Selective knowledge: Reporting biases in microfinance data

Jonathan Bauchet & Jonathan Morduch

June 8, 2009

Selective knowledge: Reporting biases in microfinance data

Jonathan Bauchet
New York University

Jonathan Morduch
New York University

June 8, 2009

Abstract

Answering surveys is usually voluntary, yet much of our knowledge depends on the willingness of households and institutions to answer. We explore the implications of voluntary reporting on knowledge about microfinance. We show systematic biases in microfinance institutions' choices about which survey to respond to and which specific indicators to report. The analysis focuses on data for 2,072 microfinance institutions from MixMarket and the Microcredit Summit Campaign databases for the years 2004-2006. In general, we find that financial indicators are more often reported than social indicators. The patterns of reporting correlate with the institutions' region of operation, mission and size. The patterns in turn affect analyses of key questions on trade-offs between financial and social goals in microfinance. For example, the relationship between operational self-sufficiency and the percentage of women borrowers is positive in the Microcredit Summit Campaign data but negative in the MixMarket data. The results highlight the conditional nature of our knowledge and the value of supporting social reporting.

1. Introduction

The creation of economic knowledge rests on developing and testing theories. To complete those tests, researchers need data, often using whatever relevant information they can find. Yet data sets often only capture a particular slice of the underlying populations. Participation in firm-level and household-level surveys is generally voluntary, and, even among those who participate, some choose to skip questions entirely. There is no reason to think that the pattern of opting in and opting out, and of answering and skipping questions, is random. High-income households, for example, are notoriously difficult for surveyors to engage. At the national level, poorer countries tend to have less-developed statistical systems and irregular surveys. Samples, whether at the level of individuals or of countries, are thus seldom fully representative of the underlying populations. Researchers make do with what they can, and they can't say much about what they do not see.

Microfinance data are not immune to these problems. Data on the performance of microfinance institutions are critical to pushing policy initiatives forward, yet, as a relatively new sector, data have been slow to accumulate. Two organizations have been critical in collecting large data sets on institutions, but both have their own foci and priorities; one centers on financial data, the other elevates the cause of social and economic change. We show that while both data projects aspire to global coverage and encourage the broad submission of data, institutions appearing in the two sets differ systematically by geographic location and by their foci on poverty and financial performance. Different answers are thus obtained when asking the same questions with the same variables but with different data sets. For example, we find a

positive relationship between financial performance and serving more women customers in one data set, but a negative relationship in the second.

The two large-scale data collection projects have admirably pushed for transparency and disseminated aggregate information on microfinance. The Microfinance Information Exchange, Inc. (MIX), through its MixMarket internet platform, collects a wide array of data on microfinance institutions, including financial and institutional data, supplemented by a limited amount of social data. The *Microbanking Bulletin* project involves data collection on a subset of institutions from MixMarket, and the statistics are then adjusted to improve comparability and reveal implicit subsidies. On the social side, the Microcredit Summit Campaign (MCS) collects a limited number of indicators, most of them related to social outreach, on a far larger number of microfinance institutions.[1]

Because reporting to any microfinance information database is voluntary, analyses based on these data are vulnerable to self-selection bias, which can be manifested in three ways. First, institutions reporting to any source are likely to be different from institutions not reporting at all. This bias is likely to be the largest in magnitude, yet it is, by definition, the most difficult to measure and mitigate. Second, microfinance institutions self-select into reporting to either one or both of the reporting outlets. Although both MixMarket and the Microcredit Summit Campaign

---

[1]These three sources of data form the backbone of most analyses of microfinance's institutional landscape. Gonzales and Rosenberg (2006), for example, pull together the three sources to examine a sample of about 2,600 unique microfinance institutions. They examine levels and determinants of profitability of the institutions, juggling between data from one or more sources depending on the type of analysis performed and the availability of data in each source. Recently, Cull et al. (2009) investigate the trade-offs between commercialization of microfinance and social outreach, using a proprietary dataset of 346 institutions in the *Microbanking Bulletin* data. While the data are high quality, adjusted by the MIX and complying with international accounting standards, they concede that the data are "skewed toward institutions that have stressed financial objectives and profitability." (p. 6) Tucker and Miles (2004) examine institutions' financial performance by geographic area, comparing data from the *MicroBanking Bulletin* with data reported by commercial banks. They too note that their results might suffer from selection bias because data in the *MicroBanking Bulletin* is self-reported and institutions "who do provide data tend to be more successful." (p. 45)

are open to all institutions and aim to increase the number of reporting institutions, it is up to

each microfinance institution to decide whether to participate and which database to report to. As

a consequence, institutions reporting to one outlet might differ from institutions reporting to

another, or from institutions reporting to more than one outlet. Finally, institutions may report

some indicators but not others, and some years but not others. MixMarket's and the Microcredit

Summit Campaign's interest is to collect as much information as possible, but with varying

incentives for the reporting institutions to do so. This paper examines the latter two sources of

bias.

We use data on 2,072 microfinance institutions for the years 2004-2006. All data are

provided by institutions to the MixMarket online database and/or the Microcredit Summit

Campaign. We show that the two databases attract substantially different institutions, consistent

with the goals of each organization. Institutions reporting to the Microcredit Summit Campaign

are typically larger, more focused on reaching poor borrowers, and are more likely to operate in

South Asia. Institutions reporting to MixMarket are more financially-focused, as indicated by

their higher average operational self-sufficiency ratio (a measure of financial performance), and

are more likely to operate in two broad regions: "Latin American and the Caribbean" and

"Eastern Europe and Central Asia."

The reporting of specific indicators in each database follows similar trends. Notably,

institutions operating in South Asia and institutions that are already sensitive to reaching the poor

are more likely to report on the percentage of poor borrowers that they serve. Microfinance

institutions operating in Latin America and the Caribbean are less likely to report on the

percentage of poor borrowers, but they tend to be more "professional" in the quality of their

financial reporting. The trends in reporting, both inter-database and intra-database, influence

4

analyses of microfinance, leading to different conclusions depending on the data sets used. As illustrated in Figure 1, data from MixMarket, for instance, indicate trade-offs between reaching profitability (as indicated by the operational self-sufficiency index, OSS, which indicates whether institutions are covering all of their operational costs with revenues—given by OSS > 100) and serving a greater share of women borrowers. The data in the Microcredit Summit Campaign instead show no sign of a trade-off. In Latin America and the Caribbean, the MixMarket data depict a statistically significant U-shaped relationship, with the percentage of women borrowers declining up to an OSS ratio of about 130, then increasing for higher OSS ratios. On the contrary, the Microcredit Summit data suggest that the percentage of women borrowers *increases* when the OSS ratio is lower than 120, and *decreases* afterwards, although this relationship is not statistically significant.

The remainder of the paper is organized as follows. Section 2 describes the Microcredit Summit Campaign, MixMarket, and the dataset used in this paper. Section 3 presents the results. Section 4 discusses the results and concludes the paper.

2. Data

Microcredit Summit Data

We use two sources of data on microfinance institutions. Both sources are open to and encourage all microfinance institutions to report to them, regardless of their mission or performance. The first source is the Microcredit Summit Campaign, a high-profile advocacy organization actively seeking to bring together actors involved in microfinance around four core themes: "reaching the poorest, reaching and empowering women, building financially self-sufficient institutions, and ensuring a positive, measureable impact." (Microcredit Summit Campaign 2008) In 1997,

participants in the first Microcredit Summit launched a campaign to reach 100 million poor families by 2005. The goal was later expanded to 175 million poor families by 2015. To report on that goal, the Microcredit Summit Campaign invites all microfinance institutions to report to its database, in large part to maximize the count of borrowers and "poorest borrowers." Institutions typically report to the Microcredit Summit Campaign annually, with spikes in reporting before the periodic summits[2].

The Microcredit Summit Campaign takes two steps to improve the quality of the submitted data. First, a data verification process takes place for about 40 percent of reporting institutions. All reporting institutions are asked to provide a third-party contact person who can confirm the validity of four numbers provided: the total number of borrowers, the number of poorest borrowers, the percentage of borrowers who are women, and the percentage of poorest borrowers who are women. Microcredit Summit staff members reach out to verifiers for institutions reporting more than 2,500 poor clients and either record the lowest number provided by the two sources or exclude the institution from the database if the verifier is unable to confirm the existence of the institution. The use of a third-party verifier helps improve the quality and accuracy of the data, provided that the verifiers are sufficiently knowledgeable about the institution. Second, the Microcredit Summit Campaign carefully examines the list of institutions submitting data and adjusts the numbers to avoid double-counting borrowers when, for instance, a network of institutions and the members of the same network both report.

