Predicting and Modeling Income Stability in Credit Decisions[1]

AUSTIN LEOPOLD[2], CHRIS PENTO[3], RAFIQ SITABKHAN[4]

[1] Lehigh University, Bethlehem, PA 18015, USA. Course listing: "FinTech Capstone"; FIN 388. Advisors: Donald Bowen, Lehigh University; Aneesh Varma, Aire. Last revised: May 4, 2020.

Acknowledgments: The authors would like to thank Aneesh Varma for providing insight and guidance throughout our research on this topic.

[2] Finance, College of Business, Lehigh University. Email: ajl321@lehigh.edu and austin.leopold50@gmail.com
[3] Finance, College of Business, Lehigh University. Email: cnp221@lehigh.edu and cpento2@gmail.com
[4] Finance, College of Business, Lehigh University. Email: rss321@lehigh.edu and rsitabkhan24@gmail.com

Table of Contents

1. Executive Summary

New regulations and lending guidelines are prompting lenders to evaluate an applicant's "ability to pay." Existing ability to pay assessments are costly, as they are typically ad-hoc and labor intensive. Thus, a scalable solution has the potential to both reduce costs and improve credit models significantly for lenders. Our team was tasked by Aire, a company that provides lenders with credit insights about individual applicants, to model applicant income stability, a key ingredient in their ability to pay model.

- Credit lending shifted from using *soft* information to *hard* information during the 20th century, accelerated by the explosion in computing power and data.
- The extreme growth in the amount of available data has prompted lenders to use artificial intelligence and machine learning to better understand applicant risk profiles.
- New regulations are prompting lenders to look at "ability to pay," which is an assessment of how an applicant's income can support their personal expenses and debt burden.
- Personal expenses are stable and relatively easy to predict, and thus an assessment of "ability to pay" hinges on an estimate of future income stability.
- We propose an income stability model that considers possible changes in employment status, reductions in earnings, and, in the event of a job loss, how labor market conditions impact the likelihood of getting a replacement job at or near their current income level.
- On implementation: Aire, a FinTech startup, can use the model within the current user interface, which protects existing customer throughput and revenue generation.
- We discuss how such a model fits into the existing and changing regulatory framework, and avenues for improving the model.

2. Introduction

Regulatory developments and increasing demand from applicants impacts the credit lending industry yearly. Lenders must adapt to these changes, while balancing profitability and risk management. The key avenue to address these challenges has been the introduction of financial technology to leverage “Big Data” during the credit decision process.

Traditionally, credit decisions hinged on an analysis of an applicant's historical record to assess an individual’s financial well-being. However, credit is a forward-looking agreement and therefore, lenders who are able to incorporate forward-looking projections about an applicant’s ability to pay debt burdens will be able to price the risk of such loans more accurately. Existing models for creating forward looking projections are costly, as they are typically ad-hoc and labor intensive. Thus, a scalable solution has the potential to both reduce costs and improve credit models significantly for lenders, in addition to helping applicants that are overlooked by existing credit models.

For several months, our group has researched regulations in the credit industry, the challenges at hand for credit rating companies, and current solutions to solving the forward-looking aspect of credit. We propose an initial income stability model that is a function of location and job-specific labor market conditions. Our proposed product can be computed over various forecasting horizons and income stability thresholds. This model can be estimated with information currently collected in the Aire UI, and we discuss both extensions to the model and how it fits within the changing regulatory framework for credit decisions.

3. The Regulatory Environment for Credit Decisions

The Credit Decision Process

The credit decision process originates with the customer applying to a lender for credit, after which the lender begins to assess the applicant for their credit worthiness. Creditworthiness is an assessment of the risk of an applicant, which is characterized by asymmetric information: The applicant knows more about their risk than the lender.

The historical solution to this process involved loan officers at local branches. They would physically meet with the applicant and acquire *soft* information, such as personal characteristics and character. This process is associated with large costs (physical branch locations and the labor of the loan officers) and the dominant model has shifted away from in-person lending decisions based on soft information towards using *hard* information, such as a credit score.[5] Hard information is quantitative, easy to store and transmit in impersonal ways, and its information content is independent of its collection (Liberti and Petersen).

Implications of Paradigm Shift

The switch from *soft* information, involving a personal relationship with a banking officer, to *hard* information has important implications for the structure of the lending industry. First, the use of hard information generates the potential for tremendous economies of scale relative to the classical lending model of local loan officers. This impacts the structure of the lending market, because it "transformed banking from an exclusively local and personal market to a nation, competitive, and in some cases impersonal market" (Liberti and Petersen). In this impersonal market, lenders must establish ways to scale their technology for large scale usage with a plethora of candidates from different backgrounds applying. In order to build models that perform well across a national market covering a heterogeneous pool of applicants, the acquisition of data becomes paramount. The value of data is heightened in the new paradigm because of the ability of machine learning to exploit enormous datasets.

[5] Credit scores are provided by three main players (Equifax, Experian, and Transunion).

