Lying about Borrowing[*]

Dean Karlan
Yale University
Innovations for Poverty Action

Jonathan Zinman
Dartmouth College
Innovations for Poverty Action

November 2007

ABSTRACT

We compare survey self-reports with administrative data and find that nearly 50% of recent borrowers do not report their high-interest consumer loans. Under-reporting appears to be correlated with several characteristics of interest, in particular gender: 62% of women, when interviewed by men, under-report whereas 42% of women interviewed by women under-report. On the other hand, 40% of men under-report, irrespective of the gender of the interviewer. As such relying strictly on self-reported data may lead to biased inference, and we outline some methodological implications for identifying impacts of credit access on borrower behavior and outcomes. Matching female surveyors to female respondents appears to be a low-cost mitigating strategy, but clearly the best strategy is to make sure one has administrative data from a lender to measure actual borrowing history.

---

1. Introduction

Social scientists rely heavily on self-reported data. But can respondents be trusted to report the truth? This is a particularly important issue when the variable of interest is used, uninstrumented, as a right-hand-side variable in empirical analysis. Mean-zero errors or lies in self-reports will bias estimates of the correlation of interest toward zero. Systematic and heterogeneous tendencies to under-report or over-report will bias the estimated correlation in an indeterminate direction. We explore the extent and implications of respondent mis-reports of recent borrowing in a sample of borderline consumer loan applicants in South Africa.

Prior research suggests that survey respondents are likely to under-report activities viewed as socially undesirable. Studies comparing self-reports to administrative or observational data have found, e.g., that respondents underreport criminal activity (Wyner 1980), and smoking (Means, Habina, Swan and Jack 1992),[1] and over-report voting (Parry and Crossley 1950) and good academic performance (Cassady 2001).[2] Credit card borrowing is under-reported by a factor of two in the U.S. (Zinman 2007), while also being widely viewed as undesirable (Durkin 2000).[3] Borrowers may also decline to report if they suspect that survey information will be used by debt collectors.

We measure under-reporting by comparing survey self-reports with administrative data in a setting where there is likely stigma and suspicion: the "cash loan" market in South Africa. Our sample includes first-time applicants for four-month installment loans at roughly 200% APR.

We find that nearly 50% of respondents known to have borrowed recently from the Lender do not report *any* recent borrowing from a financial institution. We label this under-reporting "lying," since the high cost and amount of transactions with the Lender makes it implausible that respondents simply forgot about their borrowing. We also bound the prevalence of *over-reporting* at relatively low levels by identifying individuals who report borrowing in the survey, but who did not borrow from our Lender (they may have borrowed elsewhere, hence this is simply the upper bound of over-reporting of borrowing).

While the prevalence of lying in survey responses is interesting in its own right, we are primarily interested in the implications for statistical inference.

---

[1] Both of these studies find that the probability of under-reporting varies inversely with self-reported perceived desirability of the activity.

[2] Parry and Crossley's paper also found that easily verifiable information, such as age or telephone number, was reported accurately.

[3] Other plausible explanations for the under-reporting of credit card debt include forgetting (Soman 2001) and survey fatigue (given that most households have several credit cards).

Consequently we also explore *who* lies. More formally, we are interested in whether heterogeneity in reporting that would typically be unobserved (i.e., whether the respondent lies) is correlated with respondent characteristics that often interest researchers studying credit access and behavior.

We find that female borrowers are significantly more likely to lie *than* males, and significantly more likely to lie *to* males (i.e., females lie more when interviewed by a male, while male lying is invariant to surveyor gender). We also find suggestively large differences across several other demographics, choices, and well-being measures, although in most cases our small cell sizes produces differences that are not statistically significant.

Overall, the results suggest that researchers should think twice before using borrowing self-reports for statistical inference. We outline some alternative strategies for measuring borrowing, and its relationship to other choices and outcomes, in the final section.

2. Setting and Sample Characteristics

Our sample is drawn from marginal, first-time applicants for expensive consumer credit from a leading South African firm in the "cash loan" market.