MixMarket

---

[2] The first Microcredit Summit was held in 1997. Additional Global Microcredit Summits were held in 2002 and 2006, and regional Summits almost every year in between. The complete list is available at http://www.microcreditsummit.org/summit/previous.htm.

The second source of data is the not-for-profit organization Microfinance Information Exchange,

Inc. through its MixMarket web-based information platform. MixMarket was created to (a)

"enable institutions to compare their financial and outreach performance with that of their peers,"

and (b) "attract more public and commercially-oriented investors to microfinance by promoting

financial transparency, accountability, and increased disclosure standards." (MixMarket 2008)

The MixMarket data are provided by the institutions themselves, and are verified in two ways.

First, the financial data of most of the institutions reporting to MixMarket is supported by

Audited Financial Statements or rating reports, which are established by a third party (since

2000, 61 percent have been verified in this way). Second, before publication, MixMarket

analysts reclassify the financial data and accounts according to international norms for

international comparability, and review the information for coherence and consistency.[3] This

verification process applies mostly to financial data.

Only institutions for which microfinance represents 91-100 percent of the activity were

included in the analysis below. This restriction is imposed by MixMarket's Trend Analysis tool

to improve the comparability of results; but it excludes 277 institutions, among them ProMujer

Bolivia, the ProCredit Bank network in Eastern Europe only, SKS Bangladesh, and Fondo

Esperanza in Chile. (The full list of excluded institutions is available upon request and on the

MixMarket website.) Because reporting depends on characteristics of the institutions, this

exclusion introduces a sample selection problem. The type of bias introduced by this restriction

is difficult to predict, however, because institutions for which microfinance represents less than

91 percent of the activity are very diverse. For example, these institutions include for-profit

commercial banks, whose main business is not microfinance, and poverty-focused NGOs who

---

[3] The data from the MixMarket that we use for this paper were downloaded from www.mixmarket.org on June 20,
2008.

7

provide health training and healthcare along with loans. The geographic distribution of these institutions also influences reporting, but is not known.

The combined data set

The data obtained from both the Microcredit Summit Campaign and MixMarket were combined into one institution-level dataset for analysis covering the period 2004-2006, the most recent years for which both reporting organizations provide data. A microfinance institution is counted as "reporting" a particular indicator if it provided data for at least one of these three years. In the analyses, values for the most recent year available are used. For example, for a given institution the number of borrowers might be from 2006, operational self-sufficiency might be from 2004, and gross portfolio might be from 2005. This feature introduces some imprecision but allows us to maximize sample size.[4]

Microfinance institutions from North America and Western Europe were dropped from the dataset in order to increase comparability between regions and between data sources given that most data are from developing countries. The sample of North American and Western European institutions is small in the Microcredit Summit Campaign data (2.7 percent of institutions in that database). The dataset used in the analyses therefore contains 2,072 observations: 1,471 institutions from the Microcredit Summit Campaign and 978 institutions from MixMarket. Out of the 2,072 institutions, 1,094 report only to the Microcredit Summit, 601 report only to MixMarket and 377 report to both.

The MixMarket dataset provides richer detail on the area of its focus. For instance, the MixMarket data provide three indicators of clients' poverty status, compared to one in the

---

[4] Most of our results, and all our main findings, are robust to restricting the data to 2006 only. Tables are available from the authors upon request.

8

Microcredit Summit Campaign's data. However, all three indicators in the MixMarket dataset are much less reported than in the Microcredit Summit Campaign dataset. The strengths of the Microcredit Summit Campaign dataset are the number of organizations reporting to it and the diversity of these organizations. The Microcredit Summit database collects data on many small institutions and a few very large ones, whereas the MixMarket database includes a set of medium-sized to large microfinance institutions.

Definitions of Key Variables

We mainly focus on common indicators in microfinance, but three deserve further definition and clarification. First, the "percentage of poor borrowers" is defined in slightly different but comparable ways in the Microcredit Summit Campaign and MixMarket databases. "Poorest borrowers" in the Microcredit Summit Campaign are defined as families living with less than US$1/day in purchasing-power parity terms *or* families whose income is in the bottom 50 percent of the population living below the national poverty line when they started the program. No indication is provided about which definition is used by which institution. This indicator does not have a direct equivalent in the MixMarket data, which provide detailed information on three, more precisely defined indicators: (i) the percentage of clients in households earning less than US$1/day per household member, (ii) the percentage of clients below the poverty line of US$2/day, and (iii) the percentage of clients in the bottom half of the population below the poverty line of US$2/day. To analyze the reporting of the percentage of poor borrowers in the MixMarket data, we created a new reporting indicator mirroring the one from the Microcredit Summit Campaign database. It is a dummy variable set to one if the institution in MixMarket reports the percentage of clients in households earning less than US$1/day per household

member *or* the percentage of clients in the bottom half of the population below the poverty line of US$2/day. When using the percentage of poor borrowers as a correlate on the right-hand side of regression equations, the percentage of clients in households earning less than US$1/day per household member is used because its is close to the Microcredit Summit Campaign's definition and it is reported by more institutions than the percentage of clients in the bottom half of the population below the poverty line of US$2/day. Results are robust to the use of various definitions.

Second, we analyze below correlates of "professionalism in reporting." The measure of professionalism is the number of diamonds awarded by MixMarket[5]. MixMarket awards diamonds to all institutions that report to it, on a scale of one (worst) to five (best).[6] The first, second, and third diamonds are awarded for disclosing general, outreach and financial information, respectively. The fourth and fifth diamonds are awarded to institutions that submit adjusted financial data such as audit and benchmarking reports. Diamonds are cumulative; in other words, an institution that received four diamonds disclosed general, outreach and financial information, plus they also submitted audited financial statements for a minimum of two consecutive years.

For this analysis, we restricted our dataset to the 94 percent of institutions reporting to MixMarket that were awarded three or more diamonds and created a dummy indicator of having four or five diamonds. Because the fourth and fifth diamonds reward the quality of disclosure more than the quantity of indicators disclosed, we consider our dummy variable to be a measure of professionalism in reporting. It should be noted that this indicator rewards reporting financial information but does not take into account social reporting. The fourth diamond is granted for

---

[5] A description of the diamond system is available at http://www.mixmarket.org/en/diamond.system.asp.
[6] One institution in our sample – GLOPEC MICRO FINANCE in Ghana – did not receive a diamond notation.

providing audited *financial* statements, including auditors' opinion and notes, and the fifth

diamond is awarded for submitting ratings, due diligence, and "other benchmarking assessment

reports." While the former may include an evaluation of social performance, they typically do

not.

Finally, we also consider the impact of institution type on reporting. This information is

provided only in the MixMarket database, which categorizes institutions as banks, non-bank

financial institutions, cooperatives/credit unions, not-for-profit (NGO), rural banks, and others.

We recoded rural banks and other types together due to small sample sizes, with the omitted

NGO binary variable serving as the reference category in all regressions.

3. Results

This section is organized into three sub-sections. The first two describe patterns in reporting,

beginning with where institutions report and proceeding to the correlates of which specific

indicators are reported. We then illustrate how non-randomness in reporting can influence

analyses of microfinance by contrasting results on the relationship between increasing outreach

to women and increasing operational self-sufficiency.

Who reports to whom?

Three distinct patterns emerge in microfinance institutions' choice of reporting to the

Microcredit Summit Campaign or to MixMarket. First, the two major reporting databases attract

different kinds of microfinance institutions. Institutions reporting to the Microcredit Summit

Campaign tend to be more socially-focused, as seen by the percentage of women and poor

borrowers served. In contrast, institutions reporting to MixMarket tend to be more financially-focused, as measured by their relatively higher operational self-sufficiency ratios.