As the industry has shifted towards hard information, there has been a concurrent explosion in the amount of data at their disposal. These two trends have prompted the industry to leverage artificial intelligence methods and machine learning models in order to better understand the risk profiles of applicants.

However, the use of machine learning alongside big data can lead to models that run afoul of regulatory protections. AI methods are often referred to as "black box" models. In order to prevent such models from exploiting information about applicants that fall under regulatory protections (e.g. age or race), lenders remove such characteristics from datasets that are used in credit models. However, this might not be sufficient to ensure that AI models do not discriminate against protected classes. For example, models with enough information can uncover the race of an applicant and act "as though" they know. This, in turn, can lead AI models to generate credit assessments discriminatorily. As such, lenders and financial technology companies, such as Aire, need to remain cognizant of the potential biases black models might create (Klein).

Key Regulations Impacting the Space

The CARD Act has changed the credit lending decision for many lenders due to the complex rules and regulation within the Act itself. The Act seeks to protect consumers from unfair practices in the credit card industry, by preventing lenders from targeting vulnerable consumers or making decisions with the knowledge that the person will not be able to make such repayments.[6]

In the CARD Act of 2009, there is a provision known as the "ability to pay", which requires card issuers to consider a consumer's ability to make the required payments on the account (U.S. G.P.O.). This standard applies to both increases in an existing credit line or new cards. There have been subsequent modifications to the CARD Act, most notably in section 226.51 of the Truth in Lending Act. Section 226.51 contains requirements for issuers to "consider repayment ability

[6]By prohibiting double cycle billing, interest rate increases at will, and unfair fees, while also helping consumers manage their accounts and offering special protections for students and young people, this act also opens the door to the "ability to pay" provision.

based on the consumer's income, assets, and current obligations (U.S. G.P.O.). Issuers must establish reasonable policies and methods for considering one of the following: the ratio of debt obligations to income; the ratio of debt obligations to assets; or the income the consumer will have left after paying debt obligations" (Consumer Action).

In 2012, Regulation Z was amended to allow applicants more flexibility in reporting income and overall net worth, in order to maintain a better chance to qualify for credit. Under this new amendment, credit card issuers can consider assets or income that an applicant has as a "reasonable expectation of access" (Mishkin). This means that an individual can, for example, report the income of a non-applicant, such as a spouse, if they share a joint account. This once again plays into the review of income and an applicant's ability to pay, with new regulations ensuring that lenders are getting the most accurate information and representative look into the current economic situation of an applicant.

The Equal Credit Opportunity Act (ECOA), which was implemented by Regulation B, applies to all creditors. Within this act, the statute makes it "unlawful for any creditor to discriminate against any applicant with respect to any aspect of a credit transaction (1) on the basis of race, color, religion, national origin, sex or marital status, or age (provided the applicant has the capacity to contract); (2) because all or part of the applicant's income derives from any public assistance program; or (3) because the applicant has in good faith exercised any right under the Consumer Credit Protection Act" (CFPB). While this act seems to focus on the disparate treatment of applicants, it also covers disparate impact.[7] As a result, lenders must have strong procedural safeguards in order to not only protect themselves but also the consumer (from models that disparately impact protected classes).[8]

[7] Disparate impact occurs when neutrally designed policies, practices, or rules unintentionally and disproportionately affect a member of a protected group of people (CFPB).
[8] Models with inputs correlated to protected classes of people are legally permissible in models as long as the business need is justified. Such as a price model using zip codes.

Impact On Industry and Aire Compliance

As the rules in these regulations promulgate and become effective, lenders will need to adjust. Traditional credit analysis will not be completely sufficient, with new rules requiring them to make assessments about customers' future ability to pay. One option for lenders is to partner with financial technology firms that specialize in acquiring and using data to make the necessary assessments.

This space is where Aire is focusing, to "deliver fully auditable credit insight for lenders, built specifically to comply with FCRA, ECOA and other consumer protection laws" (Aire). While the lenders and industry itself are still using a patchwork of bespoke manual solutions in order to predict ability to pay, the idea behind Aire's main product, *Pulse,* is to create modules (a set of statistics and estimates) that give better insights into the future outlook of an applicant. These insights can then be built into the lender's credit decision process in order to both improve lending decisions and improve their regulatory compliance systems.

Following the 2008 financial crisis, new regulations in the credit lending industry were implemented to ensure that applicants had the financial health to meet debt burdens before they could be granted a loan. The ultimate goal of these new regulations was to protect consumers from predatory lending and unfair lending practices. By using Pulse for credit insights, lenders can ensure that they will be fully compliant with regulatory guidelines in the credit lending industry while also receiving a complete evaluation of a consumer's ability to pay.

