The Practice of Credit Risk Modeling for Alternative Lending

Keith Shields, Chief Analytics Officer, Magnify Analytic Solutions, Detroit, Wilmington, Charlotte
Bruce Lund, Consultant, Magnify Analytic Solutions, Detroit, Wilmington, Charlotte

ABSTRACT

In recent years, data scientists in the credit risk profession have experienced less freedom to deviate from industry-accepted practices because the biggest users of credit risk models (banks and large lenders) are the very institutions that face increased regulatory pressure since the 2008 crisis. Regulators have thus been less anxious to bless practices that deviate from the "scorecard and cutoff score" paradigm. But things are changing. The lending industry is demanding innovation and creativity from its data scientists more than ever. The rise of alternative lending, driven mainly by Marketplace lenders (aka Pier-to-Pier lenders), is built on the notion that web-based platforms can integrate customer-facing technology and Big Data to acquire, fund, and underwrite loans in a manner that is more targeted and efficient than banks currently do today. Big Data techniques and "freeware" have increased in popularity, but using SAS® to perform logistic regression and survival analysis on structured data is as good an option for quantifying credit risk as ever. In the "new world" of lending it is not the model fitting techniques that need to change, as much as it is the treatment of samples, potential predictors, and model refits.

INTRODUCTION

A credit risk model is a linear statistical model that uses combination of credit, contract, and personal attributes to predict the likelihood that a loan applicant will default (fail to pay back the loan). The statistical model gives rise to a credit risk scorecard. The scorecard assigns points to the credit application and those applicants whose points are above the "cut-off" are given a loan or line-of-credit. This paper describes the credit risk scorecard and the usage of scorecards in traditional lending by banks and captive finance companies.

But traditional lenders are now competing with the alternative lending industry and new start-up alternative lending companies. Many start-up alternative lenders have distinguished themselves from banks by heavier use of technology, removal of traditional loan structures, and tapping into overlooked sources of collateral and non-traditional funding sources. In this paper the alternative lending industry is surveyed and emerging trends are discussed.

But this paper also discusses what is required for us (statisticians, data scientists, analysts) to have a career in alternative lending. What are the personal traits and statistical skill set that we need? Surely knowledge of scorecards and statistical procedures for model fitting are a requirement, and expertise in the usage of SAS® gives an enormous advantage. But, additionally, personal traits of risk taking and curiosity are needed to meet the demand for Creativity and Disruption that drives success in this industry.

CREDIT RISK SCORECARDS AND TRADITIONAL LENDING

A credit risk scorecard is a system of additive weights that, when applied to a credit applicant at the time of application, creates a credit score, which quantifies an applicant's credit worthiness and provides a convenient means for rank-ordering applicants based on their risk of default (failure to pay back the loan).

When appropriately scaled, credit scores are used by lenders to accept or decline credit applications, and to price loans in accordance with lenders' risk-adjusted return targets.[1]

---

[1] See Siddiqi (2006) for an introductory book on credit risk models.

An Example of an "accept/decline" application is given below:

- *Assume we make $1,000 profit on "good" loans and suffer a $10,000 loss on "bad" loans. Further assume we are charged with determining which of the following 5 applications we should approve:*

Credit Score	% of Applicants	Good Rate	Loss (Bad) Rate	Expected Net Profit	Cumulative Approval Rate	Cumulative Loss Rate
500	40%	99.5%	0.5%	$945	40%	0.5%
400	20%	96.0%	4.0%	$560	60%	1.7%
300	15%	92.0%	8.0%	$120	75%	2.9%
200	15%	87.0%	13.0%	-$430	90%	4.6%
100	10%	75.0%	25.0%	-$1,750	100%	6.7%

- *Some simple math on each credit tranche leads us to very **useful conclusions**:*
    - *Cutoff score = 300. We lose money on borrowers who score below 300.*
    - *We will decline roughly 25% of our applicants.*
    - *Our loss rate will be **2.9%** (as opposed to 6.7% if we approved everyone).*
        - *(40%*.5%+20%*4%+15%*8%) / 75% = **2.9%***
    - *Our profit per contract is $677 (as opposed to $268 if we approved everything).*

The additive weights in a credit scorecard are typically derived from a linear statistical model (called a risk model), which is most often developed by regressing a binary default indicator on combination of credit, contract, and personal "attributes" (in the context of a risk model, the term "attributes" refers to the independent variables). Examples of attributes that are often predictive of default:

- Credit Score

- Loan-to-Value Ratio (LTV)

- Payment-to-Income Ratio (PTI)

- Utilization = (balance on trades / credit limit on trades)

- Recent Credit Inquiries

A simplistic example is given below of converting a risk model to a scorecard. Here, there is only one attribute: X_1

- *Credit Risk Models are converted to "scorecards" by creating a point system whereby the parameter estimates of the credit risk model are multiplied by the possible values of the independent variables to create points.*

- *Example: $Y = B_0 + B_1*X_1$. Suppose X_1 can take on a value of 1, 2, or 3. Scorecard:*

X_1	Points
1	$B_1 * 1 * 100$
2	$B_1 * 2 * 100$
3	$B_1 * 3 * 100$

Over the past several years, credit risk scorecards have become an integral part of how consumer and commercial lending is done.

The pervasiveness and importance of risk models in the lending industry is well evinced by the industry's increasing demand for "data scientists", focus on model governance and "self-policing" (note the Basel III accord), and scrutiny from regulators (note the OCC's *Supervisory Guidance on Model Risk Management:* http://www.federalreserve.gov/bankinforeg/srletters/sr1107a1.pdf)

ALTERNATIVE LENDING INDUSTRY LANDSCAPE

What is an alternative lender? The simplest definition is that an alternative lender is any lender that isn't a bank (*editorial note: we except captive lenders, such as Ford Credit, GMF, etc. … from this definition, in that they function more like traditional lenders*). Historically, the customers of alternative lenders have been people or small businesses who are unable to secure a bank loan. This is still somewhat the case, but not nearly to the extent it once was.

Lending by non-bank entities has grown substantially in recent years. There are several reasons for this:

- Banks became very cautious after the 2008 crisis, leaving a gap between supply and demand for loans, a gap which was rapidly filled by alternative lenders.

- Increased regulation on banks has made it difficult for them to safely return to the pre-2008 underwriting standards.

But lastly, the most compelling reason for the growth of alternative lenders is that they are more progressive in their deployment of technology, big data, and analytics to adjudicate, price, and approve loans in a flexible manner. How does this manifest itself?

- "FINTECH" platforms that allow for easy online loan application, fast response, and lower costs due to the reduction of overhead incurred by allowing the statistical models and technologies within the platform to "decide everything" (or almost everything). Examples: Prosper, Lending Club.

- Removal of traditional loan structures, and tapping into overlooked sources of collateral and non-traditional funding sources. Examples:
  - SoFi: funds raised from alumni to fund consolidation loans for recent graduates: this is classic *Peer-to-Peer lending (https://en.wikipedia.org/wiki/Peer-to-peer_lending)*
  - Kabbage, Credibly: To pay back small business cash advances, the merchant splits the proceeds from credit card transactions with the lender until the advance is paid back.

- Banks repositioning themselves in the marketplace by extending more large commercial loans and fewer consumer loans. The commercial loans often take the form of large warehouse lines of credit to alternative lenders who can exercise more liberal lending practices and achieve scale quickly.

- A renaissance for credit risk modelers and data scientists. Alternative lenders, particularly Fintech lenders, are highly experimental and aggressive about their development and deployment of predictive models. The costs saved from the removal of manual underwriting are largely being redeployed to acquire Analytics talent. This creates huge opportunities and challenges for Analytics professionals – see the next section.

OPPORTUNITIES AND CHALLENGES FOR "SAS PROFESSIONALS"

Opportunities

- **SAS is pervasively used** by alternative lenders, just as it has been pervasively used by banks for years. Most credit risk models are still developed using logistic regression.[2] Some very common syntax for specifying a risk model (using the attributes noted above) follows:

---

[2] See Allison (2012) and SAS Institute (2012) for references in using logistic regression in SAS.

```
PROC LOGISTIC DATA = Loans;
MODEL Default = ln_PaymentToIncome ln_LoantoValue InqL6 RevUtil_ge80;
OUTPUT OUT = Scoredloans / p=PD;
RUN;
```

Where:

ln_PaymentToIncome = log (MonthlyPayment / MonthlyIncome);

ln_LoantoValue = log (Loan Amount / CollateralValue)