Table 1 shows that, conditional on reporting, institutions reporting to the Microcredit Summit Campaign serve a higher percentage of poor borrowers than institutions reporting to MixMarket. The average percentage of poor clients in the Microcredit Summit Campaign data (62 percent) is higher than the average percentage of poor clients in the MixMarket data, regardless of the definition used in MixMarket (29, 37 and 53 percent). Figure 2 further illustrates the database specialization. It suggests that institutions reporting to the Microcredit Summit Campaign also serve more women borrowers on average than institutions reporting to MixMarket (75 percent and 65 percent of borrowers are women, respectively). In contrast to both these figures, the average operational self-sufficiency ratio of institutions reporting to the Microcredit Summit Campaign is 95 percent, against 115 percent for institutions reporting to MixMarket. All of these differences are highly statistically significant (all p-values < 0.001).

Second, the geographical distribution of microfinance institutions is very different in each database, and a Chi-square test of the distribution of regions in each database is statistically significant (p-value < 0.001). While the share of South Asian institution is more than twice as large in the Microcredit Summit Campaign than in MixMarket (32.8 percent and 14.5 percent, respectively), Eastern European and Central Asian institutions are almost exclusively reporting to MixMarket: with more than 88 percent of them reporting exclusively to MixMarket, they represent only 1.5 percent of institutions in the Microcredit Summit Campaign database and 18.4 percent of institutions in the MixMarket database. Latin American and Sub-Saharan African institutions, for their part, are about equally represented in both databases: they constitute about

one-third and one-fourth of all institutions, respectively, in the Microcredit Summit Campaign as well as in MixMarket.

Finally, microfinance institutions reporting to the Microcredit Summit Campaign serve more borrowers, on average, than institutions reporting to MixMarket (92,200 and 59,900 respectively). The median (4,921 and 7440, respectively) and standard deviation (992,800 and 375,800 respectively), however, indicate that the Microcredit Summit Campaign database collects data from many small institutions and a few very big ones while MixMarket attracts mostly medium-sized institutions.

Most of these patterns are robust to confounding factors. Table 2 presents the results of multinomial probit regressions, where the dependent variable is a trichotomy variable indicating whether the institution reports to the Microcredit Summit Campaign only, MixMarket only, or to both. The explanatory variables are all the variables that are present in both databases: measures of microfinance institutions' size, outreach to women and poor borrowers, operational self-sufficiency, and region of operation. The results confirm that poverty-focused institutions are more likely to report to the Microcredit Summit Campaign and less likely to report to MixMarket. Both relationships are strongly statistically significant, although the magnitude of the coefficients remains small. The probability of reporting only to MixMarket, for instance, decreases by 7.7 percentage points (from 10.6 percent to 2.9 percent) when the percentage of poor borrowers increases from 30 to 90 percent (the sample's 25th and 75th percentile), holding all other variables at their mean. In addition to poverty-focused institutions, the Microcredit Summit Campaign attracts institutions with lower levels of operational self-sufficiency and institutions from South Asia. MixMarket, on the other hand, attracts institutions serving fewer

poor borrowers and institutions from Eastern Europe and Central Asia. The latter are about 50 percentage points more likely than South Asian institutions to report to MixMarket.

The multinomial probit analysis also allows us to analyze correlates of reporting to both databases. It shows that institutions that report to both the Microcredit Summit Campaign and MixMarket are statistically significantly different from those reporting to one database only, in three different ways. They are more likely to be bigger institutions, to have achieved a higher level of operational self-sufficiency, and to be from East Asia, Latin America and the Caribbean, or Sub-Saharan Africa. While strongly statistically significant, these "biases" exert a small impact on analyses of microfinance since these data are captured regardless of the source of data used. They do, however, stress the importance of avoiding double-counting institutions when using various data sources.

One characteristic of microfinance institutions stands out in Table 2 for its lack of relationship with the choice of what database to report to. The percentage of women borrowers does not seem to be a source of difference among institutions reporting to one or another database, despite the overall higher average percentage of women borrowers in the Microcredit Summit Campaign than in MixMarket.

In summary, the choice of what database to report to is strongly and systematically related to several characteristics of microfinance institutions. These differences are likely to impact analyses of microfinance because neither database provides a representative view of the global microfinance landscape.

Item non-response

Intra-database reporting, or item non-response, refers to the microfinance institutions' choice of what specific indicators to report. This analysis focuses on the reporting of indicators and not their actual values; dependent variables are dummy variables indicating whether indicators were reported. We examine the reporting of outreach and financial indicators, as well as our measure of professionalism in reporting, in both the Microcredit Summit Campaign data and the MixMarket data. We found important patterns of reporting by microfinance institution focus, region of operation and size, but no strong pattern by type of institution.

Understanding the reporting on poor borrowers

Before summarizing the main significant trends in reporting the percentage of poor borrowers, we succinctly describe levels of reporting in both databases. We showed in the previous section that institutions serving a higher (lower) percentage of poor borrowers are statistically significantly more likely to report to the Microcredit Summit Campaign (MixMarket). This distinction extends to the reporting of the percentage of poor borrowers in each database. Over 92 percent of institutions reporting to the Microcredit Summit Campaign report the percentage of poor borrowers, whereas less than 16 percent of institutions reporting to MixMarket report this indicator and less than 20 percent of the latter report any poverty indicator. These simple averages are not surprising given the focus of each database – reaching the poorest is the number one core theme of the Microcredit Summit Campaign, for instance – but they work in conjunction with the specialization of each database to reinforce the biases that arise by using only one data source when studying the microfinance landscape.

The statistical analysis of the correlates of reporting the percentage of poor borrowers reveals two important trends: institutions are more likely to report this indicator if they are already sensitive to reaching the poor, or if they operate in South Asia. First, data on the percentage of poor borrowers comes from microfinance institutions that set a lower average first loan size. The average first loan size, collected by the Microcredit Summit Campaign only, is different from the average loan balance. It is chosen by microfinance institutions so as to be low enough to attract poor borrowers and high enough to meet transaction costs related to servicing the loans. Table 3 shows that the average first loan size for institutions that report the percentage of poor borrowers is US$278, versus US$741 for institutions that do not report this indicator; the regression results in Table 4 show that this difference is statistically significant when controlling for several of the microfinance institutions' characteristics, although the magnitude of the association is very small. In addition, this finding is reinforced by the negative coefficient on the dummy variable, indicating institutions that do not report this indicator. Microfinance institutions that choose not to report the average first loan size are 20 percentage points less likely to report the percentage of poor borrowers on average, keeping all other variables at their mean, which is approximately equivalent to an increase in the average first loan size from US$100 to US$8,000.

Second, the probability of reporting the percentage of poor borrowers varies importantly by geographical area of operation. Both the Microcredit Summit Campaign and the MixMarket data show that South Asian institutions are statistically significantly more likely than institutions in any other region to report the percentage of poor borrowers. Indeed, Table 4 shows that the probability of reporting this indicator in the Microcredit Summit Campaign database is 10 to 28 percentage points higher for South Asian microfinance institutions than for institutions in other regions. Latin American and Caribbean institutions, for their part, are the least likely to report the

percentage of poor borrowers. In the MixMarket data, for instance, the probability of reporting poor borrowers is almost 15 percentage points lower on average for institutions in Latin America than for institutions in South Asia. This is a very large impact since only 20 percent of all institutions in MixMarket report any poverty indicator.

Finally, several simple differences between reporting and non-reporting institutions are not robust to controlling for microfinance institutions' characteristics. For instance, on average, institutions reporting the percentage of poor borrowers to the Microcredit Summit Campaign serve more borrowers, have a lower operational self-sufficiency ratio, and serve more women than institutions not reporting this indicator (Table 3). None of these characteristics are statistically significant in the regression analysis presented in Table 4. Similarly, the size and type of institutions are not significant correlates of reporting on the poverty of borrowers in the regressions using MixMarket data. The regional dummy variables in the regressions are thus defining much of the broad distinctions.