Regulatory compliance can be satisfied through Pulse because Aire's model is designed to be fully traceable. Since we know that the Fair Credit Reporting Act (FCRA) requires that rejected applicants receive clear information as to why they have been rejected, the traceability of Pulse will allow lenders to accurately portray these reasons to consumers. Thus, lenders who use Pulse will be in compliance with FCRA regulations. Additionally, since we know that the ECOA prevents discrimmination against protected classes, the elimination of bias in Pulse's model can ensure compliance with ECOA regulations. Therefore, Pulse can not only provide a lender with a better understanding of the complete picture of an applicant's ability to pay, but it can also ensure compliance with key regulations in the credit lending industry.

4. Value Proposition

Measuring Ability to Pay

There are four main factors that play a role in the credit decision for a lender: past credit behavior, ability to pay, identity, and fraud risk (CFPB). The three main consumer credit reporting rating agencies are Equifax, TransUnion, and Experian. While all three play a huge role in the credit industry, they mainly specialize in the "past credit behavior" aspect of the credit decision. Additionally, companies like LexisNexis and Onfido are positioned in the "identity" aspect of the credit decision for verifying the identity of credit applicants. And companies like IBM and SAS play large roles in fraud risk assessment.

The one aspect of the credit decision for lenders that has yet to establish a clear leader is the ability to pay components, and this is where Aire focuses its business model. The ability to pay aspect of the credit decision was a new requirement implemented into the credit decision process following the 2008 financial crisis, the fallout from which changed the credit industry entirely. Procedures around credit issuance were tightened (for example, so called "NINJA" mortgages are no longer issued), and creditors had to start placing much more emphasis on the financial well-being of potential debtors.[9]

Presently, there simply aren't many competitors in the ability to pay space. Equifax has an initiative called "The Work Number" that provides access to income and employment data for an individual applicant, but The Work Number does not directly assess ability to pay. Thus, creditors have largely relied on ad-hoc and manual assessments of ability to pay, with lenders using employees to manually reach out to applicants.[10] This status quo is both costly and adds little value to the credit risk modeling process.

---

[9] "NINJA" is an abbreviation for "no income, no job, and no assets." Such mortgages were extended to borrowers that stood little chance of being able to repay the loan. Lenders also frequently did not confirm the applicant's assets.

[10] Frequently, this takes the form of (potentially long) individual phone calls.

A systematic product that can analyze ability to pay at scale has the potential to generate large value for creditors.[11] By replacing manual outreach, a scalable product can save lenders time and money. Aire's internal estimates claim it can save lenders "up to $78 per existing customer" (Aire), which can result in massive cost reductions when applied to an entire customer base. Beyond labor costs, automating the ability to pay assessment within the lending process speeds up lending decisions and makes building evidence for regulators that lending decisions meet standards easier.

Given the business need, and the lack of incumbents, Aire launched a product called *Pulse* targeting this market. Pulse is a service that provides "actionable insight on the current financial situation of your existing customers, fast" (Aire). It focuses on three main areas of a customer's financial health: validated total income, non-discretionary expenditure, and a measure of engagement.[12] The final piece of the Pulse product is a prediction of a credit applicant's ability to pay, which can be used by a credit issuer in the credit decision process.

Competitive Landscape

As previously mentioned, the ability to pay factor in the credit space does not have a clear leader. *Pulse* by Aire is targeting this sector. Their main competitor, *The Work Number* by Equifax, has proven itself useful within the space, but not specifically targeting forward-looking ability to pay. *The Work Number*'s main purpose is validating an applicant's identity, education status, and tax data. *The Work Number* also utilizes its technology to match an individual with the Social Security Administration, giving lenders the ability to obtain tax and wage statements, painting a better picture of the candidate's historical records (The Work Number). While *The Work Number* produces information adjacent to the ability to pay, *Pulse* is the only product directly focusing on

[11] Such a shift would echo the larger history of the banking industry, from decision-making at local levels using in-person processes with soft information to automate processes driven by data.
[12] For validated total income, "Pulse returns a USD value for the gross income that a consumer receives, or has reasonable expectation of access to," which is a value that is validated by Aire's unique set of data. The non-discretionary expenditure value provides "[a]n individualized view of expenditure that gives greater insight into a consumer's ability to service additional debt, and withstand shocks" (Aire).

creating forward looking projections to advise lenders on how affordable credit is for a given applicant.

However, while the current landscape is favorable for developing an ability to pay model, the key risk is the entrance of competitors.

The recent explosion in available data, easily scalable cloud computing, and the wider use and understanding of AI algorithms has dramatically lowered the barriers to entry. As such, innovation at each stage of the lending process competition has become more competitive and fluid. Indeed, according to the AI Opportunity Landscape Research conducted by Emerj Research Analytics, "approximately 15% of the venture funding raised for AI vendors in the banking industry is for lending solutions" (Faggella).[13]

To show why these shifts can be significant, we highlight a recent paper discussing the relationship between digital footprints and credit scores. The paper looks at five variables describing the "digital footprint" of a loan applicant: The operating system the borrower uses when applying for a loan, the type of device, time of application, the email domain (e.g. gmail, hotmail, edu, etc), and whether your email contains your name (Berg et. al). The researchers demonstrate that these variables can outperform traditional credit score models in predicting if a loan will be paid back.

