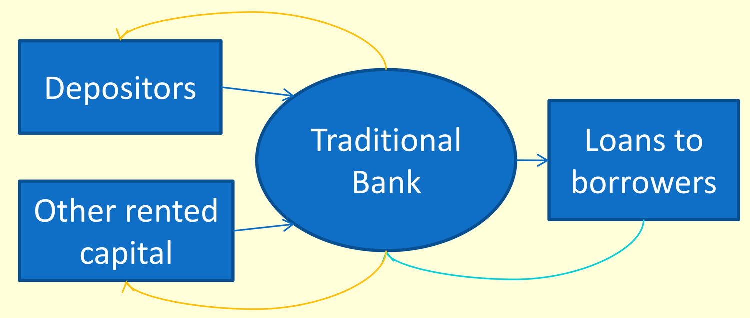
Measuring the Likelihood of Small Business Loan Default: *Community Development Financial Institutions (CDFIs) and the use of Credit-Scoring to Minimize Default Risk*

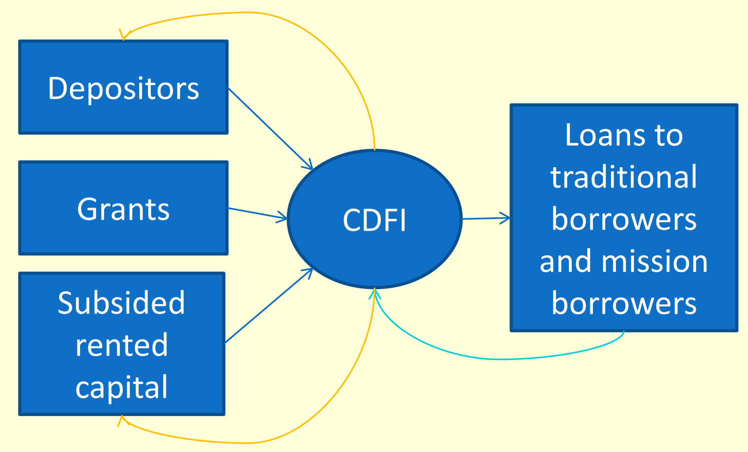
Andrea Coravos, Duke University

Introduction

CDFIs provide financial services to underserved markets and populations. The availability of these services depends on grants and subsidized rented capital. A CDFI receives these investments if extends enough credit to "mission" clientele, which includes women, minorities and low-wealth individuals.

Given the recent recession and the pressure to reduce underwriting costs, many CDFIs are looking for tools to manage the risk in their small business loan (SBL) portfolios. Credit scoring is a statistical technique used to quantify the risk of a loan. Theses models apply relative weights to the explanatory variables. For instance, the borrower's personal consumer credit score, henceforth called a FICO score, is often the most predictive of the small business loan default (Cowan and Cowan 2006). Other predictive variables can include outstanding debt, business type, management experience, and macroeconomic influences (Mester 1997), (Berger and Barrera 2005). An in-house credit-scoring model designed for a specific CDFI's portfolio can increase underwriting efficiency and help extend more credit into the small business community. But will it force a CDFI to "mission-drift" away?





Problem *In a CDFI's portfolio, what influences small business loan default?*

Model

Dependent Variables

Strong	Medium	Weak
Never delinquent, never modified	Ever 30 days late >1x, or more than 60 days late	Ever 90+ days late, charged off

Because the dependent variable in this dataset has three outcomes, this provides an opportunity to run three separate regressions: (**1**) a binary model comparing "strong" loans to "weak and medium" loans, (**2**) a binary model comparing "weak" loans to "strong and medium," and (**3**) a multinomial logistic regression with all three outputs.

Independent Variables

- **Borrower-Specific Characteristics** Such as corporate structure, FICO score, education and industry
- **Loan-Specific Characteristics** Such as guarantee percentage, loan amount, and interest rate
- **Lender-Specific Characteristics** Such as loan-officer identity, bank type, and region
- **Macroeconomic Variables** Such as changes in the business cycle and in unemployment

$$P\left(\frac{SBL_i}{Default}\right) = \beta_0 + \beta_1 * borrower_i + \beta_2 * lender_i + \beta_3 * loan_i + \beta_4 * macroeconomic_i + \varepsilon_i$$

