Improving Access to Credit for SMEs:
An Empirical Analysis of the Viability of Pooled Data SME Credit Scoring Models in Brazil, Colombia & Mexico

Margaret Miller
(The World Bank)
and
Dina Rojas
(Fair Isaac Corporation)
October 2004
DRAFT
I. Introduction

Lending to small business has traditionally been a very time consuming and costly proposition for banks and other financial intermediaries. Small firms lack proper accounting procedures and owners frequently mix their business and personal finances so financial statements are unreliable. Prior to the days of credit scoring, credit grantors had to evaluate each credit application individually, asking questions such as:

Do I know the person who runs this business?

Have I ever loaned money to this business before? Was the money paid back and was it paid back on time?

How much is the business owner asking for—pocket change or a large amount?

What other debts does this business have to pay off?

It was a long process. Applicants often had to wait days, or even weeks, for loan approvals. Credit analysts had to use their own judgment in evaluating applicants but since asymmetric information between borrowers and lenders is so acute in the small business credit market, these judgments were often poorly informed representing loan officers’ biases and unique experiences rather than a systematic evaluation of risk. The information asymmetries also served to reduce competition in the SME lending market so firms which were good credit risks often paid handsomely for access to credit, even when the bank or firm lending to them had established their creditworthiness.

Over the past fifteen years, this situation has gradually improved in many developed country markets as small business credit scoring (SBCS) technologies have been adopted
by banks and other lenders. SBCS tools enable lenders to rapidly evaluate the risks associated with different borrowers according to objective and statistically validated criteria. As a result, risks can be better managed, borrowers can be more confident that loan decisions are based on their qualifications and the credit granting process has been streamlined with processing times falling from days and weeks to a matter of minutes in some cases.

In the United States, SBCS gained widespread acceptance in the mid 1990s with the introduction of an off-the-shelf product, the Small Business Scoring Solution produced by the Fair Isaac Corporation. This product was initially developed through a consortium of 17 unaffiliated banks in the United States. The pooled data model which resulted was more robust because of the quantity of data used to create it and was more economical than a custom application. The Fair Isaac pooled data SBCS tool is still the industry standard in the United States small business credit market and this technology has been applied in a limited number of other countries including Canada and Japan.

In most developing country markets, however, SBCS technologies still have only limited use and there has been virtually no development of pooled data SBCS models. This is likely due to a variety of factors including poor lending practices (insider lending, fraud, etc.), small market sizes which do not justify the product development expenses for SBCS and incentives for maintaining relatively labor intensive loan processing technologies rather than automating (low skilled labor costs, institutional incentives to maintain high levels of employment, etc.).
Although some large financial institutions in developing countries have implemented custom designed or expert model SBCS technologies, pooled data models have not been adopted. One reason may be reluctance on the part of lenders to share data on their small business customers, since these borrowers typically pay relatively high interest rates and represent a profitable market segment. The relatively high levels of concentration in some Latin American banking systems further exacerbate this problem, as market-leading financial institutions jealously guard their “information advantage” from their large customer base. They may view the development of SBCS tools which could improve risk analysis throughout the financial system as a way for competitors to close the gap and thus consider them inimical to their self interest.

In this research project, the World Bank has teamed together with the Fair Isaac Corporation to determine the feasibility of developing a pooled data small business credit scoring tool in several Latin American countries: Brazil, Colombia and Mexico. The convening power of the World Bank will reduce the cost of creating a consortium and gaining access to data – part of the high sunk costs for development of these models. Fair Isaac is responsible for assembling the data and data analysis. In this research project the goal is not to develop a functioning product, but rather to determine the feasibility of a pooled model approach with a limited set of data from approximately three to five lenders in each market.
The remainder of the paper is organized as follows. Section II is a brief literature review of both research on the value of shared credit information for small business access to finance and credit scoring literature. Section III discusses international experiences with small business credit scoring. Section IV presents the methodology and project design used in this research. Empirical results (yet to be developed) are in Section V and Section VI provides conclusions.

II. Literature Review

Asymmetric information between borrowers and lenders, which is a defining characteristic of all credit markets, is particularly acute in the “informationally opaque” market for small business credit. Small businesses pose a particular challenge for lenders because of their lack of audited financial statements, commingling of the owner’s personal finances and those of the business and because of their diversity. To evaluate small business creditworthiness, lenders have traditionally relied on three approaches: financial statement lending (focused on a firm’s financial statements); asset-based lending (collateralized lending); or relationship lending (where assessments of the business owner’s character and other informal information are a key part of the lending decision).\(^1\) Of these three lending technologies, relationship lending is the one which is often identified as most characteristic of the small business loan environment.