This style of credit assessment is already being put into the marketplace by FinTech startups. For example, one startup that is aggressively using AI in the credit sector is Lenddo. Similar to the study above, Lenddo creates a credit score using the digital footprint of applicants by examining 12,000 factors that includes social media account use, internet browsing, geolocation data, and other smartphone information. Recently, FICO, the global credit agency, announced partnerships with Lenddo to use their technology in the new FICO scoring systems in India (FICO).

For startups and incumbents experimenting with new ways to use Big Data and AI in the lending process, especially in the U.S. and Europe, the key hurdle is contending with existing regulations

[13] Startups are not the only firms using AI in this space. Equifax has several patented products using AI techniques, such as neural networks.

and norms. For example, while the digital footprint paper's methodology has predictive power, many "digital footprint" variables will likely fall under protected classes (Berg et. al). Moreover, achieving widespread adoption by customers in the U.S. would likely prove difficult due to privacy concerns. The model we discuss below uses data that in its analogous implementation in the U.K. does not prompt privacy concerns. The remaining regulatory hurdles represent both a (solvable) problem to which we dedicate much discussion below and also a barrier to entry for new companies which benefits Aire.

5. Modeling Income Stability

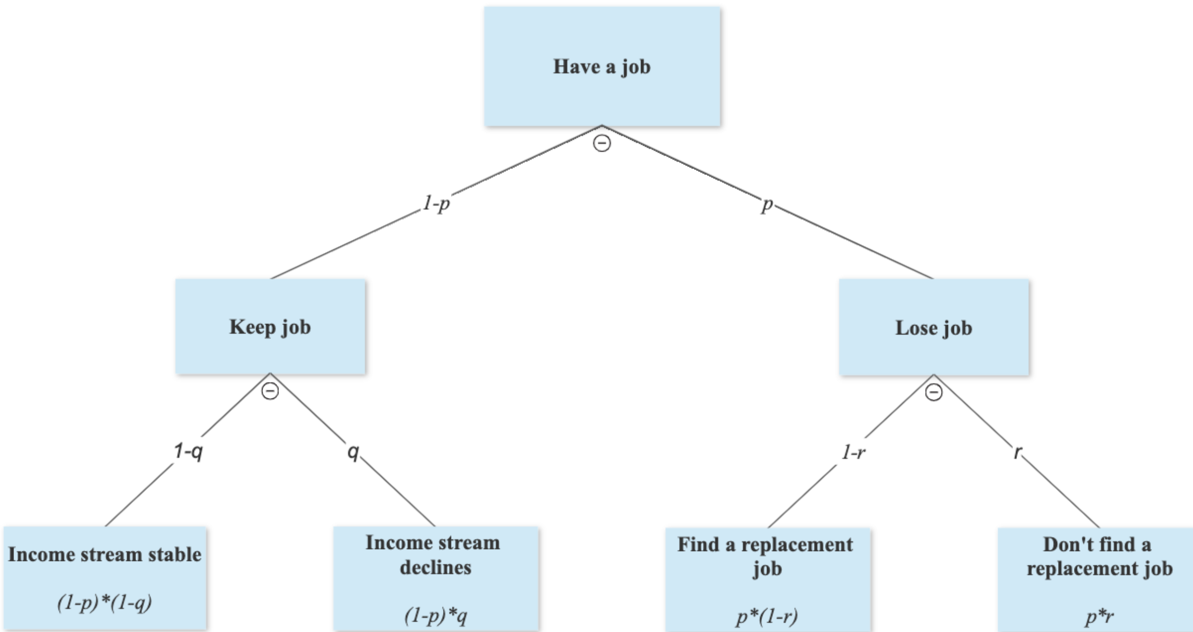
Ability to pay is defined as future income less expenses. **Because expenses remain relatively stable over time for most consumers, we focus on modeling income stability.** Below is an in-depth analysis of the income-stability sub-problem and our proposed solution. However, before we dive into the factors that affect the stability of future income, we first must define what we mean by the term "income stability."

We define income stability as the likelihood that an individual's future income over the next two years will be greater than 90% of the individual's current income.

We focus on the scenario of a loan applicant with one job, which covers the majority of users. We model the probability that their income remains above 90% of their current income as a function of whether they remain in their current job (and its income remains at current levels) or they lose their job but find a replacement. **Figure 1** depicts the model.

The first branch of the tree addresses the probability that an applicant loses his or her job and is denoted by "p." This probability is estimated based on the user's current job title and employer data, which are gathered by the existing Aire user interface. This information is matched to the sector in which the applicant works in order to better determine the probability of keeping a job in said sector. In other words, "p" is a function of sector-level employment conditions. We assume that the probability of losing a job remains constant throughout the forecast horizon.[14]

Figure 1: Probability Tree Underlying Model for Applicant with One Job



To account for the fact that earnings for a given job might not be fixed (e.g. if the applicant's job is paid hourly or on commission), we introduce the branches off of the "Keep job" node. With some probability "q", their current job income declines. For simplicity and to make initial execution easier, we can use the simplifying assumption that the incomes of a given job do not fall ("q=0"). This assumption implies that no job has downward surprises in income and is demonstrably false. However, this assumption likely performs well for a large portion of applicants, particularly those with salaried jobs, and can be implemented immediately.

