Scoring the Unscorables: Credit Analysis for Underbanked Households

The decision to make credit available for many households rests in large part with the perceived risk of that individual. Many lenders measure this risk –and make decisions based on this risk - by requesting credit reports from credit reporting agencies (CRAs). Over the last 25 years, this practice has become increasingly popular, especially in the consumer credit space. However, consumers with limited or no credit history face substantial challenges in accessing various credit opportunities. For example, if a consumer does not have a credit record with one of the CRAs, lenders often respond by denying credit or overprice them by charging higher than normal interest rate. Therefore, in a way, the credit models does not accurately capture or even worse, excludes certain population.

Specifically, these underbanked consumers are divided into two groups: the "unscorables" and the "credit invisibles." The former are comprised of individuals with insufficient CRA data. This means they do not have enough credit or payment history for the model to generate a reliable credit score. A typical example is a recent college graduate, some "millennials" and those recovering from earlier financial troubles. The latter are individuals with no available CRA information, and include new immigrants and refugees. FICO estimates the first group to be 28 million and the second group to be 25 million[1].

Traditionally, credit-scoring models impose minimum scoring criteria. In the case of FICO, they require three[2]; consumer must be alive, the credit file must have one trade that's more than 6 months old and the credit file must have a reported trade line within the last 6 months. A new immigrant will only pass a third of the requirement.

According to the CFPB[3], there has been a significant effort to develop models that can capture individuals with little or no credit history by using Alternative Data. Alternative Data is consumer information that is not in the traditional CRA file, and therefore not incorporated into credit score models. They include information relating to non-financial activities such as energy utilities, telecoms, pay TV and rent among others[4].

The proposal for this research is to blend two sources of data: traditional CRA data and Alternative Data. We will construct a benchmark-scoring model that optimally draws from all the variables in the CRA set. We then restrict the variable space to be only the subset of variables that are also available for the

[1] http://www.fico.com/en/latest-thinking/white-papers/can-alternative-data-expand-credit-access
[2] http://www.fico.com/en/wp-content/secure_upload/70_Insights_To_Score_or_Not_To_Score_3009WP.pdf
[3] http://files.consumerfinance.gov/f/201505_cfpb_data-point-credit-invisibles.pdf
[4] http://www.perc.net/wp-content/uploads/2015/03/ResearchConsensus.pdf

Alternative set. We will define a measure of sub optimality between the benchmark and the restricted model and assess the quality of the suboptimal scores.

We then score the subjects in the Alternative file using the model calibrated on the CRA and finally compare these scores with scores obtained from the Alternative set only.

We will also explore the notion of "risk equality" – does a 750 on a traditional CRA file represent the same level of risk as that of someone whose score was constructed using the Alternative Data? Other important aspects include understanding the demographics and characteristics of individual with little or no credit history.

We should be able to study the quality of creditworthiness predictions of individual with little or no credit scoring.

The data comes from a leader in the consumer credit modelling space.

Budget Breakdown

Activity	Cost
Access to the Center for Statistics and Analytical Research (for analytical support), including access to student research assistants.	• $12,000
Full access to the super computer cluster at the university for one semester.	• $1,500
Funding for travel to present research to relevant conferences/organizations.	• $1,000
Total	• **$14,500**

Edwin Baidoo

734 Wilkes Drive
Loganville, GA 30052
ebaidoo@students.kennesaw.edu 347-994-4425

EDUCATION

Doctor of Philosophy
Analytics and Data Science, May 2019
Kennesaw State University
Kennesaw, GA

Master of Quantitative Finance, May 2014
Rutgers Business School
Rutgers University, NJ

Bachelor of Arts, June 2011
Brooklyn College, NY
Major: Mathematics

RESEARCH ACTIVITIES

Research Assistant - Kennesaw State University – 2016 - Present
• Business Cycle and Consumer Default Behavior

SAS Student Symposium - 2017
- Modelling Crime: The case of Philadelphia

SAS Global Forum - 2016
- An analysis of Credit Scoring Accuracy: Comparing Time Series and Logistic Regression Models

SAS Analytics Conference - 2015
- Analyzing the impact of Patient Demographics on Hospitalization Duration in Georgia (with a colleague)

Center for logic, algebra and computation - NYC College of Technology - 2009
- Mathematics research in Digital Signature Cryptography

HONORS | AWARDS | ORGANIZATIONS

- Top 8 Student Symposium Team – SAS Student Symposium
- Dean's List, fall 2010, spring 2011
- Golden Key International Honor Society
- International Association of Financial Engineers
- New York State Leaders of Tomorrow Scholarship

TEACHING EXPERIENCE

Lecturer – Business Calculus 08/2016 – 12/2016
Kennesaw State University, Kennesaw, GA

- Fundamentals of calculus - Derivatives, Rate of Change, Integration, etc.
- Application of fundamental concepts to business problems