A significant body of research exists on how the relationship between loan officers and small business owners affects the loan decision and characteristics of the loan market. A decade ago Petersen & Rajan (1994), Berger and Udell (1995), Miller (1995) and others

\(^1\) See Berger and Udell (2002) for more detailed descriptions of these lending types.
discussed the importance of a firm’s proximity to the bank and other aspects of the bank-borrower relationship (such as number of years of the relationship) in obtaining credit. Other authors focused on the negative relationship between bank size and SME lending in the United States. For example, Peek and Rosengren (1995) posited that smaller financial institutions enjoyed an advantage in relationship lending to small firms and that this explained the larger share of SME loans in their portfolios.

More recently, however, researchers have been revisiting the role of relationship lending in the U.S. small business credit market. New empirical evidence suggests that the importance of relationship lending may be diminishing, due in part to the adoption of new lending technologies such as small business credit scoring (SBCS). Petersen and Rajan (2002) revisited their earlier work on the relationship between a firm’s distance to its bank and access to credit and found distances increasing. They suggest that expanding access to credit information from credit reports and the adoption of technologies based on this information, such as credit scoring tools, are behind these findings. Dell’Ariccia and Marquez (2003?), Hauswald and Marquez (2002) and Brevoort and Hannan (2004) further analyze the evolving nature of relationship lending in the U.S. small business credit market and find that lending markets are becoming segmented with large, nationally active lenders employing new technologies such as SBCS to evaluate applicants while local lenders continue to rely on relationship lending.

There is an abundant technical literature on credit scoring technologies, much of it focusing on the relative merits of different statistical techniques and approaches. The key
innovation which made possible the application of credit scoring to the small business credit market was the realization that data on the business owner could be used as a valuable input to these models. Mester (1997) discusses how data on business owners such as their monthly income, outstanding debt, financial assets, home ownership and previous payment history can all be used in small business credit scoring to improve the predictive power of the models. This innovation was pioneered and commercialized by Fair Isaac Co. which led to the widespread adoption of SBCS in the U.S. market.

In an attempt to better understand the role of SBCS in U.S. banks, the Federal Reserve Bank of Atlanta surveyed 200 banks in January 1998, of which 99 responded. The data collected in this survey indicates that larger banks are more likely to adopt SBCS: see Akhavein, Frame and White, (2001). Of particular interest are findings by Frame, Srinivasan and Woosley (2001) based on this survey data that SBCS is associated with an increase in lending volumes to small businesses of 8.4% which translates to an average of $4 billion dollars in additional lending per financial institution. Berger, Frame and Miller (2002) build on these results and find that the increased lending volumes associated with adoption of SBCS serve to increase access to credit by marginal or riskier borrowers who would otherwise be rationed.

Unfortunately, in developing country markets where credit rationing of small business borrowers is acute there have been only limited inroads of SBCS technologies and virtually no development of the more economical pooled data models. One reason may be the high concentration levels in many developing country banking sectors and the
reluctance of dominant lenders to lose their information advantage. Recent research by Beck, Demirguc-Kunt and Maksimovic (2004) shows that credit reporting can reduce financing obstacles faced by firms in concentrated financial systems. One could conjecture that decision tools which make credit reporting data more valuable, such as SBCS, would further reduce the market power of large lenders and increase access to finance.

This paper contributes to the literature by analyzing the viability of small business credit scoring, and particularly, SBCS based on scorecards with pooled data from multiple lenders in a developing country market. The findings will provide insights as to the value of SBCS in emerging markets, focusing specifically on two Latin American countries – Colombia and Mexico. The analysis will also address whether pooled data from institutions in different countries can be used to create robust scorecards.

III. International Experiences with Small Business Credit Scoring
Credit scoring technologies help to simplify the credit origination process by using statistical models to estimate probabilities of default for risk classes of borrowers. For each application, a credit scoring model generates a credit score. Portfolio managers benefit by maintaining more control over the risk they’re willing to accept and more closely matching underwriting processes to portfolio objectives. By systematically quantifying, or rank-ordering, the risk of each application, credit scoring speeds the decision process while simultaneously bringing greater accuracy and fairness to each decision.
Generally speaking, most credit scoring models are developed and designed to help credit
grantors predict the outcome of making a loan to a business. The model is composed of
several questions (characteristics) about the applicant. Different answers (attributes) are
rated on a point system and assigned score weights. An applicant’s score is the sum of all
of his or her attribute points—the higher the score, the lower the risk. If the score is equal
to or higher than the score an organization has established as the “cutoff,” the applicant
presents an acceptable level of risk and the institution may decide to extend credit to that
applicant. In an automated system, scoring takes place instantaneously, allowing lenders
to assess risk and make account origination decisions more quickly, accurately, and
objectively.