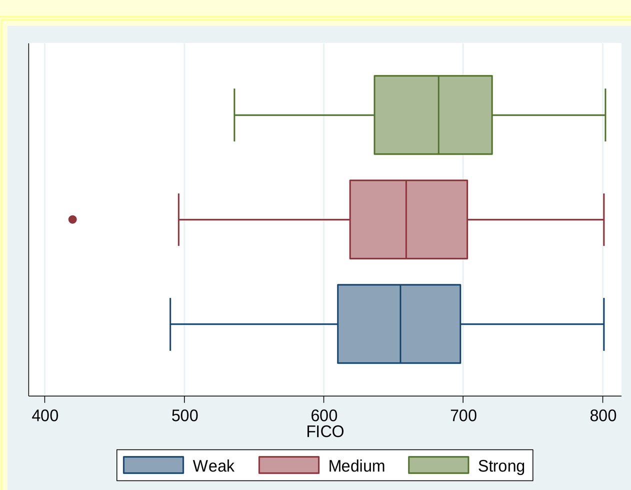
The probability of small business loan (SBL) default depends on four sets of variables. ***Borrower***$_i$ is a vector of self-perceived borrower and business-specific covariates, ***loan***$_i$ is a vector of loan-specific covariates, ***lender***$_i$ is a vector of lender-specific covariates and ***macroeconomic***$_i$ is a vector of macroeconomic covariates. $\boldsymbol{\beta_0}$ is the constant and $\boldsymbol{\varepsilon_i}$ is the error term.

All Loan Results

The dataset contains 530 loans that originated between 2002 and 2007. Of those, 229 are Small Business Administration (SBA) government guaranteed loans. The loans in the dataset have a relatively high default rate: 26% are classified as weak and medium and only 39% are classified as strong.

In the *Strong/Not Strong* model, borrowers with higher **FICO** scores were much more likely to repay their loans.

Figure 1. Better FICO scores are mildly predictive of loan repayment as of Oct 2009

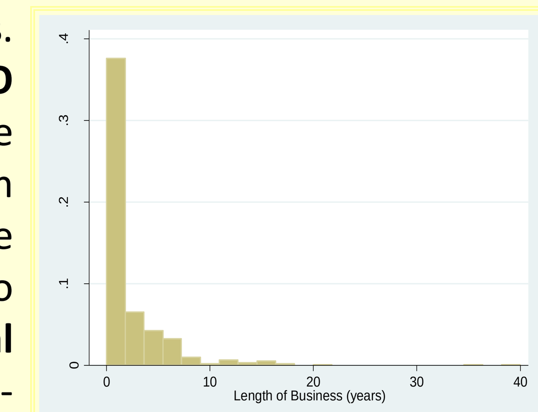


Even though **FICO** is the most predictive, it only explains a small part of the overall default rate (see *Figure 1*). Additionally, borrowers with **management experience** are more likely to be strong. Businesses that experience **large peaks in local unemployment** are more likely to default. Loans with large **government guarantees** are more likely to be weak. This could be due to a moral hazard, or it could be because these government-backed loans are riskier due to unobserved characteristics.

Start-Up Loan Results

Start-up loans are defined as loans given to firms in business for **one year or less.** Compared to *all loans*, **FICO** scores were more predictive of startup SBL default than non-startup. Start-ups were also much more sensitive to changes in the **local unemployment rate.** Start-up borrowers are less sensitive to **interest rates.**

Figure 2. Most of the businesses in the CDFI's portfolio are young or start-ups



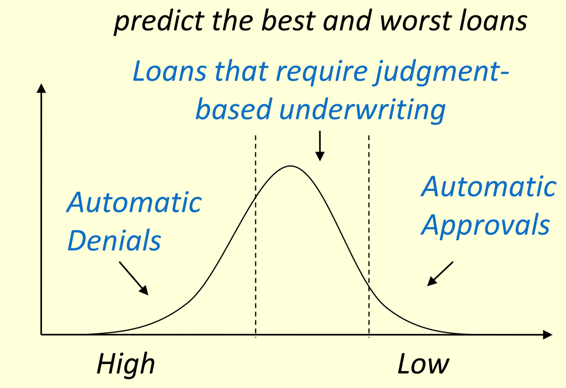
Microloan Results

The industry standard definition of a microloan is **$35,000 or less**. Microloan default is not as well predicted by **FICO** score, likely because micro-borrowers are not well-represented in national credit databases. Micro-borrowers are also more sensitive large deviations in the **interest rate** compared to the Fed Prime rate and to the type of interest rate (**fixed** or **variable**). Compared to other types, microloans given in **strong economic climates** were more likely to default.