WORK EXPERIENCE

Data Science Intern 05/2016 – 08/2016
Equifax, Alpharetta, GA

- Performed clustering analysis on consumer credit to analyze aggregate consumer credit path
- Performed a time series analysis on each credit path
- Performed segmentation analysis and modeling of aggregate consumer bucket

Market Analyst 08/2014 - 05/2015
InterContinental Exchange, Atlanta, GA

- Analyzed and settled futures prices for various markets on the exchange as well as interacting with customers to answer questions regarding the settlement procedure
- Analyzed and used market trends to settle options prices for various markets such as cocoa, U.S Dollar Index and Sugar
- Developed a pricing model for the settlements of Natural Gas Liquid (NGL) contracts to eliminate pricing errors and contribute to its effectiveness

Summer Analyst Intern 06/2013 – 12/2013
Assurant Specialty Property, Livingston, NJ

- Identified trends related to causes of incomplete claim submission as well as to make proposals to senior management regarding my findings
- Created a snapshot of the number and dollar amount pertaining to claims that are pending, as well as to discover any discernible trends between 2007 through 2012
- Created a performance indicator relating to aggregate claims pool versus stand-alone pools and the conditions under which it may be prudent for clients to be under each category

Data Parsing Consultant 07/2011–08/2012
Citadel Investment Group, New York, NY

- Calculated prices of Mortgage Backed Securities and developed an excel macro to calculate it as well and also looked up prices from other sources like Bloomberg
- Maintained and corrected any errors within the database to promote efficiency as well as to develop and learn market terminology
- Coordinated with portfolio managers and financial engineers to ensure proper database security

SKILLS

- SAS, Python, Alteryx, Matlab, R, Excel

KENNESAW STATE UNIVERSITY
College of Science and Mathematics
Ph.D. in Analytics and Data Science

Denis Healy, Director
National Bureau of Economic Research
Household Finance Working Group
1050 Massachusetts Ave.
Cambridge, MA 02138

Dear Mr. Healy,

I am writing this letter of recommendation for the Household Finance Working Group research grant application of Edwin Baidoo.

Edwin Baidoo is a second year student in our Ph.D. program in Analytics and Data Science[1]. Edwin's research interests are uniquely positioned within the nexus of consumer credit/data science and mathematics. Edwin has already proved himself to be a promising researcher in this field – he recently presented his paper "*Analysis of Credit Scoring Accuracy comparing Time Series and Logistic Regression*" at the 2016 national analytics conference in Las Vegas, NV. As this paper continues to receive important feedback from both practitioners as well as from his faculty advisors in Finance, Statistics and Computer Science, Edwin is also engaged with a second, but related research– *Scoring The Unscorables: Credit Analysis for the Underbanked Households*, for which he is applying for this grant.

One aspect that makes this research potentially interesting for the NBER is the data. Our university works closely with several companies engaged in consumer lending – particularly in the sub-prime and underserved, unbanked and "credit invisible" markets. These companies have provided us with access to two substantive datasets[2] to support Edwin's research. Data related to Edwin's research can be shared with the NBER.

The first dataset is related to the subprime market. This data includes transaction data, credit attribute data and payment performance data for 17 million subprime cardholders. The second dataset is related to the payday lending domain and includes data for approximately 500,000 payday lending customers across four years, including credit attributes and credit scores. Access to these very rich datasets has formed the foundation of much of Edwin's research.

I will be serving as one of Edwin's senior advisors[3]. While I currently serve as the Director for the Ph.D. program in Analytics and Data Science at Kennesaw State University, I previously worked in the consumer credit industry for 11 years. My positions included Vice President with MasterCard and with VISA EU in London, as well as a Consultant with Accenture in the Financial Services practice. I teach courses in credit scoring and engage the students in

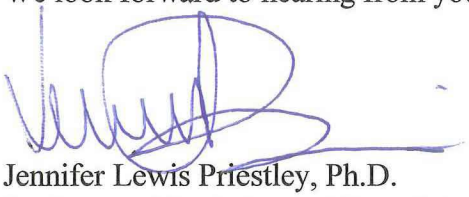
[1] http://csm.kennesaw.edu/datascience/
[2] Both datasets have been purged of personally identifiable information and are approved to be used for research purposes.
[3] https://www.linkedin.com/in/jpriestl/

KENNESAW STATE UNIVERSITY
College of Science and Mathematics
Ph.D. in Analytics and Data Science

consulting/research in consumer credit and risk scoring, with heavy emphasis on machine learning and high performance computing.

The combination of Edwin as a promising young researcher in this field, the substantive and interesting data that we can share for research, combined with my strong domain experience in consumer credit and risk scoring, hopefully strongly positions Edwin's grant application. Should you need additional information from me regarding this recommendation or our program, please do not hesitate to contact me.

We look forward to hearing from you.



Jennifer Lewis Priestley, Ph.D.
Professor of Statistics and Data Science
Director, Ph.D. Analytics and Data Science