[14] Explicitly altering the existing prediction, which is a one-month prediction, to longer horizons by looking at macroeconomic projections might be sensible. Another extension would be to estimate the uncertainty of such projections, though the "structural" model we present here is focused on a point estimate. That is, our model estimates the probability of unstable income, without error bands.

If the applicant happens to lose their job at some point over the next two years (with probability “p”), keeping a stable income requires finding a new job with sufficient pay. We propose estimating “1-r”, the probability that the applicant finds a new job with income greater than 90% of their previous job’s income, by exploiting job board data. Formally, we can write the joint probability of finding a job and that job having a sufficient salary as $1 - r = Pr(J \cap S) = Pr(S|J) * Pr(J)$, where J denotes the event of finding a job and S denotes the event where the salary is above 90% of current income.

Because the Aire UI already collects the user’s location and job title and matches it to job board data, this can be leveraged to estimate the components of “1-r” accounting for location and job-specific labor market conditions. For a given job title and location, the data contains the distribution of salary offers and information on job postings and fillings. First, this offers a direct way to measure $Pr(S|J)$: one minus the CDF value of matched salary offers at 90% times the applicant’s current income. Second, we can use this to estimate the probability of finding a job, Pr(J), given that the applicant is unemployed. An initial way to estimate this is by comparing the number of available jobs in the surrounding area with the size of the unemployed working age population.

Having estimated “p”, “q”, and “1-r”, we can compute the probability that income is unstable as the probability that a user reaches the second (“Income stream declines”) and fourth (“Don’t find replacement job”) nodes of the final level of the tree.

Formally, the probability that income is unstable, “Y”, can be written as

$$Y = (1 - p) \cdot q + p \cdot r$$

This probability can be passed to lenders as an output of the Aire API to assess the overall income stability of an applicant in order to better determine creditworthiness.

Implementation of Model

We consider the feasibility of this model in relation to its impact on Aire and the general user base of credit applicants. The key considerations are data costs, user completion rates, and model run time.

A primary consideration regarding the implementation of the model is costs, which primarily stem from buying data and the labor required to gather, clean, and integrate it with existing data. The model constructed above does not require any data not already available within the current Aire data ecosystem, and as such, costs are limited to model development and testing.

The next consideration is completion rates. Lenders send applicants to the Aire website and interview in order to gather more concentrated data to inform credit decisions. Aire is paid for passing back information to the lender, which can not be done until applicants complete the interview. Thus, completion rates are the key metric driving revenue (besides the number of initial referrals from lenders). Importantly, as introduced above, our income stability model requires no change to the user interface and as such, will not impact completion rates.

The final consideration for Aire is the run time of their models. Speed is crucial for Aire since slow run times would critically delay the delivery of data insights to lenders. With lenders increasingly competing on the speed of their lending decisions, slowdowns could put Aire in a precarious position with their clients and make them vulnerable to entrants.[15]

The income stability model above should operate at speeds comparable to existing Aire prediction models as it does not require additional data or interview questions, and all of the components are easy to calculate. "p" is already computed upstream before the income stability model, "q" is assumed as zero, and "r" has two components which can be pre-calculated and retrieved once an applicant's location and job are given.

[15] For example, this commercial for Rocket Mortgage touts the speed of refinancing decisions on their platform: https://www.youtube.com/watch?v=v7ojmCETe-0

Regulatory Discussion of Model

In order to ensure that our income stability model is in full compliance with regulations, we have done extensive research in the credit lending industry in order to better understand the legislation that Aire is bound by in the United States. As previously stated in the "Impact On Industry - Aire Compliance" section of our report, the anti-discrimination emphasis from the Equal Credit Opportunity Act (ECOA) is the major focal point for Aire's regulatory compliance efforts.

The equation for the probability that income is unstable, "Y", written as $Y = (1 - p) \cdot q + p \cdot r$ **has** all three factors ("p", "q", and "r") at play. To assess how this model fits within the existing regulatory space, we will consider each factor in turn.

As previously noted, "p" is the probability that an applicant keeps his or her job. It is essentially a function of sector-level employment conditions as the probability is estimated based on the user's current job title and employer data. Because this is operating at a sector level, it is not going to discriminate against any protected classes, which are defined at the individual level.

Next, "q" is the probability that the applicant's current job income declines. For simplicity and to make initial execution easier, we have decided to assume that the incomes of a given job do not fall ("q=0"). Therefore, since everyone is assigned the same value for "q", there is no bias in this portion of the model.