InqL6 = # of credit inquiries in the last 6 months

RevUtil_ge80 = binary variable to indicate if applicant uses more than 80% of revolving high credit

PD = point estimate of the probability of default

- Alternative lenders make **huge expenditures on alternative data** (e.g. Yodlee, Clarity, Yelp), which leads to large and difficult data integration exercises specifically aimed at producing usable data marts for Analytics. The SAS programming language is, and always has been, perfectly tailored for this exercise. The exploration of alternative data is such an important exercise that it requires many hours of SAS coding, creating a need for devoted SAS programmers who understand that the ultimate purpose of data marts is to support and deploy models and analysis. The alternative lending industry thus contains a natural separation of duties between SAS programmers and SAS analysts (this is a familiar story by now), with specific, well-defined needs for both.

- Alternative lenders usually strive for **100% technology-enabled, model-driven, loan adjudication and pricing**. This implies that an ongoing need for predictive models and quantitative technologies, and people who can create them, is inherent to the industry.

- Loan acquisition is of paramount importance to any lender. The presence of many alternative lenders is primarily online, so the problem of **loan acquisition begins to look less like a problem of sales presence and more like a problem of predictive analytics**, built upon finding signals from a combination of digital and customer-provided data.

Challenges

- SAS is only one tool in an expanding toolset, and there is **some sentiment that SAS is an overpriced, "old" technology**. R & Python are becoming popular among alternative lenders, but R & Python are, generally, complements to SAS, not substitutes.

- **Small samples** and the "overfitting contest" have aided in the rise of Machine Learning models, some of which are called by account management systems and re-trained in real time. But prior experience with this problem informs us that PROC LOGISTIC in SAS offers a better alternative as part of the following steps:

  − Fix the parameter estimates of attributes we know are important

  − Refit frequently

  − Let the refits tells us if other attributes have become important while we control for what we know to be important. An example where the parameter of CreditScore is fixed at .01 is shown below:
    ```
    DATA loans2; Set loans;
       CreditScoreFix = .01*CreditScore;
    RUN;
    PROC LOGISTIC DATA = loans2;
    MODEL Default = ln_PaymentToIncome ln_LoantoValue InqL6 RevUtil_ge80
    / OFFSET = CreditScoreFix;
    RUN;
    ```

- **Biased samples** are a problem for predictive modelers, irrespective of the tool they use. Many alternative lenders are start-ups and have managed to build successful lending businesses without custom risk models, at least initially. While in their nascent stages these lenders controlled risk not only through generic credit scores, but also by creating acquisition strategies and lending products that attracted inherently lower-risk applicants.

Here are some examples:

- SoFi: student loan consolidation built around Stanford graduates lending to Stanford graduates…this population is not representative of typical student loan risk.
- Merchant Cash Advance (Kabbage, Credibly, many others): high-risk merchants that sign up for a "split" repayment plan (where credit card transactions are split between merchant and lender) are implicitly showing confidence that they have the sales to support repayment. These are not typical high-risk merchants.
- Uber: cash advances for new cars are granted only to drivers who have demonstrated an ability to generate fares.

- **Relentless pressure to grow assets** (before regulation) and a strong appetite from secondary markets for those assets necessitate that modelers refit risk models frequently and reduce their development and deployment times. It also requires that risk models make use of any and all attributes that help lenders differentiate between applicants from the same large pool of previously underserved consumers with marginal or high credit risk.

DEMAND FOR CREATIVITY AND DISRUPTION

Growing a loan portfolio rapidly and consistently requires that a lender increase the pool of applicants to which it can lend. This usually requires that a lender either relax its cutoff scores (and take on more risk), or do a better job identifying "prime" applicants among an ostensibly "subprime" pool of applicants (credit score below 600 is generally viewed by the industry as being subprime).

To accomplish the latter, a lender must have a better risk model than its competitors, and it is difficult to compete with any large bank's ability to quantify risk, because a bank generally has the luxury of developing its models on hundreds of thousands of seasoned loans spanning over many years and business cycles.

Alternative lenders, however, enjoy the advantage of being able to deploy innovative, and sometimes disruptive approaches to risk modeling and underwriting, because they exist in start-up environments that are unencumbered (or less encumbered) by legacy systems, regulatory oversight, and mounting distrust from customers.