Our cooperating Lender identified a pool of marginal applicants as part of a field experiment on the profitability implications of relaxing its underwriting criteria, and the impact on marginal clients from receiving credit. Over three thousand first-time applications at eight participating branches during September-November 2004 were assessed using the Lender's typical underwriting process. 787 of these applications were labeled "marginal" by loan officers— rejected under the usual criteria, but potentially creditworthy and hence eligible for a randomized "second look" at the application. 325 applications were assigned a second look in this encouragement impact evaluation design, and loan officers made loans to more than 50% of these 325 cases.

We then hired a survey firm to administer a detailed household survey on financial condition and well-being to the pool of marginal applicants (surveys were completed for 626 of the 787 targeted households). The survey firm was entirely independent of the Lender, and in fact did not even know of the existence of a Lender. Individuals simply were asked to participate in a household survey for research purposes, between six and twelve months after they applied for a loan from the Lender. The survey firm succeeded in completing interviews with 80% of the individuals in the Lender's experiment. We also worked with the Lender to track loan repayment and credit scores over time using administrative data. In other work, the random assignment and follow-up outcome data is used to identify impact of credit access on various measures of economic and subjective well-being (Karlan and Zinman 2007).

Here we compare self-reports on recent borrowing history to the Lender's records on who actually got a loan. We consider the 553 cases where we

obtained self-reports from the marginal applicant (as opposed to surveying another member of the applicant's household).

We provide a summary of the loan details and application process here; for a more complete description see Karlan and Zinman (2007). The product that individuals in our sample applied for was a four-month installment loan at 11.75% monthly, simple interest (about 200% APR). Expensive consumer loans of this type are prevalent and economically important in South Africa.[4] Confirmed steady wage employment, absence of derogatory credit history, and manageable total debt service are the primary criteria for loan approval. Loan sizes tend to be small relative to the fixed costs of underwriting and monitoring them, but substantial relative to a typical borrower's income. Our sample's median loan size of 1000 Rand ($150) was nearly 50 percent of the median gross monthly income. Loan uses span a range of consumption smoothing and investment projects.

Unsurprisingly, cash loan applicants tend to be working poor and lack the credit history and/or collateralizable wealth needed to borrow from traditional institutional sources such as commercial banks. Households in our sample earn roughly the median South African income on average, but have larger-than-average households to support.

3. Measuring Under-Reporting, and Bounding Over-Reporting

First we compare self-reports on recent borrowing history to the Lender's administrative data to construct measures of lying about borrowing.

The 553 applicants in our sample were surveyed 6 to 12 months after they applied with the Lender. 156 of these applicants actually received a loan from the Lender, and we mostly restrict attention to these "known borrowers" in the rest of the paper.

Self-reports on borrowing were elicited through detailed questions on loan applications and application results during the last year.[5] From questions on the timing, disposition, and place of application we flag all the loans the respondent reports anyone in the household obtaining, and classify the loans as coming from an informal sector, microlender, or other formal sector source.

We then classify the respondent as lying if s/he is a known borrower but does not report having obtained a microloan (or any other formal sector loan) in

---

[4] Estimates of the proportion of working-age population currently borrowing in the cash loan market range from below 5 percent to around 10 percent. Aggregate outstanding loans in the cash loan market segment equal about 38 percent of non-mortgage consumer debt (sources: reports by the Department of Trade and Industry, Finscope South Africa, Micro Finance Regulatory Council, and South African Reserve Bank).

[5] The survey instrument is available from us upon request.

the survey. As noted at the outset we label under-reporting "lying" because we deem it highly unlikely that a respondent actually forgets having borrowed at 200% APR within the last 6-12 months. This is especially so given the required monthly payments, and the frequency (57%) of repeat borrowing with the Lender. So most borrowers in our sample would have had contact with the Lender no more than a few months prior to the survey.[6]

Table 1, Panel A shows that about 50% of known borrowers "lie": they do not self-report having borrowed. The first row counts the respondent as reporting truthfully only if s/he reports a "microloan". 56% of our known borrowers lie under this definition (Column 1). However, reasonable people might have different definitions of "microlender", so in the second row we only count a known borrower as lying if they do not report any formal sector loan (i.e., any loan from a financial institution). By this definition, 47% lie.