Finally, "r" is the probability that an applicant won't find a new job with income greater than 90% of their previous job's income, in the event that they lose their current job. The calculation of "r" is mapped to job board data for location and job-specific labor market conditions. As with the argument for "p", there is no opportunity for this variable to bias against individuals in protected classes.

6. Future Improvements

Improving Model Components

While our model is designed to accurately predict future income stability, we were concerned with the degree of individualized predictions. That is, will the model above assign the same income stability to all real estate appraisers in Phoenix? Banks would probably hesitate to use models that do so because sector-level predictions do not provide enough information about a credit applicant's individual situation in order to inform a credit lending decision.

The discussion of the base model proposals above suggests this might be the case, because "p" and "r" are based on information about the labor markets for specific jobs in specific locations. However, this is not the case. Crucially, "r" is conditioned on an applicant's current income. This ensures that the model generates different predictions for two individuals with the same job in the same location if they have different salaries. Moreover, the extensions discussed below would make the model's output depend on additional individual factors (but not protected factors).

Here, we discuss possible improvements and issues with the above definitions of the key model components: "p", "q", and "r".

We first propose an improvement to "p", which we call "p2". This improvement can be achieved because we believe that we can individualize the sector level prediction based on the size of an applicant's current employer. Firms of different sizes have different layoffs rates, and the ratio between layoff rates for firms in different size buckets is consistent across the business cycle (United States. Bureau of Labor Statistics). Specifically, the larger the firm size, the lower the layoff rates.

We can exploit this by matching the employer of the applicant to a database containing firm names and sizes. For example, publicly listed firms disclose the number of employees. For applicants whose companies do not have publicly available employment data, a question can be added to the Aire UI to ask for the number of employees at their firm.[16]

Our baseline model makes the simplifying assumption that "q", the probability that a current job's income declines, is zero. That is, we assume incomes of a given job do not fall ("q=0").

The key insight is that some jobs have more (downward) variability than others. Jobs that are non-salary or non-union, or whose pay significantly depends on hourly wages, tips, commissions, and bonuses are more prone to downward surprises in total income. Additionally, the total income at some jobs (e.g. gig economy jobs) is more sensitive to poor macroeconomic or company conditions.

The discussion below does not propose exact methods for measuring "q", but is more speculative and provides ideas about how to proceed.

One way to capture that intuition is to exploit the user's job title, which is already collected within the existing Aire UI, cleaned, and then matched to several job-title level databases. One option is to examine income distributions based on job title, which is already available to Aire. A second option is BLS occupation-level data on the variation of incomes for each job type.[17] This BLS dataset indicates which jobs are typically salaried. For users with job titles where income often comes from several components or varies widely (say, hourly jobs without consistent hours), the interview could add a conditional question to the Aire UI to better predict the downside risk. A third option is to leverage the employer itself, as some employers' employees will have more downside risk in wages (e.g. the likelihood of furloughs and the composition of wages will vary across industries and by firm size).

[16] One way to do this is to add a conditional drop down to the user interview. If the applicant's firm is not in a dataset with the number of employees, the dropdown would ask the applicant to select the number of employees in their firm. It would be relatively easy for the applicant if the dropdown simply listed firm size buckets that correspond to BLS firm size deciles. This would build a self-reinforcing dataset, which would be valuable down the road. Because the question would be simple and have 5-10 options, completion rates should not be affected.

[17] The BLS OEWS dataset indicates which jobs (filtered by job title) are paid in salary only versus hourly wages.

If a dataset containing total income, job type, and employer size/industry for individuals over many years can be matched to the above job level databases, it is possible to train a model to predict downward income revisions conditional on remaining in the same job.

Finally, we discuss one approach to improving the estimation of "r", which captures the probability that the applicant does not find a new job with income greater than 90% of their previous job's income in the event that they lose their original job. This does not capture the length of time spent looking for a job. The longer the individual is jobless, the higher the next job's salary will need to be to offset the lost earnings during unemployment.

By analyzing job board data regarding the average length of time that a particular industry/location/job type position sits open on a job board before it is filled, we can get a better understanding of how long it will take for an applicant to find a new job that fits this particular industry/location/job type description. The shorter the average time that said position stays open, the more quickly we can assume that the applicant will be able to attain a new job. The same logic follows for the longer average times as well. We shall call this improved version of r, "r2."

More Model Scenarios

The current version of the model begins with the assumption that an individual currently has a job. According to the March 2021 estimates of the Employment-Population Ratio from the Bureau of Labor Statistics, 57.8% of working age adults are employed. Of the remaining 42.2% of adults without a job, a non-trivial fraction will apply for credit on the basis of spousal income. This is a large blind spot for our proposed model, but it could be covered if all of the questions are changed to ask for employment information for the spouse.

