it is important to distinguish between these modalities of default, as strategic default may be driven by moral hazard and should be discouraged, while distressed default may be deserving of alleviation via loan modification or discharge through bankruptcy. We show that our model attributes much higher probability of default to borrowers who default strategically compared to conventional scoring models.

Our results can help to enhance our understanding of the determinants of consumer default as well as to predict its variation over time and across types of consumers. This can not only improve the allocation of credit, but also lead to important insights for the design of policies aimed at preventing consumer default or alleviating its effects on borrowers and lenders. They are also critical for macroprudential regulation, as they can assist with the assessment of the impact of consumer credit on the fragility of the financial system.

The second part of our research is still in progress and addresses fairness and transparency in credit scoring. Current legislation in the United States purports to prevent discrimination against protected categories by mandating that consumer lending and credit scoring models be blind with respect to those categories. Protected categories include age, race, gender, religion, and sexual orientation. However, these protections are inadequate since credit scoring models are based on outcomes that are likely influenced by historical

discrimination. This is easy to see when it comes to race. A recent report by the Brookings Institution (see Matthew, Rodrigue, and Reeves (2016)) showed that a house in a black neighborhood will be valued $48,000 less on average than the same house in a white neighborhood, amounting to $156 billion in cumulative losses for black homeowners. The impact of discrimination does not end at home values. Lower home values imply that black borrowers have smaller mortgages, home equity lines of credit and may consequently need to use credit cards more to borrow compared to whites. All these factors lower credit scores, which reduces black borrowers' access to other credit and increases the interest rates on their loans. Higher cost of borrowing increases the probability of default, other things equal, producing a low credit score-high default risk spiral, which cannot be prevented by simply prohibiting the explicit use of race in loan terms or credit scoring models. We plan to use recent advances in constrained learning (see Cotter et al. (2018), Cotter et al. (2019) and Gupta et al. (2018)) to guarantee fairness in credit allocation outcomes based on default predictions implied by our model. Our research will provide both a substantive and a methodological contribution to the literature on ethical machine learning developed in computer science. The substantive contribution will be to adopt notions of fairness that are rooted in the social choice tradition in economics, to ameliorate the pitfalls associated with the algorithmic notions typically used in computer science (see Corbett-Davies and Goel (2018)). From a methodological standpoint, we will incorporate feedback loops from our model to credit allocation outcomes that arise from equilibrium mechanism, such as adjustment of threshold rules for credit approval to changes in the distribution of default predictions under fairness constraints. The feedback mechanisms can alter the assessment of the impact of fairness constraints and are typically ignored in the computer science literature.

Data We use anonymized credit file data from the Experian credit bureau. The data is quarterly, it starts in 2004Q1 and ends in 2015Q4. The data comprises over 200 variables for an anonymized panel of 1 million households. The panel is nationally representative, constructed from a random draw for the universe of borrowers with an Experian credit report. The attributes available comprise information on credit cards, bank cards, other revolving credit, auto loans, installment loans, business loans, first and second mortgages, home equity lines of credit, student loans and collections. There is information on the number of trades for each type of loan, the outstanding balance and available credit, the monthly payment, and whether any of the accounts are delinquent, specifically 30, 60, 90, 180 days past due, derogatory or charged off. All balances are adjusted for joint accounts to avoid double counting. For mortgages, we also know when the mortgage was originated, and whether the borrower has Fannie Mae, Freddie Mac, FHA or VA mortgages. Additionally, we have the number of hard inquiries by type of product, and public record items, such as bankruptcy by chapter, foreclosure and liens and court judgments. For each quarter in the sample, we also have each borrowers's score, specifically their Vantage Score 3. The data also includes an estimate of individual and household labor income based on IRS data. Because this is data drawn from credit reports, we do not know gender, marital status or any other demographic characteristic, though we do know a borrower's address at the zip code level. We also do not have any information on asset holdings. To the best of our knowledge, this is the largest repository of credit registry data for the U.S. held by private researchers.

Computational Requirements We consider a variety of unconstrained and constrained machine learning techniques, from logistic regression, to random forests, extreme gradient boosting and deep neural networks. Our final model at this stage consists in a 7 hidden layers, fully connected deep neural network with 50% dropout rate, for which the number of layers and units and the activation function optimized via hyperparameter tuning. We employ over 150 unique features for each borrowers, and train up to 20 million samples over about 10 GB of source data, using single precision floating point operations to reduce computational burden. We run our computation on the University of Pittsburgh Center for Research Computing high performance computing cluster, using 1-2 GPU nodes with 4 NVIDIA GeForce GTX1080 GPUs, which is over 40X faster than CPU nodes. Our code is written in Python 3.6.3 (using the numpy, pandas, matplotlib packages) and our deep neural networks are built with keras running on Google TensorFlow. The computational burden is very intensive owing to the fact that we consider a large number of machine learning models for comparison purposes, the large data set we rely on, and the fact that we conduct an extensive robustness analysis.

References

Albanesi, Stefania, and Domonkos Vamossy. 2019. "Predicting Consumer Default: A Deep Learning Approach." *Manuscript, University of Pittsburgh*.

Corbett-Davies, Sam, and Sharad Goel. 2018. "The measure and mismeasure of fairness: A critical review of fair machine learning." *arXiv preprint arXiv:1808.00023*.

Cotter, Andrew, Maya Gupta, Heinrich Jiang, Nathan Srebro, Karthik Sridharan, Serena Wang, Blake Woodworth, and Seungil You. 2019. "Training Fairness-Constrained Classifiers to Generalize."

Cotter, Andrew, Heinrich Jiang, Serena Wang, Taman Narayan, Maya Gupta, Seungil You, and Karthik Sridharan. 2018. "Optimization with non-differentiable constraints with applications to fairness, recall, churn, and other goals." *arXiv preprint arXiv:1809.04198*.

Gupta, Maya, Andrew Cotter, Mahdi Milani Fard, and Serena Wang. 2018. "Proxy Fairness." *arXiv preprint arXiv:1806.11212*.

Matthew, Dayna Bowen, Edward Rodrigue, and Richard V. Reeves. 2016. "Time for justice: Tackling race inequalities in health and housing." *Brookings Report*.