Column 2 of Panel A considers the possibility that additional survey prompts might uncover additional loans. Our survey instrument asked respondents about borrowing indirectly (via questions on large expenditures) as well as directly. But it turns out that counting loans reported in the expense section reduces under-reporting very little. We will discuss the methodological implications of this below.

Column 3 of Panel A shows that the prevalence of lying about having a current outstanding loan is no different. 129 respondents were known to have a loan outstanding at the time they were surveyed; this is higher than the 57% who took an additional loan from the Lender due to delinquent borrowers.

Panel B suggests that respondents are far less likely to over-report than under-report. Although we cannot measure over-reporting precisely because we lack contemporaneous administrative data on borrowing from other lenders, we can calculate upper bounds by counting respondents who did not get a loan from the Lender but did a report a loan when surveyed. At most 28%-30% of respondents over-report based on the formal loan measures, and at most 8%-12% of respondents over-report based on the microloan measures.

---

[6] There is a potential source of minor misclassification in our definitions. We will overstate lying *per se* if respondents misremember the precise timing of loan origination (since we count only loans reportedly taken out since the date of application with the Lender). But note that we will correctly record under-reporting with respect to the treatment window of interest, and such windows are likely to be relevant in an evaluation setting. Moreover in practice the prevalence of under-reporting falls only slightly if we expand our definition of truthful reporting to count loans reportedly taken out in the last 12 months, but before the treatment date (i.e., to count loans before the date that the applicant actually applied with the Lender).

4. Who Lies?

The question of who lies has implications for statistical inference. If a valid instrument for self-reported borrowing is not available then correlations between heterogeneity in reporting that would typically be unobserved (i.e., whether the respondent lies) and other variables of interest will lead to biased inference.

Table 2 explores whether such confounding correlations exist by measuring whether lying prevalence varies with several household characteristics, choices, and measures of well-being that typically interest researchers studying borrowing behavior. Panel A considers several variables measured from administrative data.

The first row estimates that females are 13 percentage points more likely to lie than males. The difference is significant with 91% confidence. Although surveyors were instructed strictly to conduct all interviews privately, we examine whether the effect is driven by cohabiting females (or males) being more likely to lie (i.e., if surveys were not done privately perhaps individuals lied to the surveyor because their spouse did not know about the loan). This is true for men: 52% of cohabiting men lie, whereas only 32% of non-cohabiting men lie (p-value = 0.10). The opposite is true for women: 43% of cohabiting and 62% of non-cohabiting (p-value = 0.11). This pattern is interesting and suggests that some combination of interview privacy and/or gender-varying correlations between other household or respondent characteristics and co-habiting are important. We also find that matching the gender of respondent and surveyor reduces lying by females; i.e., females lie significantly more when interviewed by a male surveyor.

The fourth row considers the other margin microlenders often use for targeting: income. Borrowers with greater than (in sample) median income at the time of application are 11 percentage points less likely to lie, but the difference is insignificant statistically (p-value = 0.17). The next six rows suggest that there is little correlation between lying and some other demographics (race and marital status), or between lying and various measures of creditworthiness and repayment behavior.

Table 2 Panel B explores correlations between lying and several variables used in Karlan and Zinman (2007) to measure choices and outcomes that might be impacted by access to credit. Many of the estimated differences are large economically, but only one out of fifteen is statistically significant due to our small cell sizes.

Table 3 shows similar analysis, but using multivariate ordinary least squares. The gender result remains statistically significant when the administrative data only are included. When further independent variables are included from the survey data, the point estimate on the gender result diminishes from 0.25 to 0.16 and is no longer statistically significant.

5. Interpretation and Implications

We compare administrative data on lending to survey self-reports on borrowing and find that lying about borrowing is prevalent. About 50% of known recent borrowers from a single consumer lender do not report *any* recent formal sector loan. There is some (largely noisy) evidence that lying prevalence appears to vary with borrower characteristics, choices, and well-being.