Correlates of reporting on women borrowers

Two clear trends and two more nuanced trends emerge when analyzing the correlates of reporting on the percentage of women borrowers. The first clear trend is that financially-focused microfinance institutions, as measured by a higher operational self-sufficiency ratio, are more likely to report the percentage of women borrowers. This pattern is not obvious when comparing the average operational self-sufficiency ratios of reporters and non-reporters, especially in the MixMarket data (115 percent and 112 percent, respectively). It is strongly statistically significant, however, when controlling for a set of institution-level characteristics, although the relationship is very weak. The probability of reporting the percentage of women borrowers

increases by 5.9 percentage points (about 62 percent to 68 percent) in the Microcredit Summit Campaign data and by 2.5 percentage points (about 96 percent to 98 percent) in the MixMarket data when the operational self-sufficiency ratio increases from the 25th to the 75th percentile in each dataset, holding all other variables at their mean. One caveat is in order, however. In the Microcredit Summit data, the 40 percent of institutions that do not report operational self-sufficiency are strongly and significantly more likely to report the percentage of women borrowers. Assuming that these institutions have a lower operational self-sufficiency ratio on average, this contradicts the positive relationship between operational self-sufficiency and reporting the percentage of women borrowers that appears both in the Microcredit Summit Campaign and the MixMarket databases.

Second, similar to patterns of reporting on poor borrowers, the reporting on women borrowers follows strong regional patterns. For this indicator, however, the pattern is somewhat unexpected and instructive. The Latin American institutions that opt to report to the Microcredit Summit Campaign are the most likely to report on the percentage of women borrowers (relative to others in the Microcredit Summit Campaign database). The probability that they report on women is 18 percentage points higher than the same probability for South Asian institutions, and 65 percentage points higher than for Eastern European and Central Asian institutions. The relationship in the MixMarket database is the opposite, however, with Latin American institutions' probability of reporting on the percentage of women borrowers second to last, although this finding is not statistically robust. (In the MixMarket data, East Asia is the region that has the lowest probability of reporting the percentage of women borrowers, and the result for Latin America is likely tied to the self-selection of commercially-minded institutions that serve fewer women.)

Another correlate of reporting the percentage of women borrowers deserves careful analysis: the institutions' size. On the one hand, microfinance institutions lending to more clients have a higher probability of reporting the percentage of women borrowers. This finding is obvious in the Microcredit Summit Campaign data, which indicates that a one-percent increase at the mean of the number of borrowers (4,920) is associated with a 0.05 percentage point increase in the probability of reporting the percentage of women borrowers. The MixMarket database provides more indicators of institution size, depicting a subtler picture of the relationship. By including indicators of the number of borrowers, the loan portfolio and the average loan size, we are able to look at the influence of the number of loans and the size of each loan. The statistical significance and sign of these three coefficients suggest that institutions making fewer, bigger loans are more likely to report the percentage of women borrowers. Column (3) of Table 5 uses a specification similar to that used for the Microcredit Summit Campaign regression and fails to find a significant relationship between the number of borrowers and the probability of reporting the percentage of women borrowers. The sign of the coefficient, however, is consistent with the coefficient from the Microcredit Summit Campaign data.

Finally, the percentage of women borrowers is one of the few indicators in the MixMarket data that is reported differently by type of institution. Columns (2) and (3) of Table 5 suggest that not-for-profit microfinance institutions have a higher probability of reporting the percentage of women borrowers than any other type of institution. This finding, however, is not robust to controlling extensively for institutions' size, with the exception that non-bank financial institutions are 5.5 percentage points less likely to report the percentage of women borrowers than not-for-profit institutions, holding all other institutions' characteristics at their mean.

Although many non-bank financial institutions started as not-for-profits, this result suggests that there are significant differences between these two types, at least as far as reporting is concerned.

Reporting financial indicators

The MixMarket data's main strength is the diversity and quantity of financial indicators. We present results for four key indicators only – gross loan portfolio, total assets, average loan balance, and operational self-sufficiency – because all financial indicators follow similar trends. All financial indicators in the MixMarket database are reported by more than 80 percent of institutions. Paradoxically, this strength is also a weakness for our analysis: we cannot use regressions because there is not sufficient variation in the dependent variables. We therefore only present simple descriptive statistics, in the form of graphs, which should of course be taken with all necessary caution. The Microcredit Summit Campaign data, for their part, provide only one financial indicator, operational self-sufficiency, which is reported by about 60 percent of institutions.

Differences in reporting financial indicators in the MixMarket data by microfinance institutions' region of operation and type are summarized in Figure 3. First, the proportion of institutions in Sub-Saharan Africa that are in the MixMarket database and that also report key financial indicators is lower than the proportion of institutions in all other regions. The difference is statistically significant when tested with a Chi-square test. Second, the right panel of Figure 3 suggests that (a) banks report financial indicators more than other types of institutions, and (b) cooperatives and credit unions report less than other types. A Chi-square test, however, reveals that these differences are not statistically significant.

In the Microcredit Summit Campaign database, regression analysis indicates that bigger institutions are statistically significantly more likely to report operational self-sufficiency (Table 6), and the magnitude of the relationship is large. At the mean number of borrowers (4,920) and of all other variables in the model, an increase of 1,000 borrowers is associated with an increase in the probability of reporting the operational self-sufficiency ratio by 2 percentage points. In addition, poverty-focused institutions, as measured by their percentage of poor borrowers, tend to have a higher probability of reporting the OSS ratio. This may arise because institutions with strong social missions fear little by reporting a low OSS ratio, compared with banking-focused institutions. The relationship, however, is not robust to specification changes. Finally, unlike the reporting of all outreach indicators, there does not seem to be any regional trend in the reporting of the financial indicator in the Microcredit Summit Campaign data.

Professionalism in financial reporting

As explained in the data section, the indicators of professionalism in financial reporting are dummy variables equal to one if the microfinance institution received four or five diamonds from MixMarket. The sample is limited to the 94 percent of institutions reporting to MixMarket that received three or more diamonds. Almost 60 percent of institutions in this sample received four or five diamonds, and 28 percent received five diamonds.

The main correlate of professionalism in reporting is, once again, the region of operation, which, as described earlier, is related to the measure's bias in favor of financial information disclosure. South Asian institutions are statistically significantly less "professional" in their financial reporting than institutions in other regions. Table 7 shows that being a South Asian institution is associated with a 14 to 37 percentage point decrease in the probability that an

21

institution has four or five diamonds, as opposed to three diamonds only, keeping all other variables at their mean. This impact is large, representing 23 to 61 percent of the mean of the dependent variable. Latin American institutions are the most "professional," even after controlling for operational self-sufficiency, measures of poverty and gender outreach, and institution size. The cause of these patterns is unclear. On one hand, the reporting standards used by many of the South Asian institutions may be different enough from MixMarket's standards that conforming to both standards is excessively costly. On the other hand, the generally pro-poor nature of microfinance institutions in South Asia could mean that they are either less focused on passing financial audits or simply lack the budget to allocate to audits.

Return on assets, a measure of financial focus, is also negatively and significantly associated with microfinance institutions' professionalism in reporting. The impact of ROA is small, but statistically significant. An increase from the 25th percentile to the 75th percentile of return on assets is associated with a decrease in the probability of having four or five diamonds from 79 percent to 76 percent. This finding runs against the global picture of financially well performing microfinance institutions reaping the benefits of transparency, and remains a puzzle. Interaction variables indicated that this result is not driven by a particular region or type of institution. Cooperatives and credit unions are also statistically significantly less likely than other types of institutions to receive 4 or 5 diamonds, but the meaning of this relationship is unclear.

In summary, the results presented so far have established that non-reporting is not random. On the one hand, an inter-database bias is introduced by the significant specialization of each database. On the other hand, intra-database biases exist due to statistically significant patterns of reporting specific indicators. We illustrate in the next sub-section that these biases

have a significant impact on analyses of the microfinance landscape, in the context of the relationship between social outreach and financial performance.

Financial sustainability versus outreach to women

We use the relationship between operational self-sufficiency and percentage of women borrowers to illustrate how reporting biases influence final analyses, because this relationship is part of the current debate between commercialization and social outreach in microfinance. We are able to explore it since these two indicators are available and reported by enough institutions in the Microcredit Summit Campaign and the MixMarket databases. As shown in Figure 1, the MixMarket data indicate that reaching sustainability and serving women borrowers is incompatible, whereas the Microcredit Summit Campaign data suggest that both can be achieved simultaneously. We do not intend to establish any causality between these two indicators, but simply to highlight statistical associations, and how they vary across database.