The method of assigning the characteristics’ weights and point assignments is designed
during model development. For instance, analysts will take data—such as the number of
delinquencies an applicant has on record from past loans—and evaluate its predictive
power against many other characteristics. There may be several characteristics evaluated,
but the challenge is to find the unique combination of 8-15 characteristics which can be
combined to best determine risk. For example, if the characteristics “number of times
delinquent with other lenders” and “number of times delinquent with (the lender in
question)” are both predictive of risk, they could potentially define the same risky
individuals and may not both be used in a model. Exhibit 1 is an example of what a
credit scoring model typically looks like.
### Exhibit 1: Example Model

<table>
<thead>
<tr>
<th>Time Since Most Recent Delinquency (in months)</th>
<th>Value</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-11</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>12-35</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>36-59</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>60+</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Years in Business</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-2</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>3-5</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>&gt;5</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

This score is expressed as a 3-digit number—the lower the number, the riskier the applicant.

In 1995, Fair Isaac Corporation\(^2\) launched the Small Business Scoring Solution (SBSS) – the first widely used SME scoring tool in the U.S. market. Within a few years virtually all of the top 25 SME lenders (banks) had adopted this tool. With SBSS, Fair Isaac pioneered the concept of using consumer information about the principal owners of a small business, combined with data about the business itself, including financial and application data, to produce robust, highly predictive models of small business risk. Since the behavior of the business owner in his or her personal finances is correlated with the behavior of the firm and ample information is available from U.S. credit bureaus on consumers, this approach was feasible and added significantly to the predictive power of SBCS models. While Fair Isaac continues to be the market leader in SBCS, other firms

---

\(^2\) The Fair Isaac Corporation, headquarters in San Rafael, California, is the leading credit scoring firm in the United States and internationally. The Fair Isaac credit score developed for the leading U.S. credit bureaus, known as the FICO score, is the industry standard. Fair Isaac Corporation is active in ___ countries around the world.
also now offer small business credit scoring tools, including some of the large international credit bureau firms like Experian.

The concept of combining data on business owners with firm data in SBCS has been applied to models developed for many developed and emerging markets. For example, Fair Isaac has applied this technique in models developed for clients in Mexico, the Caribbean, Canada, China, Chile, Brazil, Columbia, Belgium, Nicaragua, Panama, El Salvador, Guatemala, Costa Rica, Indonesia, UK, Belgium, Germany, Italy, Spain, Portugal and Japan. These models fall into one of three categories:

1. **Expert scoring models.** Expert models are developed judgmentally for a specific customer situation, based on both the software developer’s experience and the requirements of the lender. Objective, experienced developers are needed to ensure that the models do not merely reproduce and codify the biases of the lender. Although expert scoring models are tailored specifically for a client, these models are not developed empirically; therefore, no predetermined odds-to-score relationship exists. This is the most common approach used in the international market.

   Expert models are the preferred choice for lenders with the following characteristics:

   - Inadequate or inaccessible historical or performance data
   - The need for an economical alternative to a custom scoring model development
   - Insufficient volume to support a custom scoring model development
2. **Custom application models.** Custom models are empirically derived statistically based tools, custom-developed by Fair Isaac from the client’s data. Scores from empirical models represent the odds that an applicant will pay as promised on a loan. For example, a certain score may represent 60 to 1 odds, meaning that if 61 applicants receive this score, 60 would be expected to pay as promised and 1 would not.

3. **Pooled data scoring models.** These models are empirically derived from a database of pooled creditor data. This option may pose initial difficulties, as data must be collected on a common basis, but will result in more powerful and accurate models.

For the development of a credit scoring model, information is needed on both accepts and rejects and it is difficult to get an adequate number of rejected applications, as well as adequate numbers of loans which are non-performing, when working with only one institution – this is the primary motivation behind pooled data models. In many international markets, credit bureau data is also limited, as it frequently contains only “negative data”—records of delinquency or default—and not positive payment data. To develop a powerful scoring system requires a large data sample with approximately 800-1200 “bad” loans and an equal or greater number of “good loans” and declined applicants. Since many institutions do not have enough data (especially bad SME loans and declined applicant information) for a custom model development, the pooled data, consortium approach can provide a workable solution to the problem of data availability.
Also, by developing models with data from multiple lenders, the resulting product should be more robust (not reflecting the idiosyncrasies of a single institution) and be cheaper than products developed for one client.