We put forward five explanations for this under-reporting. First, individuals may have forgotten about the debt. But this seems implausible given the nature, frequency and high costs of transactions with this Lender. Furthermore, over-reporting is bounded at a relatively low level. Table 3 also shows no noticeable correlation between time since application and likelihood of lying.

The second, third and fourth potential explanations for under-reporting all point to lying. Qualitative evidence indicates that there is some social stigma to borrowing in this market. Our results suggest that this stigma is more severe for females and can be mitigated by assigning female surveyors to interview female borrowers. Survey respondents may also lie if someone else (e.g., their spouse) is present at the interview and they do not want that person to know about the loan (our data support this hypothesis for male but not female borrowers). Yet another reason for lying is if respondents suspect that their financial information will be shared with debt collectors or fraudsters.[7]

A final potential explanation for under-reporting is that our surveyors may have made systematic mistakes in recording identifying or loan data. But cross-checks with administrative data indicate accurate identification of respondents. And there is little reason to think that surveyors would systematically under-record rather than over-record debts.

Regardless of the source of the under-reporting, our results cast some doubt on the feasibility of obtaining unbiased estimates of the impacts of borrowing or credit access (on, e.g., household choices or outcomes) using self-reported data on borrowing.

There are at least two complementary methodological alternatives to measuring borrowing using self-reports. One is to obtain administrative data on lending to individuals. This is likely to be useful and practical for impact evaluations conducted with particular lenders. Second, in some markets, obtaining credit bureau data may be feasible. The issue then becomes whether the credit bureau has sufficiently comprehensive coverage of borrowing activity

---

[7] Upon first contact, many targeted respondents were in fact suspicious of "research surveys" and unwilling to participate. Reluctant or unreachable individuals were contacted repeatedly by phone and in-person visits over a two month period, and ultimately 80% of targeted respondents were surveyed.

(some bureaus report only late payments) and lenders (i.e., do the relevant credit sources actually report to the bureau?).

Lying about borrowing may have some substantive import as well. In particular exploring the source of the gender differences strikes us as an interesting line of inquiry.

REFERENCES

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir and Jonathan Zinman (2007). "What's "Creative Worth"? Evidence from a Consumer Credit Marketing Field Experiment." Working Paper. August.

Cassady, Jerrel (2001). "Self-Reported GPA and SAT: A Methodological Note." *Practical Assessment Research and Evaluation* 7(12)

Durkin, Thomas A. (2000). "Credit Cards: Use and Consumer Attitudes." *Federal Reserve Bulletin* 86(9): 623-634. September.

Karlan, Dean and Jonathan Zinman (2007). "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts." Working Paper. June.

Means, Barbara, Katherine Habina, Gary Swan and Lisa Jack (1992). "Cognitive Research on Response Error in Survey Questions on Smoking." Vital Health & Statistics. Hyattsville, MD, National Center for Health Statistics. Series 6, Number 5.

Parry, Hugh and Helen Crossley (1950). "Validity of Responses to Survey Questions." *The Public Opinion Quarterly* 14(1): 61-80.

Soman, Dilip (2001). "Effects of Payment Mechanism on Spending Behavior: The Role of Rehearsal and Immediacy of Payments." *Journal of Consumer Research* 27(4): 460-474. March.

Wyner, Gordon (1980). "Response errors in self-reported number of arrests." *Sociological Methods and Research* 9: 161-177.

Zinman, Jonathan (2007). "Where is the Missing Credit Card Debt? Clues and Implications." Working Paper. July.