Additionally, the model scenario assumes that the applicant has only one job. Currently, 2.6% of working age adults have multiple jobs according to the Bureau of Labor Statistics.[18] As the "gig

[18] "The U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) defines a multiple jobholder as anyone who holds two or more jobs in a quarter and at least one of these jobs is a long-lasting, stable job — meaning a job held in the previous, current and subsequent quarters" (Bailey and Spletzer).

economy” expands, so will the number of people with multiple jobs. This poses an interesting problem for our income stability model, because having multiple jobs offers some diversification protection for income, while at the same time indicating likely exposure to downturns that differentially impact workers in service sector jobs.

Finally, the model above might not apply well to individuals whose primary income is due to passive income (e.g. real estate). This income is more sensitive to aggregate macroeconomic conditions, but the number of individuals this applies to is paltry.

Extensions to Model Output

The model estimates the probability that income remains above 90% of the current income for the next two years. Lenders might rationally expect to use different thresholds (e.g. 85% instead of 90%) and forecast horizons (e.g., 6 months, 1 year, or 2 years). The formula can be adjusted to output metrics at each of these levels, and lenders can choose according to their preferences. This would give our model more fluidity across lenders that Aire works with.

7. Conclusion

The credit industry has seen many changes in recent years with the ever increasing access to Big Data. What was once a process founded on personal relationships has transitioned to an industry centered on hard information. These changes continue as new regulations are implemented to both protect applicants and ensure the health of the U.S. credit industry. In this report, we focus on the mandate for lenders to confirm an applicant’s ability to pay. While regulation presents additional due diligence challenges for lenders, it also creates opportunities for companies to provide solutions to banks and lenders to help determine ability to pay.

We propose a forward-looking model that predicts a credit applicant’s income stability, in order to better determine ability to pay. Our model is feasible and can be efficiently implemented with very little economic cost. Additionally, the model’s design follows all regulatory guidelines to ensure that privacy is protected and bias is removed. We discuss potential improvements to our initial proposal, focusing on adding more individual factors and covering more applicant situations, such as "gig economy" workers.[19]

[19] "Gig economy" workers are individuals that work as independent contractors for various online platforms (such as DoorDash and Uber).

References

"Ability to Pay' Rules." *Consumer Action* , 13 Nov. 2018, www.consumer-action.org/helpdesk/articles/ability_to_pay_rules#:~:text=The%20Credit%20Card%20Accountability%2C%20Responsibility,in%20an%20existing%20credit%20line.

Aire, 17 Sept. 2020, aire.io/.

Bailey, Keith A., and James R. Spletzer. "A New Way to Measure How Many Americans Work More Than One Job." The United States Census Bureau, 3 Feb. 2021, www.census.gov/library/stories/2021/02/new-way-to-measure-how-many-americans-work-more-than-one-job.html.

Berg, Tobias, et al. "On the Rise of FinTechs: Credit Scoring Using Digital Footprints." *OUP Academic*, Oxford University Press, 12 Sept. 2019, academic.oup.com/rfs/article/33/7/2845/5568311?login=true#204669164.

CFPB. "CFPB Consumer Laws and Regulations ECOA." CFPB Consumer Laws and Regulations, June 2013, files.consumerfinance.gov/f/201306_cfpb_laws-and-regulations_ecoa-combined-june-2013.pdf.

Conlan, Rae. "Q&A: Machine Learning and Explainable AI in Credit Risk." *Equifax Insights Blog*, 30 Oct. 2020, insight.equifax.com/qa-machine-learning-explainable-ai-in-credit-risk/.

"Consumer Lending: Understanding Today's Empowered Borrower." *PWC*, 2015, www.pwc.com/us/en/consumer-finance/assets/pwc-consumer-lending-radar.pdf.

Credit Card Accountability Responsibility and Disclosure Act of 2009, U.S. G.P.O., 2009. https://www.ftc.gov/sites/default/files/documents/statutes/credit-card-accountability-responsibility-and-disclosure-act-2009-credit-card-act/credit-card-pub-l-111-24_0.pdf.

"Fair Lending." *OCC*, 6 Apr. 2019, www.occ.gov/topics/consumers-and-communities/consumer-protection/fair-lending/index-fair-lending.html.

"FICO and Lenddo Partner to Extend Credit Reach in India." *FICO*, 3 Oct. 2016, www.fico.com/en/newsroom/fico-and-lenddo-partner-extend-credit-reach-india.

Faggella, Daniel. "Artificial Intelligence Applications for Lending and Loan Management." *Emerj*, Emerj, 3 Apr. 2020, www.emerj.com/ai-sector-overviews/artificial-intelligence-applications-lending-loan-management/.

Fay, Bill. "Truth in Lending Act (TILA) – Consumer Rights & Protections." *Debt.org*, 15 June 2020, www.debt.org/credit/your-consumer-rights/truth-lending-act/.

Kagan, Julia. "NINJA Loan Definition." *Investopedia*, Investopedia, 14 Dec. 2020, www.investopedia.com/terms/n/ninja-loan.asp.