Application for Credit Scoring

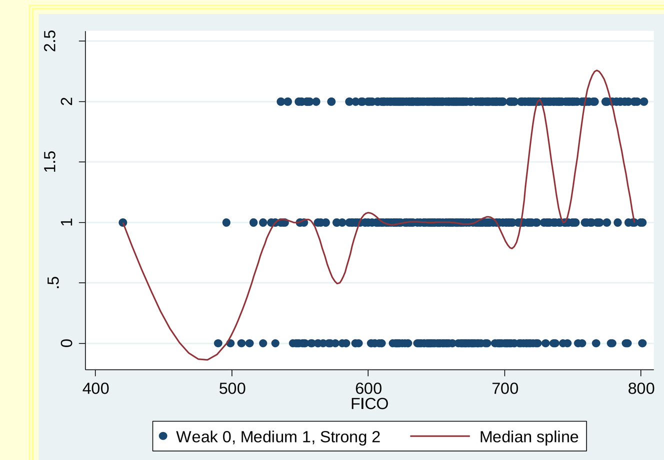
The coefficients from the OLS regressions provide the relative weights for a CDFI credit-scoring model. Another method is to create the probabilities for each loan using its odds ratio outputs from the logistic regression, and then set a threshold probability.

Figure 3. *A credit-scoring model can predict the best and worst loans*



The best loans predicted from the *Strong/Not Strong* model can be used for the automatic approvals. The worst loans in the *Weak/Not Weak* model can be automatically denied. Generally, a CDFI would want to automatically deny more loans than it automatically accepts. The credit-scoring cut-off deserves some attention. If the cut-off score is too high, it will exclude many good loans, which is called a "**Type I**" selection error. A high cut-off will significantly reduce the amount of profit a bank can expect to make on their loans. However, if the cut-off is too low, it will include many bad loans, which is referred to as a "**Type II**" selection error. Low cut-offs erode profits. In this dataset, the thresholds were not naturally specified, and the CDFI would have to determine its acceptable level of risk exposure when setting a cut-off.

Figure 4. Even a spline function cannot isolate natural cut-offs between FICO and loan strength



Conclusion

Even though credit-scoring technologies for consumers and credit cards have been well-developed over the past few decades, credit-scoring is relatively new for small businesses. In 2007, Berger and Barrera noted that no micro-lenders used credit scoring. In the past 3 years, more intuitions have adopted these technologies, including CDFIs, whose portfolios expand beyond microloans. The big question is: do in-house credit-scoring models have inherent biases and disadvantage the target mission clientele? Depending on how the data are sliced, it appears that females are sometimes associated with stronger loans. Furthermore, the minority dummy variable is never significant, even when not fully controlling for other factors. These results suggest that employing this credit model would not cause the CDFI to "mission drift" away. Future research in this area should include a time-sensitive analysis, because many SBLs default in the first few months. This time-effect can a powerful explanatory variable. In addition, these models should be regularly updated as the CDFI digitalizes more data from its loan files and as it underwrites more loans.

Literature Cited

Cowan, C. & Cowan, A. 2006. "A Survey Based Assessment of Financial Institution Use of Credit-scoring for Small Business Lending." U.S. SBA. Retrieved Sept 2009.

Mester, L.. 1997, September."What's the point of credit-scoring?" Business Review, Federal Reserve Bank of Philadelphia, 3-16.

Berger, A. & Barrera, M. 2007. "Credit Scoring for Microenterprise Lenders". Microenterprise FIELD. [http://accionnm.org/PDF/CreditScoring.pdf]

Acknowledgements

I am grateful to Professor Charles Becker for his year-long encouragement and advice. I would like to thank X CDFI for providing the data and incredible support for this project. My thesis Professors Kent Kimbrough and Michelle Connolly substantially shaped these models. I would have not been able to write such an in-depth statistical analysis without the help of Kofi Acquah. My peers in the Economics Workshops 198 and 199 provided great support all year. Thank you to Professor Lori Leachman for granting the Research in Practice Program funding for this work.

Contact Information

Andrea Coravos
PO Box 99905
Durham, NC 27708
arc14@duke.edu