In terms of innovation - while there is much attention paid to modeling techniques (machine learning is growing in popularity now as neural nets did in the 1990's), most of the recent innovation is centered on data: more data, better data, and the intelligent structuring of unstructured data. Examples:

- Yodlee (www.yodlee.com) reads bank statement PDF's in real time and allows for the creation of attributes such as average monthly balance, max deposit, standard deviation of deposits, etc., …all of which reduce our reliance on self-reported income.
- Many alternative data providers now offer extensive predictive attributes around driving record, address changes, public record items, etc., …
- There is widespread experimentation with the unstructured data available on Yelp and Twitter:
  - Yelp & small business lending: # of times the word "terrible" has been found in reviews in the last three months
  - Twitter & student lending: # times "broke" is in the same tweet as "student loans"

Alternative lenders also tend to be more innovative not only about the data component of the risk modeling problem, but also the deployment component. Rapid refit and relearn processes are being perfected as more data from disparate sources are being discovered, acquired, and integrated. It is relatively easy to refit routinely when data are appropriately structured, and it is relatively easy to deploy when loan origination platforms are built specifically to house predictive models, massive amounts of consumer data from multiple sources, and enable analytic-driven acquisition strategies in real-time.

Analytic professionals, irrespective of their industries, quickly learn that our predictive modeling abilities are not measured by what we can develop, but by what we can deploy. Alternative lenders' innovative approaches to data integration and model deployment (and NOT TO MODELING TECHNIQUES) are what made them a disruptive influence to Analytics and lending.

SUMMARY: BANKS AND ALTERNATIVE LENDING CONTRASTED

Here we offer a summary of the discussion we began in the last section…there are distinguishing differences between how banks and alternative lenders handle their custom risk models, and they are highlighted below:

Figure 1. Analytics: Banking vs. Alternative Lending

Issue	Banking	Alternative Lenders	How do alternative lenders handle the modeling problems?
Sample Sizes	Large	Small	Strong priors Fix parameter estimates "Wider" samples, not longer
History	Spans multiple economic cycles	1-3 years, these are mostly start-up companies	"Look-alike samples": BUY DATA Refit and relearn Introduce products that control risk without having to model it (e.g. SoFi) Change the target event to something that happens quickly (e.g. For student loans, does the student drop out of school?)
Regulation	Onerous	Light	Actively explore social media (Yelp) Huge investments in data that banks and captives don't think they can deploy
Legacy code	Lots of old SAS processes	No legacy processes	Technology "free for all": R processes are daisy-chained with SAS, then connected to Python. "The IT nightmare": 100% deployment flexibility

HOW "BIG DATA" AND "ALTERNATIVE DATA" ENABLE A RISK MODEL

The denormalization and aggregation processes that create the account level data mart enables a regression of the default indicator on structured data. A logistic regression model is developed (as shown below) and the Regression Equation is determined and is coded into the Acquisition and Scoring System.