The creation of a pooled dataset for development of SBCS products was also an innovation pioneered commercially by Fair Isaac in the Small Business Scoring Solution product line. They brought together 17 unrelated financial institutions in the early 1990s to participate in the product development phase and provide information on both SME loans which they had made and on those loans which they had rejected. Subsequent improvements of the model have increased its predictive power. These improvements include teaming with Dun & Bradstreet to include their trade data in the model and re-engineering the model with data from increasing numbers of financial institutions; 35 lenders participated in the latest revision.

In the United States, 90% of the top small business lenders use Fair Isaac’s pooled data models. In addition, the SBSS 5.0 models that were developed based on United States data have been validated on Canadian small businesses data and the majority of Canadian lenders are using these models. Pooled scoring models exist in only a few other countries. Their development is typically fostered by a lenders’ association or partnership. For example, Fair Isaac has a pooled small business model available for use in Japan. This model was developed using over 3,500 small businesses accounts contributed by 13 participants. This model is highly predictive and incorporates characteristics from the consumer, bureau information, financial ratios and application data. A consortium model
is also currently in development by Fair Isaac for the Hong Kong market. Six of the region’s top lenders are participating in this endeavor.

Despite the benefits and positive market response to off-the-shelf pooled data SBCS products in markets where they are available, this approach is still uncommon, especially in developing countries. There are several reasons for this, beginning with the fact that SBCS is still a fairly new lending technology. Second, credit scoring models require access to local data, in this case on SMEs and their owners.\(^3\) The investment required to develop SME scoring tools may not be justified in smaller economies and credit markets where lending volumes are low.

Larger, emerging market economies however, could conceivably support the development of pooled SME scorecards; this includes countries in Latin America such as Argentina, Brazil, Chile, Colombia and Mexico. Both lenders and policymakers in these countries are interested in promoting small business access to finance, so why has so little been done to automate decisions in this part of the credit market? One reason may be reluctance on the part of lenders to share data on their small business customers, since these borrowers typically pay relatively high interest rates and represent a profitable market segment. The relatively high levels of concentration in some Latin American banking systems\(^4\) further exacerbate this problem, as market-leading financial institutions

---

\(^3\) The importance of adapting models to the local experiences cannot be overstated. For example, in the early to mid 1990s a European firm sold an off-the-shelf SME scoring product in Latin America which had not been reengineered with local data and thus had limited predictive power. This experience soured many bankers in the region on the usefulness of SME scoring solutions.

\(^4\) For example, based on data from Bankscope from 1995 – 1998, the average combined market share of the three largest banks were as follows in select Latin American countries: Argentina – 34%; Brazil – 40%; Chile – 46%; Colombia – 32%; and Mexico – 64%.
jealously guard their “information advantage” from their large customer base. A few large lenders in Latin America have developed proprietary SME scorecards based on their own small business portfolios. However, large banks may see participation in the development of a pooled small business scorecard as inimical to their self-interest. Generic SME scorecards based on pooled data serve to level the playing field between different financial institutions while they also empower borrowers by providing a way to communicate their quality in the financial market. Such decision tools make it easier for a wide array of lenders to profitably lend to SMEs, potentially strengthening the competition that dominant lenders face. This is in contrast to the situation in much less concentrated financial markets, such as the United States, where SBCS and most other new lending technologies are quickly available to a large number of lenders and thus do not represent a tool for perpetuating or increasing market dominance.

Even so, there are many reasons why large lenders in emerging markets could be interested in participating in the development of pooled data SBCS. Large institutions might decide to participate in a pooled model if this provides them the opportunity to improve their knowledge of the market especially if they want to go into sectors that they traditionally do not venture into (i.e. micro lending). They might use their participation as a way of getting more insight into a different modeling technique, a different set of characteristics, new sources of data, and an expanded segmentation scheme that would only be possible when a large amount of data is available. In the event that a pooled model is required to securitize loans, large lenders would also want to influence the model used by embedding in it their portfolio experience.
IV. Research Methodology and Project Design

The approach to be used here is to gather data on common characteristics used and tracked by multiple banks, develop a generic scorecard, and then validate against each lender’s portfolio to determine feasibility of a pooled model approach. We decided on this approach because full data gathering for a statistically valid feasibility study across three countries would run into a high six-figure number which was in excess of our resources. If the empirical tests are promising we would hope that private banks in the participating countries would consider further investments in credit scoring technologies for the SME sector, including possible pooled-data, consortium approaches.