This relationship is particularly interesting among Latin American and Caribbean microfinance institutions. As illustrated in the right panel of Figure 1, the MixMarket data show a strong, U-shaped relationship between operational self-sufficiency and percentage of women borrowers in Latin America. At low levels of operational self-sufficiency, Latin American and Caribbean institutions reporting to the MixMarket database lend to a smaller percentage of women, but they are able to reverse the trend once they become fully sustainable. The turning point is at the operational self-sufficiency ratio of 130 percent. In the Microcredit Summit Campaign data, however, the relationship is the inverse, following an inverted-U shape, though it is much smaller in magnitude and not statistically significant. In addition, the relationship between operational self-sufficiency and the percentage of women borrowers does not hold

23

statistically in other regions. This suggests that it is driven by the financial-performance orientation of Latin American and Caribbean institutions and institutions reporting to MixMarket.

In summary, institutions reporting to the Microcredit Summit Campaign seem to be able to reconcile sustainability and social outreach, but the MixMarket data present the opposite picture, especially among Latin American and Caribbean institutions. The database, and the choice to report, matter, even though both surveys attempt to pull together a wide range of global data.

4. Discussion and Conclusion

Leading practitioners and policymakers fought hard in the 1990s to improve financial reporting and transparency in the microfinance sector. Institutions upgraded their accounting procedures and disclosure protocols, and, in the process, many raised their efficiency levels and financial performance as well. The campaign was bold, and its success demonstrates the power of improving transparency. The initiative set off ripples that are now felt widely.

Recently, a second campaign has been underway, this time focused on the social dimensions of development finance. Important work has been completed at the Institute for Development Studies through its Imp-Act initiative, CGAP (the Consultative Group to Assist the Poor), the Ford Foundation, among other sites. The initiatives are set to deliver a stream of new insights and measurement tools over the next five to ten years.

For now, we remain faced with data sets that are not representative – often in ways that are impossible to fully describe. Some institutions do not report at all, some report only to selected databases, and some report some indicators and not others. These patterns introduce

biases in the analyses since, as we show, missing data are nonrandom. While we focus on microfinance data, these kinds of selective data problems emerge across social science research.

A simple analysis of the relationship between a proxy for profitability (the operational self-sufficiency ratio) and the percentage of women borrowers reveals the potential impact of the reporting biases. Data from MixMarket indicate that reaching sustainability and serving women borrowers are incompatible, whereas the data collected by the Microcredit Summit Campaign suggest that both can be achieved simultaneously. The conclusions diverge even more for institutions in Latin American and the Caribbean. While the MixMarket data depict a statistically significant U-shaped relationship, the Microcredit Summit data suggest that the percentage of women borrowers first *increases* when the OSS ratio is lower than 120, then *decreases* afterwards, although this relationship is not statistically significant.

Three patterns emerge. First, poverty indicators are particularly under-reported: the evidence indicates that these indicators are most likely to be reported by South Asian institutions and by institutions that appear sensitive to reaching the poor. Institutions reaching more poor borrowers are significantly less likely to report to MixMarket. Second, although we categorize the percentage of women borrowers as a measure of outreach, its reporting patterns are more similar to those of financial indicators. Both the MixMarket and the Microcredit Summit Campaign databases show that institutions that have achieved a higher operational self-sufficiency ratio are significantly more likely to report this indicator. Finally, and more surprisingly, institution type does not statistically significantly affect the reporting of most indicators. Banks, for instance, are not more professional in their reporting, as measured by MixMarket's diamonds ranking of reporting quality, and while they are more likely to report financial indicators than other institution types, the difference is not statistically significant.

These results indicate that neither of the two data sources analyzed here provides a comprehensive view of the state of microfinance when analyzed in isolation. Analysts need to take caution when making inferences and conclusions about the general state of microfinance and the needs of microfinance institutions. This advice also applies to the present study given that the present analysis could not take into account the many institutions that do not report to either MixMarket or the Microcredit Summit Campaign. We have analyzed differences in patterns found in these two important data sets, which between them include all leading microfinance institutions. But we have no way to compare the evidence to that for the actual, full universe of institutions.

Overall, the fact that the reporting of financial indicators presents fewer significant trends than the reporting of outreach indicators suggests that the global push for transparency and financial sustainability has been successful at increasing the amount and quality of financial information disclosed by institutions, particularly to MixMarket. This result suggests that increasing the emphasis on social reporting could similarly help institutions increase social outreach. Social data are often more difficult to collect than financial data, partly because social measures require that institutions collect data beyond those recorded in typical administrative and financial records. But poverty measurement tools have been devised and are already used by many institutions, impact measurement methodologies have made rapid progress in validity and simplicity (although sometimes with a trade-off between the two features), and information systems are being developed to relatively easily report basic outreach indicators. These tools can generate data that both the Microcredit Summit Campaign and MixMarket could collect and make available.

Donors and policymakers have the power to support institutions in using these tools, and to help data collection organizations harmonize their data. Donors and researchers can also join to encourage non-reporting institutions to begin reporting. Our results suggest that doing so would enrich--and likely re-shape--understandings of the current microfinance landscape.

References

Gonzalez, Adrian and Richard Rosenberg. 2006. "The State of Microfinance—Outreach, Profitability, and Poverty (Findings from a database of 2600 microfinance institutions)." Presentation at World Bank Conference on Access to Finance. May 30, Washington, D.C.

Cull, Robert, Asli Demirgüç-Kunt and Jonathan Morduch. 2009. "Microfinance Meets the Market." *Journal of Economic Perspectives*, 23(1): 167-92.

Microcredit Summit Campaign. 2008. "About the Microcredit Summit Campaign" website page. Retrieved June 20, 2008 (http://www.microcreditsummit.org/aboutmicrocreditsummit.htm).

MixMarket. 2008. "About MIX MARKET" website page. Retrieved June 20, 2008 (http://www.mixmarket.org/en/about_mix_market.asp).

Tucker, Michael and Gerard Miles. 2004. "Financial Performance of Microfinance Institutions. A Comparison to Performance of Regional Commercial Banks by Geographic Region." *Journal of Microfinance*, 6(1): 41-54.