<div align="center">Table 1. Lying
Proportions</div>

	Since Application with Lender		At Survey (~12 months after Application
	First Pass	More Prompts	First Pass
Panel A. Underreporting of Borrowing	(1)	(2)	(3)
Does Not Report any Microloan; Known to Borrow according to Lender's Administrative Data	0.5641	0.5577	0.5581
Does Not Report any Formal Sector Loan; Known to Borrow according to Lender's Administrative Data	0.4744	0.4423	0.4419
Number of Respondents Known to Borrow according to Lender's Administrative Data	156	156	129
Panel B. Overreporting of Borrowing (upper bound)			
Reports any Microloan; No Loan according to Lender's Administrative Data	0.1083	0.1184	0.0755
Reports any Formal Loan; No Loan according to Lender's Administrative Data	0.2771	0.2947	0.2759
Number of Respondents without a Loan according to Lender's Administrative Data	397	397	424

Sample frame is individuals who applied for a 200% APR loan from a South African consumer lender in September-November 2004. All applicants had not borrowed from the Lender previously. The 156 individuals in Panel A received loans as a result of the randomization of marginal applicants described in Karlan and Zinman (2007a). The 397 individuals in Panel B did not receive a loan from the Lender. Self-reports are from a survey on household finance and well-being conducted by a survey firm (with no stated relationships or link to the Lender) one year after the loan application. Here we only include surveys where the targeted respondent (i.e., the person who applied with the Lender) was surveyed.

Table 2. Proportion of Respondents who Underreported, by Demographic, Economic and Wellbeing Measures
Comparison of Proportions

		Proportion that Underrerported [Number of observations]		Proportion that Underrerported [Number of observations]	T-stat and (p-value) for comparison of proportions
Panel A: Variables from administrative data		(1)		(2)	(3)
Gender	Male	0.40 [68]	Female	0.53 [88]	1.70* (0.09)
Co-habitating (females only)	Cohabiting	0.43 [35]	Not cohabiting	0.60 [53]	1.62 (0.11)
Co-habitating (males only)	Cohabiting	0.52 [27]	Not cohabiting	0.32 [41]	-1.67* (0.10)
Surveyor Gender (female respondents)	Male surveyor	0.62 [52]	Female surveyor	0.42 [36]	-1.85* (0.07)
Surveyor Gender (male respondents)	Male surveyor	0.38 [32]	Female surveyor	0.42 [36]	0.35 (0.73)
Income at Loan Application	Above Median:	0.43 [89]	Median or below:	0.54 [67]	1.37 (0.17)
Married	Yes	0.46 [57]	No	0.48 [99]	0.34 (0.73)
Race	African	0.46 [101]	other race	0.50 [54]	0.53 (0.60)
Paid Back Initial Loan from Lender in Full	yes	0.47 [111]	no	0.49 [45]	0.23 (0.82)
Creditworthiness at Loan Application	Closer to bar:	0.45 [85]	Farther from bar:	0.51 [71]	0.74 (0.46)
Credit Score 13-15 months after loan	Above 600	0.47 [94]	600 or less:	0.48 [62]	0.19 (0.85)
Credit Score 25-27 after loan	Above 600	0.48 [99]	600 or less:	0.46 [57]	-0.34 (0.73)
Panel B: Variables from self-reports from follow-up survey					
Hunger	none in past 30 days	0.50 [139]	yes in past 30 days	0.29 [17]	-1.58 (0.12)
Food Quality	improved over last year	0.51 [43]	did not improve	0.46 [112]	-0.53 (0.60)
Employed	yes	0.45 [139]	no	0.67 [15]	1.58 (0.12)
Poverty line	above	0.46 [112]	below	0.54 [41]	0.89 (0.38)
High school graduate	yes	0.47 [100]	no	0.48 [56]	0.14 (0.89)
Anyone in household is a university student	yes	0.54 [13]	no	0.51 [94]	-0.19 (0.85)
Bought or improved dwelling since loan	yes	0.38 [56]	no	0.53 [100]	1.87* (0.06)
Anyone in household self-employed	yes	0.50 [26]	no	0.47 [125]	-0.26 (0.80)
Decision-making scale	Above median	0.47 [34]	Median or below:	0.36 [28]	-0.89 (0.38)
Optimism scale	Above median	0.45 [69]	Median or below:	0.49 [86]	0.48 (0.63)
Position on community socio-economic ladder	Above median	0.50 [78]	Median or below:	0.45 [78]	-0.64 (0.52)
General health is "very good"	yes	0.48 [85]	no	0.46 [71]	-0.22 (0.83)
No household member sick in previous 30 days	yes	0.53 [77]	no	0.42 [79]	-1.44 (0.15)
Depression	Less than Median	0.63 [40]	Median or above	0.45 [40]	-1.57 (0.12)
Stress	Less than Median	0.58 [38]	Median or above	0.51 [43]	-0.60 (0.55)

University attendance measured only for households with members between the ages of 18 and 30. Depression, stress, optimism, and decision-making measured using multi-question modules; see Karlan and Zinman (2007) for details. Decision-making questions asked only of married respondents. Depression and stress questions asked only of a subset of respondents.