The initial project design focused on small business lending in three countries: Brazil, Colombia and Mexico. Unfortunately, outreach to the banking sector in Brazil was unsuccessful, due at least in part to concerns there with sharing data because of the country’s bank secrecy laws. Financial institutions in Colombia and Mexico have agreed to participate in this project, and are in the process of preparing their data for analysis. Part of this preparation required them to provide information about their SME portfolios.

There is no uniform definition of what constitutes a small business loan in our survey sample. Table 1 below compares the definitions provided by participating institutions in each market. Although institutions in both countries considered small businesses to include sole proprietorships, partnerships and corporations, there were significant differences with regard to firm size. In Colombia firm sizes were relatively larger than in
Mexico, but interestingly the loan size that lenders used to define small business lending – US $200,000 – was the same.

<table>
<thead>
<tr>
<th></th>
<th>Colombia</th>
<th>Mexico</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Number of Employees</td>
<td>200</td>
<td>100</td>
</tr>
<tr>
<td>Maximum Annual Sales</td>
<td>US $1 million</td>
<td>US $500,000</td>
</tr>
<tr>
<td>Maximum Loan Size</td>
<td>US $200,000</td>
<td>US $200,000</td>
</tr>
</tbody>
</table>

In the sector distribution of SME portfolios between the Colombian and Mexican institutions, the differences are less significant. In both cases wholesale businesses dominate the sample and other important sectors are services, construction, manufactures and transportation. The main difference concerns the virtual lack of lending to financial or insurance businesses in Mexico, which in the Colombian sample represent 11% of the portfolio.
Acceptance and booked rates for the banks in our sample also are similar and provide
evidence that bank credit is available for SMEs in Colombia and Mexico. In both
countries almost 75% of the small businesses applicants are accepted and almost 85% of
them are booked. SME loans in Colombia, however, are more likely to go bad. As can be
seen in Table 3 below, Colombia has a higher level of delinquency, with a rate of 2.7%
after one cycle compared to only 1.2% in Mexico. The samples for both countries,
however, do not demonstrate a normal distribution - where delinquency rates would fall
from one to two cycles and from two cycles to being written off. This is most probably
due to the delinquency buckets not being cumulative; there is no requirement that an
account that is currently in written off status also be considered one cycle and two cycles
at that point in time.

![Table 3: Distribution of Delinquencies](chart)

The processing time for SME loans in Colombia is also significantly greater than in
Mexico. Participating institutions in Colombia report that it takes between five and 25
days to process a loan application compared with a range of only two to four days for the
Mexican institutions. In Colombia small business credit scoring is not yet used by
participants whereas in Mexico some of the participants are using custom risk models.
It may be that not only does this impact processing times, but also contributes to the higher delinquency rates in Colombia compared with Mexico, even though institutions in both countries have similar rates of loan acceptance and loan booking.

In general, the basic portfolio information collected from the Colombian and Mexican institutions thus far suggests that the individual loan data, when it is received, will permit the planned analysis of the feasibility of pooled SME scorecards. The next steps in this project are described below.

Participating financial institutions are currently preparing demographic and account performance data which will be used to study the feasibility of creating predictive risk models for small business in Latin America. The requested data includes:

1. Data from the application file: demographic data;
2. Data from master file: account performance (delinquency and payment history); and
3. Data on applicants which were rejected, booked (approved and became customers), not booked (approved, but did not become customers).

The window of time from which the documentation that matches the sample accounts is pulled will go back 24 months.

The business owner’s or principal’s information is a strong predictor of the potential risk of a small business. If the principal is unable to manage his/her personal obligations it is very likely that he/she will not be able to manage the company’s obligations. The
application data (demographic data), payment history data (master file) and credit bureau data are related to the principal(s). The data from the principal(s) is relevant to evaluate the credit risk of the small business, however, when the company is bigger the data from the principal(s) loses predictive power to quantify the risk. In general, if the company is small data from principal is relevant, if the company is large then firm data is recommended.

The analysis of detailed loan data from SME portfolios in this project will provide insights as to the viability of SBCS in Latin America, with a specific focus on models developed using pooled data. In addition, the research will identify predictive characteristics for SME loans in each country as well as common characteristics. Participating lenders will receive limited consulting on their individual SME portfolios from Fair Isaac Co. as an additional benefit for their contribution to this research.

IV. Empirical Results (forthcoming)

V. Conclusions (forthcoming)
References