Figure 1. Relationship between operational self-sufficiency ratio and
percentage of women borrowers.
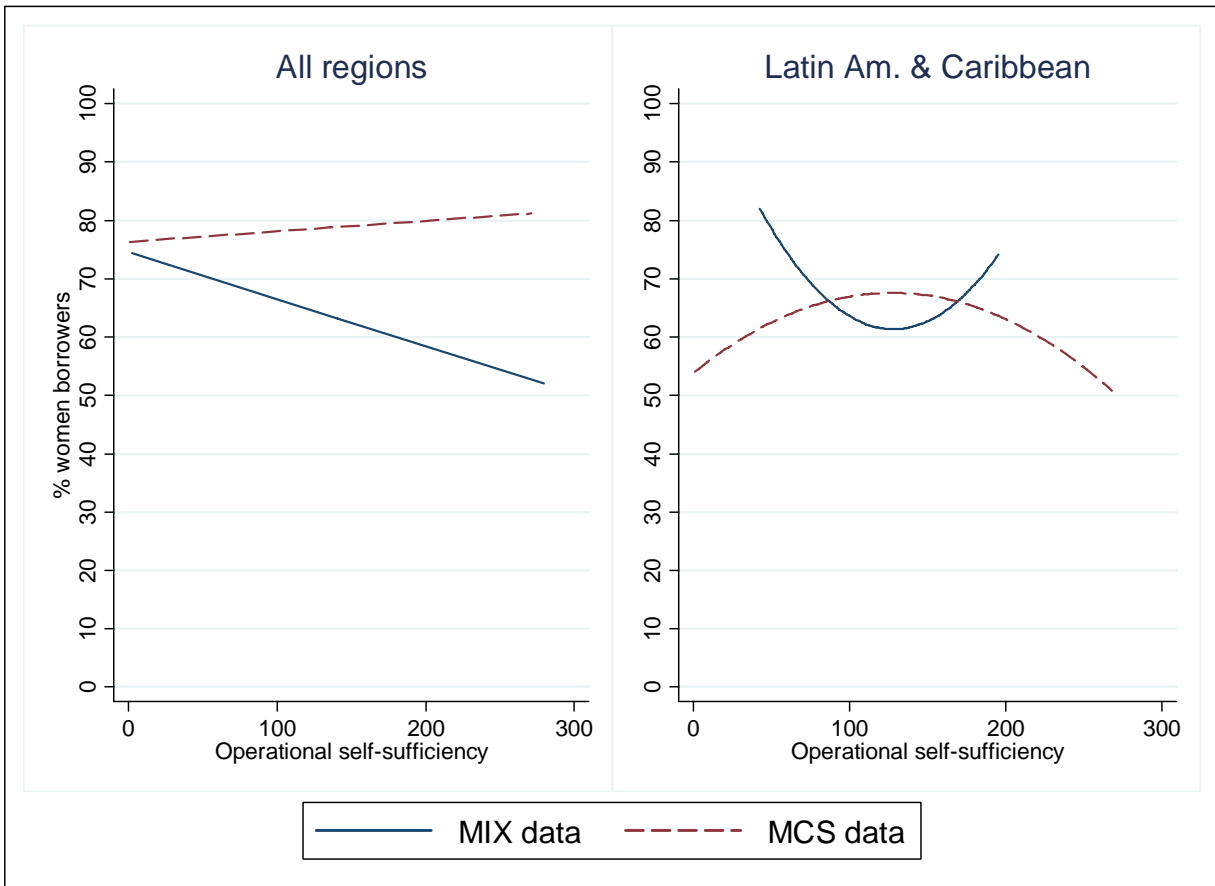
Figure 2. Differences between Microcredit Summit Campaign and MixMarket databases.
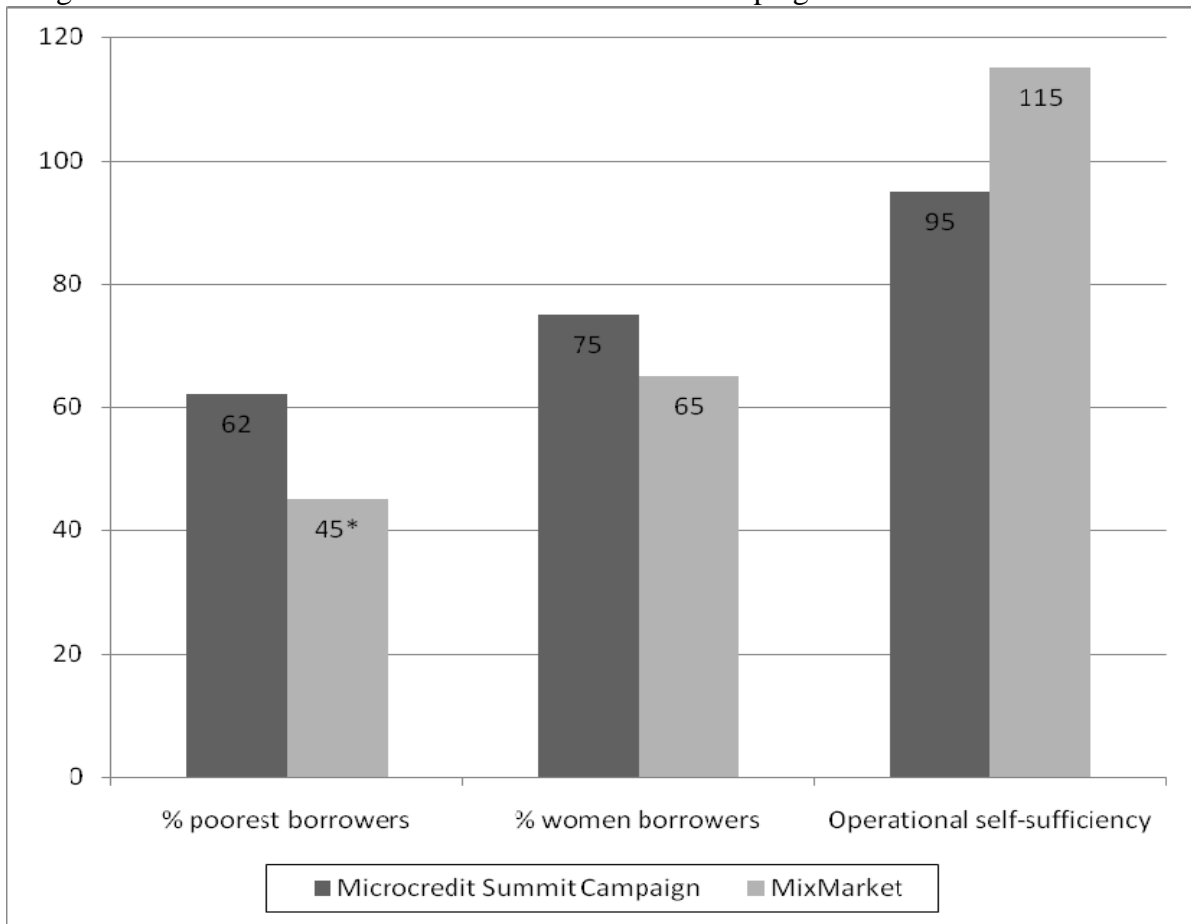
* This value comes from averaging the three poverty indicators provided in the MixMarket data (values: 29, 37 and 53 percent); it is a composite index that does not have a concrete interpretation but allows for a direct comparison between MixMarket and the Microcredit Summit Campaign.

Figure 3. Reporting of selected financial indicators in the MixMarket data.
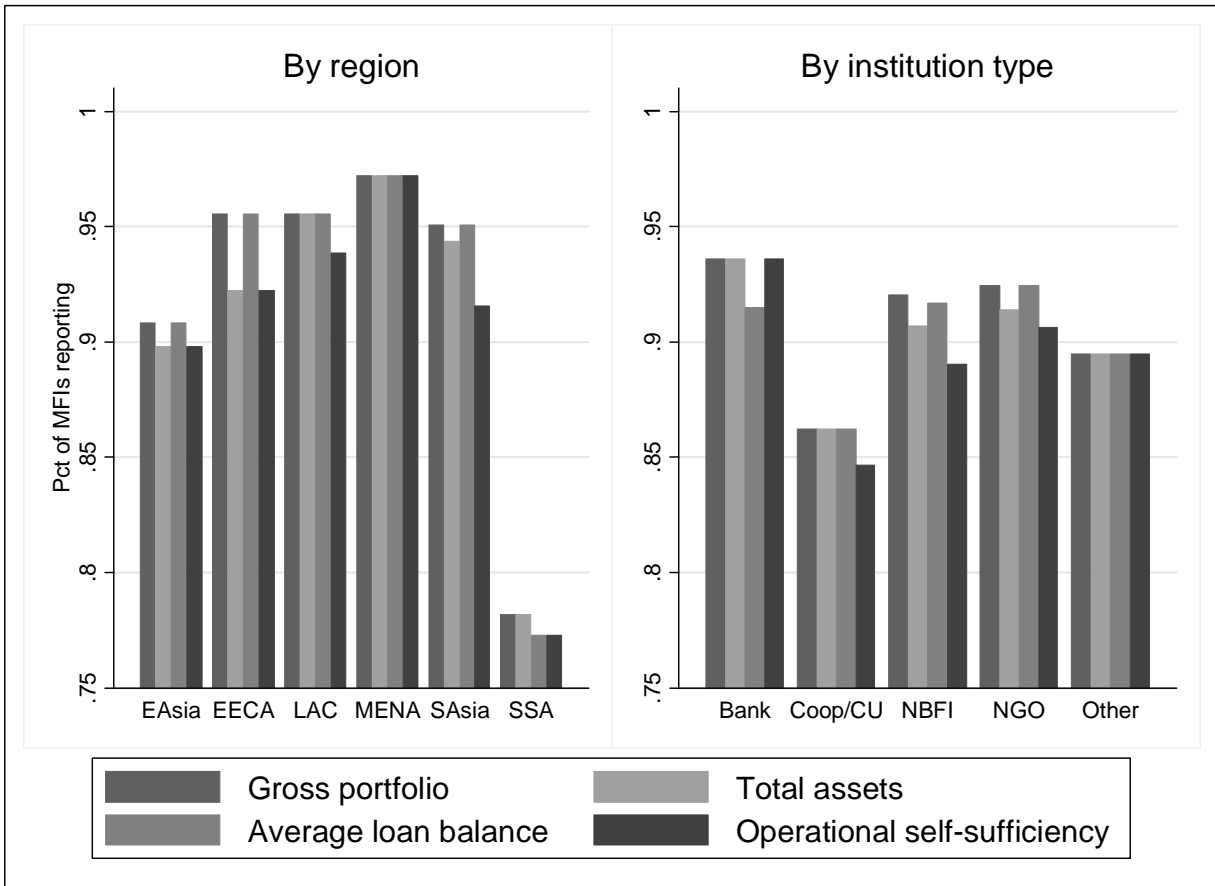
Region abbreviations: E Asia: East Asia; EECA: Eastern Europe and Central Asia; LAC: Latin America and Caribbean; MENA: Middle East and North Africa; S Asia: South Asia; SSA: Sub-Saharan Africa.
Institution types abbreviations: Coop/CU: Cooperatives/Credit Unions; NBFI: Non-Bank Financial Institutions; NGO: Non-Government Organization.