Table 3. Predictors of Underreporting
OLS

Dependent variable = 1 if respondent underreported borrowing	Independent Variables Incldue Administrative Data Only		Independent Variables Include Administrative and Survey Data
	(1)	(2)	(3)
Female	0.25192*	0.24973*	0.15957
	(0.13732)	(0.13916)	(0.15255)
Cohabitating	0.12410	0.12655	0.03326
	(0.15851)	(0.15849)	(0.22406)
Female*cohabitating	-0.24016	-0.23798	-0.12583
	(0.18570)	(0.18707)	(0.20521)
Female surveyor	-0.10938	-0.11353	-0.00620
	(0.15438)	(0.15845)	(0.19570)
Female respondent*female surveyor	-0.16746	-0.16742	-0.16065
	(0.17660)	(0.17871)	(0.19479)
Number of dependents	-0.01197	-0.01275	0.03346
	(0.03475)	(0.03493)	(0.04581)
Income >= median in sample	-0.09184	-0.09279	-0.06097
	(0.08886)	(0.09123)	(0.10253)
Race==African	-0.03396	-0.03645	0.09032
	(0.10692)	(0.10797)	(0.13904)
Creditworthiness at application: closer to bar	-0.10998	-0.11056	-0.12250
	(0.09431)	(0.09667)	(0.10764)
Months since most recent loan	0.00594	0.00688	-0.02574
	(0.06011)	(0.06197)	(0.06588)
Months since most recent loan, squared	0.00092	0.00088	0.00406
	(0.00562)	(0.00575)	(0.00604)
Paid loan in full	-0.03133	-0.03133	-0.01070
	(0.11108)	(0.11077)	(0.11057)
Credit score 13-15 months after application: above 600		0.00942	0.03953
		(0.09725)	(0.10944)
Credit score 25-27 months after application: above 600		-0.02267	-0.12281
		(0.09974)	(0.10660)
No hunger in household in past 30 days			0.43821***
			(0.13418)
Household food quality imporved over last year			0.15611
			(0.11938)
Employed			-0.37354**
			(0.17139)
Household is above poverty line			-0.05458
			(0.11810)
High school graduate			-0.06573
			(0.09895)
Anyone in household is enrolled in higher education			-0.02738
			(0.19857)
Bought or improved dwelling since loan			-0.17239
			(0.11836)
Any self-employment			0.00574
			(0.12875)
Decision making power scale: above median			0.23845
			(0.15137)
Optimism scale: above median			-0.11932
			(0.10438)
Position on community socioeconomic ladder: above median			0.02449
			(0.09117)
General health is "very good"			0.01311
			(0.09359)
No household member sick in past 30 days			0.02759
			(0.10570)
Depression scale: less than median			0.24197*
			(0.13466)
Stress scale: less than median			-0.22579
			(0.13871)
f-test p-value on month of application dummies	0.79	0.77	0.94
f-test p-value on branch dummies	0.21	0.24	0.01
r-squared	0.139	0.139	0.309
Mean of dependent variable	0.47	0.47	0.48
N	155	155	145

* p<0.10, ** p<0.05, *** p<0.01.

OLS with robust standard errors. Survey variables pertain to respondent unless otherwise noted (in which case question queried the respodents about the entire household). Number of dependents top-coded at 6. All regressions also include dummies for branch where respondent applied, and month-of-application. Regressions also include dummy variables for missing values that enable us to keep respondents who were not asked the stress, depression, or decision making variables in the sample.