Klein, Aaron. "Credit Denial in the Age of AI." *Brookings*, Brookings, 25 Oct. 2019, www.brookings.edu/research/credit-denial-in-the-age-of-ai/.

Liberti, Jose Maria, and Mitchell A. Petersen. "Information: Hard and Soft - Kellogg School of Management." *Kellogg School of Management* , June 2018, www.kellogg.northwestern.edu/faculty/petersen/htm/papers/hard%20and%20soft%20information.pdf.

"Links to Charts and Tables for a Firm Size Class." U.S. Bureau of Labor Statistics, U.S. Bureau of Labor Statistics, 27 Jan. 2021, www.bls.gov/bdm/bdmfirmsize.htm#SIZE4.

Mishkin, Barbara S. "CFPB Eliminates Independent ability to pay Requirement for Credit Card Applicants 21 or Older." *Consumer Finance Monitor*, 2 May 2013, www.consumerfinancemonitor.com/2013/05/02/cfpb-eliminates-independent-ability to pay-requirement-for-credit-card-applicants-21-or-older/.

"OEWS News Release." *U.S. Bureau of Labor Statistics*, U.S. Bureau of Labor Statistics, Mar. 2021, www.bls.gov/oes/home.htm.

Shrm. "What Are Disparate Impact and Disparate Treatment?" *SHRM*, SHRM, 4 Jan. 2021, www.shrm.org/resourcesandtools/tools-and-samples/hr-qa/pages/disparateimpactdisparatetreatment.aspx.

Steele, Jason. "What Is the Credit CARD Act of 2009?" *Experian*, Experian, 20 May 2019, www.experian.com/blogs/ask-experian/what-is-the-credit-card-act-of-2009/.

The Work Number, Equifax, 2021, theworknumber.com/.

Tracy, Paul. "Ability to Pay." *InvestingAnswers*, 16 Jan. 2021, investinganswers.com/dictionary/a/ability-pay#:~:text=Ability%20to%20pay%20refers%20to%20a%20borrower%E2%80%99s,capacity%20to%20make%20good%20on%20his%20loan%20obligations.

U.S. Bureau of Labor Statistics, Employment-Population Ratio [EMRATIO], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/EMRATIO, April 19, 2021.

U.S. Bureau of Labor Statistics, Multiple Jobholders as a Percent of Employed [LNS12026620], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/LNS12026620, April 19, 2021.

U.S. G.P.O., U.S. "§ 226.51 Ability to Pay. ." *Govinfo*, 1 Jan. 2011, www.govinfo.gov/app/details/CFR-2011-title12-vol3/CFR-2011-title12-vol3-sec226-51/summary.

Biographies

Student Biographies

Austin Leopold is a senior at Lehigh University studying Finance, Financial Technology and Business Information Systems. He has held various roles in a number of University groups, including Lehigh TAMID, Vice President of the Interfraternal Council, and actively participating in the Lehigh Coders Club. This past summer, Austin was prepared to intern at SkyBridge Capital, a fund-of-funds based in New York City. However, due to the COVID-19 pandemic his internship was cancelled along with other hiring programs within the firm. Upon graduation, Austin will join Crowe LLP as a financial services technology consultant, focusing on internal technology audit for financial firms. Austin hopes to pursue career steps in the financial technology realm and beyond.

Chris Pento is a senior at Lehigh University majoring in Finance and double minoring in Financial Technology and Business Information Systems. He also is set to receive a certificate in Business Analytics. He is a member of Lehigh University's business honor society – Beta Gamma Sigma – and is a member of the Finance Club at the school as well. This past summer, he interned at Alpine Global, which is a small investment firm in New York City. Set to graduate in May, Chris is still unsure of his plans post-graduation. He may return to school for a fifth year on the Presidential Scholarship – a scholarship awarded to those who complete a degree with above a 3.75 G.P.A. - to obtain another undergraduate major, or he will pursue one of the several job opportunities that he is currently still interviewing for in the finance industry. Chris hopes to obtain his Series 7 license within two years of graduation and plans to get a Masters of Business Administration at some point in the future.

Rafiq Sitabkhan is a senior at Lehigh University majoring in Finance and minoring in Financial Technology. He is a member of the Lehigh TAMID chapter in the investment fund group, a member of the Finance Club, and a member of the Investment Management Group. This past summer Rafiq interned at The Vanguard Group as a Corporate Finance Intern. Upon graduation, he will join the rotational Financial Leadership Development Program (FLDP) as a Financial Analyst at Johnson & Johnson in New Brunswick, NJ. Rafiq hopes to pursue further education by obtaining his MBA after gaining industry experience.

Faculty Advisor Biography

Donald Bowen joined the Perella Department of Finance at Lehigh University in 2019. His research focuses on corporate finance, with particular focus on the interplay between corporate investment, innovation, patent markets, venture capital, and IPOs. Professor Bowen's research has been published in Management Science and presented broadly, including the Western Finance Association, NBER, UBC Winter Finance, Finance Research Association (FRA), and Midwestern Finance Association.