Table 1. Summary statistics.

|	% of institutions reporting	Number of observations	Conditional mean	Conditional std. dev.
MixMarket data (978 institutions)				
Scale indicators				
Gross loan portfolio in million US$	91.0	890	23.2	133.2
Number of borrowers	92.4	904	59,863	375,768
Outreach indicators				
% clients in bottom half of the population below poverty line (US$2/day)	13.3	130	28.9	31.5
% clients living with less than US$1/day	14.6	143	37.0	35.8
% clients below poverty line (US$2/day)	17.7	173	52.5	36.0
% poorest borrowers[a]	15.7	n/a	n/a	n/a
% women borrowers	85.3	834	65.5	26.0
Selected financial indicators				
Return on assets	84.5	826	7.2	10.9
Operational self-sufficiency ratio	89.1	871	114.9	37.8
Average loan balance in US$	90.8	888	953	1,785
Other indicators				
Number of diamonds	n/a	977	3.8	0.96
Region (%) East Asia	100	978	10	n/a
Eastern Europe & Central Asia	100	978	18.4	n/a
Latin America and Caribbean	100	978	30	n/a
Middle-East and North Africa	100	978	3.7	n/a
South Asia	100	978	14.5	n/a
Sub-Saharan Africa	100	978	23.4	n/a
Type (%) Bank	100	978	4.8	n/a
Cooperative / Credit Union	100	978	19.3	n/a
Non-Bank Financial Institution	100	978	30.8	n/a
Non-Governmental Organization	100	978	39.3	n/a
Other / Rural Bank	100	978	5.8	n/a
Microcredit Summit Campaign data (1,471 institutions)				
Scale indicator				
Number of borrowers	97.0	1,427	92,219	992,888
Outreach indicators				
% poorest borrowers	92.1	1,355	62	34
Average first loan amount	94.6	1,392	298	774
% women borrowers	57.1	840	75.1	23.2
Financial indicator				
Operational self-sufficiency ratio	60.6	891	95.2	53.5
Other indicator				
Region (%) East Asia	100	1,471	7.7	n/a
Eastern Europe & Central Asia	100	1,471	1.5	n/a
Latin America and Caribbean	100	1,471	30.9	n/a
Middle-East and North Africa	100	1,471	2.4	n/a
South Asia	100	1,471	32.8	n/a
Sub-Saharan Africa	100	1,471	24.7	n/a

Means and standard deviations are conditional on reporting.
[a] This indicator is not provided by MixMarket but was created to be comparable to the same indicator in the Microcredit Summit Campaign. The dummy variable is set to one if either one of the following two indicators were reported: "report the percentage of clients in households earning less than US$1/day per household member" or "report the percentage of clients in the bottom half of the population below the poverty line of US$2/day."

Table 2. Choice of reporting outlet.

Dependent variable: institution reports to…	MCS only	MIX only	Both
Mean(dependent variable):	0.528	0.290	0.182
Log(number of borrowers)	-0.030**	-0.029***	0.059***
	(0.013)	(0.006)	(0.013)
% women borrowers	-0.001	0.001	0.001
	(0.001)	(0.000)	(0.001)
% poor borrowers	0.002**	-0.001***	-0.000
	(0.001)	(0.000)	(0.001)
Operational self-sufficiency	-0.003***	-0.0001	0.003***
	(0.001)	(0.0003)	(0.001)
East Asia	-0.252***	0.023	0.229***
	(0.059)	(0.047)	(0.071)
Eastern Europe & Central Asia	-0.491***	0.509***	-0.018
	(0.027)	(0.125)	(0.126)
Lat. America and Caribbean	-0.401***	-0.040	0.441***
	(0.046)	(0.025)	(0.049)
Middle East & North Africa	-0.351***	0.151	0.201*
	(0.059)	(0.097)	(0.106)
Sub-Saharan Africa	-0.404***	0.035	0.369***
	(0.042)	(0.038)	(0.054)
Number of observations		742	

note: 0.01 - ***; 0.05 - **; 0.1 - *;
Coefficients are marginal effects after multinomial probit regressions.

Table 3. Mean values of selected indicators, by reporting status and database.

Microfinance institutions' characteristics	Microcredit Summit data		MixMarket data	
	Report (1)	Do not report (2)	Report (3)	Do not report (4)
Reporting % women borrowers				
Number of borrowers	132,836	34,261	57,066	93,196
Operational self-sufficiency	101.1	84.5	115.1	112.1
% women borrowers	75.1	-	65.5	-
% poorest borrowers[a]	60.3	63.6	37.1	35.1
% of South Asian institutions	35.9	28.8	16.2	4.9
% of Latin American institutions	37.5	22.2	31.5	20.8
% of Eastern European institutions	0.6	2.7	19.7	11.1
Reporting operational self-sufficiency				
Number of borrowers	80,384	111,831	59,545	68,007
Operational self-sufficiency	94.4	-	114.9	-
% women borrowers	78.1	69.7	65.3	86.1
% poorest borrowers[a]	63.5	58.3	37.5	24.4
% of South Asian institutions	40.7	20.8	14.9	11.2
% of Latin American institutions	24.9	40.1	31.6	16.8
% of Eastern European institutions	1.7	1.2	19.1	13.1
Reporting % poorest borrowers				
Number of borrowers	92,544	86,095	54,131	61,040
Operational self-sufficiency	94.2	101.0	115.7	114.7
Average first loan size (US$)	277.8	741.3	n/a	n/a
% women borrowers	75.3	67.2	66.2	65.4
% poorest borrowers[a]	61.6	-	37.0	-
% of South Asian institutions	35.1	6.0	21.4	13.2
% of Latin American institutions	29.4	49.1	13.0	33.1
% of Eastern European institutions	1.2	5.2	33.8	15.5
Was awarded 4 or 5 diamonds by MixMarket				
	Yes	No	Yes	No
Number of borrowers	n/a	n/a	109,907	130,221
Operational self-sufficiency	n/a	n/a	118.1	105.1
% women borrowers	n/a	n/a	69.9	76.7
% poorest borrowers[a]	n/a	n/a	48.9	43.4
% of South Asian institutions	n/a	n/a	12.3	37.0
% of Latin American institutions	n/a	n/a	49.8	27.0
% of Eastern European institutions	n/a	n/a	4.0	2.0

[a] In the MixMarket data, this is the percentage of clients living with less than US$1/day (this table shows actual values and not reporting indicators). The three regional indicators (% of institutions in South Asia, Latin America, and Eastern Europe) do not sum to 100 given that East Asia and "Middle East and North Africa" and Sub-Saharan Africa are omitted).

Table 4. Reporting the percentage of poor borrowers.