The Regression Equation is applied to the "scoreable" data set as shown as the final result of the data flow of Figure 2.
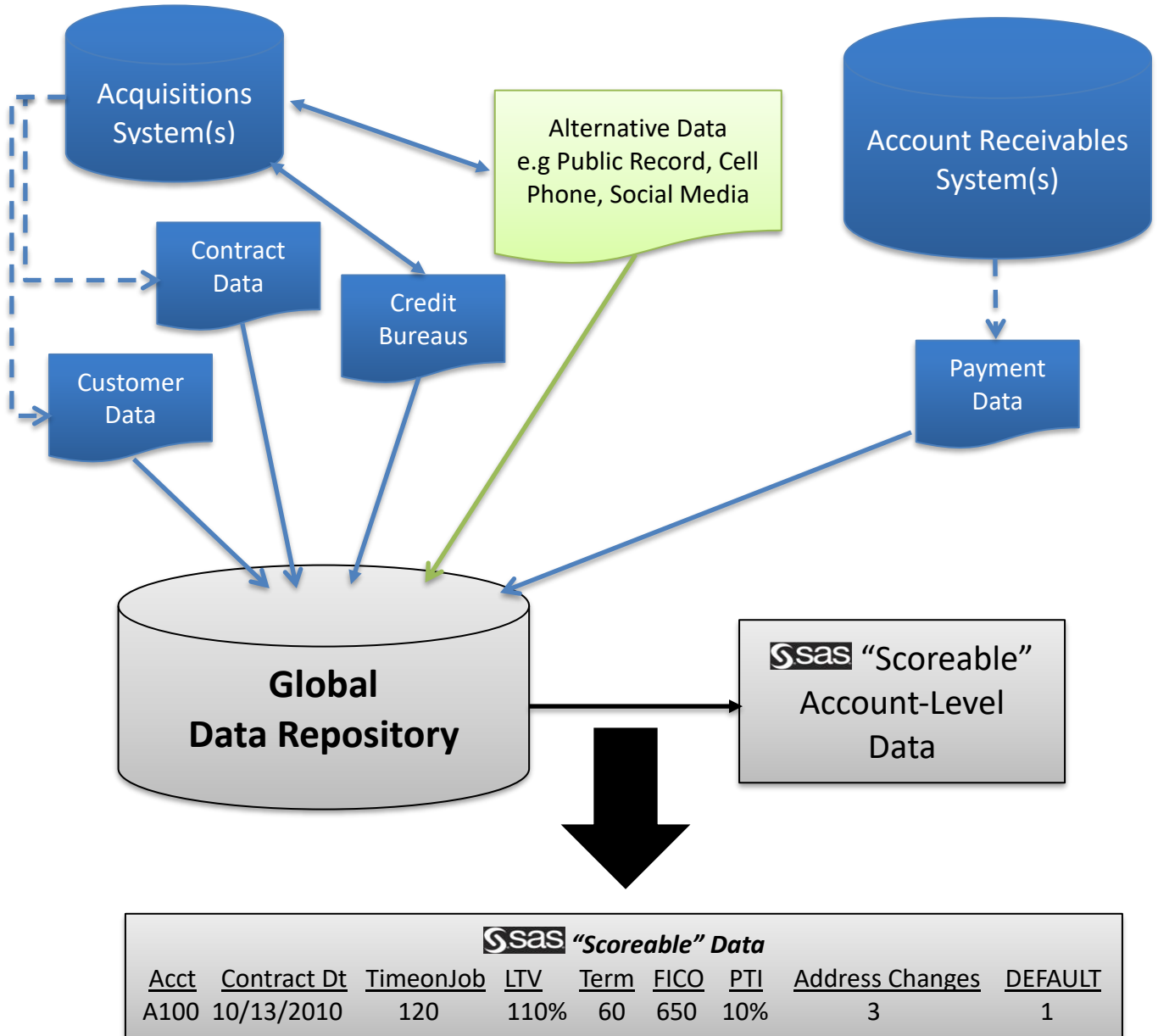
The Logistic Regression Model:

```
PROC LOGISTIC DATA = datamart;
MODEL Default = TimeOnJob FICO Term PTI Address_Changes;
OUTPUT OUT=SCORED / p=PD;
RUN;
```

Regression Equation: $PD = \frac{\exp(Xbeta)}{1+\exp(Xbeta)}$

Where: $Xbeta = b_0 + b_1*TimeOnJob + b_2*FICO + b_3*Term + b_4*PTI + b_5*Address\_Changes$

Figure 2. Acquisition and Scoring System

		$\circledS$ sas *"Scoreable" Data*						
Acct	Contract Dt	TimeonJob	LTV	Term	FICO	PTI	Address Changes	DEFAULT
A100	10/13/2010	120	110%	60	650	10%	3	1

REFERENCES

Allison, P.D. (2012), *Logistic Regression Using SAS: Theory and Application 2nd Ed.*, Cary, NC, SAS
    Institute Inc.
SAS Institute (2012). *Predictive Modeling Using Logistic Regression: Course Notes*, Cary, NC, SAS
    Institute Inc.
Siddiqi, N. (2006). *Credit Risk Scorecards*, Hoboken, NJ, John Wiley & Sons, Inc.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Keith Shields, kshields@magnifyas.com at Magnify Analytic Solutions

Bruce Lund, blund@magnifyas.com at Magnify Analytic Solutions

All code in this paper is provided by Marketing Associates, LLC "as is" without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. Recipients acknowledge and agree that Marketing Associates, LLC shall not be liable for any damages whatsoever arising out of their use of this material. In addition, Marketing Associates, LLC will provide no support for the materials contained herein.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.