Data source:	MCS	MixMarket			
Dependent variable: institution reports…	% poorest borrowers	% poorest borrowers[a]	% clients living with less US$1/day	% clients below poverty line (US$2/day)	% clients in bottom half pop. below pov. line
Mean(dependent variable):	0.921	0.157	0.146	0.177	0.133
Average first loan size ('000 US$)	-0.010***				
	(0.004)				
Did not report average first loan size	-0.206***				
	(0.073)				
Operational self-sufficiency	3.56E-06	0.0003	0.0002	0.0001	0.0001
	(0.00012)	(0.000)	(0.0004)	(0.0004)	(0.0004)
Did not report OSS	-0.010				
	(0.015)				
Log(number of active borrowers)	-0.003	-0.095	-0.072	0.031	-0.448
	(0.002)	(0.266)	(0.257)	(0.280)	(0.341)
% women borrowers	-0.013	0.0001	0.0004	0.0013*	-0.0001
	(0.028)	(0.0007)	(0.0007)	(0.0007)	(0.0006)
Did not report % women borrowers	-0.045				
	(0.029)				
Log(gross loan portfolio in US$)		0.072	0.049	-0.059	0.422
		(0.265)	(0.257)	(0.280)	(0.341)
Log(average loan balance)		-0.100	-0.065	0.031	-0.452
		(0.265)	(0.257)	(0.280)	(0.341)
Portfolio at risk >30 days ratio		0.0004	0.0003	0.0007	-0.0003
		(0.0014)	(0.0014)	(0.0014)	(0.0014)
Return on assets (%)		0.001	0.001	0.0001	0.0001
		(0.001)	(0.001)	(0.0013)	(0.0011)
Bank		0.114	0.069	-0.005	-0.010
		(0.096)	(0.089)	(0.085)	(0.077)
Cooperative/Credit Union		-0.020	-0.033	-0.017	-0.014
		(0.042)	(0.039)	(0.045)	(0.040)
Non-Bank Financial Institution		0.010	0.008	0.018	0.023
		(0.036)	(0.035)	(0.038)	(0.034)
Rural bank & Other		-0.029	-0.046	0.089	-0.013
		(0.059)	(0.053)	(0.079)	(0.057)
East Asia	-0.108*	-0.107***	-0.089**	-0.122***	-0.097***
	(0.055)	(0.034)	(0.035)	(0.035)	(0.030)
Eastern Europe & Central Asia	-0.288**	0.033	0.030	0.028	-0.013
	(0.128)	(0.066)	(0.064)	(0.068)	(0.055)
Lat. America and Caribbean	-0.132***	-0.148***	-0.139***	-0.165***	-0.124***
	(0.032)	(0.039)	(0.038)	(0.041)	(0.036)
Middle East & North Africa	-0.176*	0.029	0.026	0.127	0.037
	(0.097)	(0.071)	(0.069)	(0.087)	(0.068)
Sub-Saharan Africa	-0.057**	-0.069*	-0.058	-0.039	-0.090***
	(0.028)	(0.041)	(0.040)	(0.047)	(0.033)
Number of observations	1,427	761	761	761	761
Pseudo-R2	0.177	0.077	0.075	0.100	0.076

note: 0.01 - ***; 0.05 - **; 0.1 - *;
Omitted region: South Asia. Omitted institution type in MixMarket data: NGO.
Total number of observations in the Microcredit Summit Campaign dataset: 1,471; in the MixMarket dataset: 978.
Coefficients are marginal effects after probit regressions.
[a] This indicator is not provided by MixMarket but was created to be comparable to the same indicator in the Microcredit Summit Campaign. The dummy variable is set to one if at least one of the following two indicators were reported: "report the percentage of clients in households earning less than US$1/day per household member" and "report the percentage of clients in the bottom half of the population below the poverty line of US$2/day."

Table 5. Reporting the percentage of women borrowers.

Dependent variable:	institution reports the % of women borrowers		
	MCS data	MixMarket data	
Mean(dependent variable):	.571	.853	
	(1)	(2)	(3)
Average first loan size ('000 US$)	0.031		
	(0.021)		
Operational self-sufficiency	0.0011***	0.0008***	0.0009***
	(0.0004)	(0.0002)	(0.0002)
Did not report OSS	0.152***		
	(0.043)		
Log(number of active borrowers)	0.052***	-0.042***	0.002
	(0.008)	(0.014)	(0.003)
% poorest borrowers	0.042		
	(0.049)		
Log(gross loan portfolio in US$)		0.041***	
		(0.013)	
Log(average loan balance)		-0.057***	
		(0.018)	
Portfolio at risk >30 days ratio		-0.000	-0.000
		(0.001)	(0.001)
Return on assets (%)		-0.000	0.001
		(0.001)	(0.001)
% clients living with less US$1/day		-3.19E-05	3.84E-04
		(0.001)	(0.001)
Did not report % clients living with less US$1/day		-0.027	-0.010
		(0.016)	(0.020)
Bank		-0.084	-0.173**
		(0.063)	(0.079)
Cooperative/Credit Union		-0.043	-0.075**
		(0.029)	(0.033)
Non-Bank Financial Institution		-0.055**	-0.077***
		(0.024)	(0.026)
Rural bank & Other		-0.088	-0.154**
		(0.065)	(0.077)
East Asia	0.024	-0.476**	-0.506***
	(0.054)	(0.186)	(0.161)
Eastern Europe & Central Asia	-0.473***	-0.089	-0.136
	(0.086)	(0.091)	(0.095)
Lat. America and Caribbean	0.179***	-0.105	-0.164**
	(0.035)	(0.074)	(0.076)
Middle East & North Africa	0.014		
	(0.093)		
Sub-Saharan Africa	-0.221***	-0.134	-0.148
	(0.038)	(0.101)	(0.091)
Number of observations	1,332	856	857
Pseudo-R2	0.112	0.312	0.250

note: 0.01 - ***; 0.05 - **; 0.1 - *;
Omitted region: South Asia. Omitted institution type in MixMarket data: NGO.
Total number of observations in the Microcredit Summit Campaign dataset: 1,471; in the MixMarket dataset: 978.
Coefficients are marginal effects after probit regressions.

Table 6. Reporting operational self-sufficiency ratio (Microcredit Summit Campaign data only).

Dependent variable:	Institution reports OSS	
Mean(dependent variable):	.605	
Average first loan size ('000 US$)	-0.009	0.016
	(0.019)	(0.021)
Log(number of active borrowers)	0.108***	0.129***
	(0.008)	(0.010)
% women borrowers	0.069	0.045
	(0.086)	(0.091)
Did not report % women borrowers	-0.090	
	(0.069)	
% poorest borrowers	0.083*	0.193***
	(0.049)	(0.066)
East Asia	-0.046	-0.008
	(0.060)	(0.072)
Eastern Europe & Central Asia	0.121	0.106
	(0.121)	(0.248)
Lat. America and Caribbean	-0.112***	-0.068
	(0.041)	(0.050)
Middle East & North Africa	0.105	0.082
	(0.091)	(0.109)
Sub-Saharan Africa	-0.053	0.081
	(0.040)	(0.057)
Number of observations	1,332	803
Pseudo-R2	0.168	0.244

note: 0.01 - ***; 0.05 - **; 0.1 - *;
Omitted region: South Asia.
Total number of observations: 1,471.
Coefficients are marginal effects after probit regressions.

Table 7. Professionalism in reporting (MixMarket data only).

Dependent variable:	Dummy 4+ diamonds	Dummy 5 diamonds
Mean(dependent variable):	0.599	0.282
Log(number of borrowers)	1.469*	-0.406
	(0.833)	(0.519)
Log(average loan balance)	1.451*	-0.387
	(0.833)	(0.518)
Log(gross loan portfolio in US$)	-1.366	0.533
	(0.833)	(0.519)
Operational self-sufficiency	0.0004	0.0010
	(0.0005)	(0.0006)
Portfolio at risk >30 days ratio	0.0005	-0.003
	(0.0018)	(0.002)
Return on assets (%)	-0.005**	-0.005*
	(0.002)	(0.003)
Women borrowers (%)	0.001	0.003***
	(0.001)	(0.001)
% clients living with less than US$1/day	0.002	-0.000
	(0.001)	(0.001)
Did not report % clients living with less than US$1/day	-0.037	-0.077
	(0.061)	(0.072)
Bank	0.061	-0.099
	(0.105)	(0.078)
Cooperative/Credit Union	-0.233***	-0.204***
	(0.066)	(0.045)
Non-Bank Financial Institution	0.099**	-0.001
	(0.042)	(0.047)
Rural bank & Other	-0.030	-0.091
	(0.084)	(0.076)
East Asia	0.141***	0.117
	(0.046)	(0.095)
Eastern Europe & Central Asia	0.205***	0.252**
	(0.051)	(0.103)
Lat. America and Caribbean	0.373***	0.298***
	(0.040)	(0.077)
Middle East & North Africa	0.222***	0.315***
	(0.030)	(0.109)
Sub-Saharan Africa	0.268***	0.267***
	(0.032)	(0.085)
Number of observations	751	751
Pseudo-R2	0.275	0.230

note: 0.01 - ***; 0.05 - **; 0.1 - *;
Sample consists of 923 institutions with 3 or more diamonds.
Omitted region: South Asia. Omitted institution type: NGO.
Coefficients are marginal effects after probit regressions.