Pre-Analysis Plan
Debt Restructuring in the Shadow of Moral Hazard: A Field Experiment with Small Businesses
August 6, 2024

Frequency/time-horizon of treatment effect estimates

We will estimate by both month-by-month and 12-month treatment effects for each outcome, with exact definitions as specified below. We plan to present month-by-month estimates only graphically, and mostly for descriptive purposes. In fact, we do not expect that the month-by-month estimates will be well-powered for most outcomes. Our tables will focus on 12-month outcome measures.

First Stage

1. Monthly payment ("first-stage")

Monthly definition: Monthly installment (payment) owed by the borrower to the lender on the minibus loan.
12-month definition: Average monthly payment owed across the 12-month follow-up period.

Motivation: Payment Reduction Treatment is designed to reduce this quantity relative to the Restructuring plan (and the control group). The randomization is subject to operational constraints, so compliance will be imperfect => important to quantify strength of first stage.

2. Interest Rate ("first-stage")

Monthly definition: Current interest rate for the loan.
12-month definition: Average interest rate across the 12-month follow-up period.
Motivation: See 1. Above

Dealing with unbalanced panel: Some loans will leave the books during the 12-month period (due to pre-payment, repossession, loan maturing and paid in full, etc.). For first-stage outcomes we will use only on-books months and not impute outcomes for off-books months.

Primary Outcomes

1. Index of Loan performance: we will examine an index of loan performance, which combines information from three of the lender's key metrics to assess performance. To be precise, the index is constructed as the (equally weighted) average of three metrics defined below, after they are standardized to have zero mean and standard deviation one. The components are:

   i. Indicator variable of whether the account is not current.
   Definition: An account is not current if Total arrears < 100 ZAR in the month, with unmet obligations > 30 days late considered arrears.

   ii. Total Arrears Amount.
   Definition: Total Arrears Amount reported by the lender.

iii. Total Arrears Amount, Scaled
Definition: Total Arrears Amount divided by Monthly Owed Installment.

Monthly loan performance: measured as a snapshot taken each monthly reporting date.
12-month loan performance: measured as a snapshot taken on the 12th-month reporting date.

Dealing with unbalanced panel of loan performance. If a loan leaves the books, we will impute that month and subsequent months as follows:

- Current and zero arrears, for a loan paid in full
- Not current and the last observed arrears amount, for a loan not repaid in full

2. Index of Labor Supply: we examine an index of labor supply, which combines information from four key metrics constructed using the GPS data on borrowers' driving. As before, the index is constructed as the (equally weighted) average of the metrics defined below, after they are standardized to have zero mean and standard deviation one. The components of the index are:

i. Distance driven (in kms).
Definition: Sum of the total distance covered by the vehicle.

ii. Number of hours driven.
Definition: Sum of the total time the vehicle was active.

iii. Number of days worked.
Definition: Sum of total number of days worked in the month. A working day is defined as the day in which non-zero trips were made.

iv. Number of hours on the job:
Definition: The total duration between vehicle first trip of the day and the end of the last trip of a day.

Monthly labor supply measures: Total for a given month.
12-month labor supply measures: Summed across all months.
Dealing with unbalanced panel of labor supply measures: We will only have data for drivers while their loan is on the books. We do not plan to impute missing values, given our focus on the principal-agent problem—the lender cares only about driver effort while the borrower is obligated to repay.

Final note on GPS data: For these outcomes we are fairly certain that the following characteristics indicate GPS device measurement error and/or non-work trips, and we plan to drop these trips with (at least) one of these characteristics from our analysis data:

(i) > 1000 km distance
(ii) duration more than 8 hours
(iii) implied speed > 200 km/hr for trips

We are still working with the lender to improve our understanding of the GPS data, in particular w/r/t identifying non-work trips in the data. As a result, we may also present results with more stringent filters, while being transparent about the impact of these dealings on the results.

Secondary Outcomes

(A) Main Outcome Index component variables

(B) Other Labor Supply Outcomes:
Index of risky driving behavior: we will examine an index of risky driving behavior, which combines information from two key metrics constructed using the GPS data on borrowers' driving and their filed accidents reports. As before, the index is constructed as the (equally weighted) average of the metrics defined below, after they are standardized to have zero mean and standard deviation one. The components of the index are:

i. Number of accidents:
Definition: based on insurance claims filed with the lender.

ii. Number of instances of over-speeding:
Definition: Number of times the GPS device in the vehicles recorded driving more than 120 km/hr.

(C) Credit access index: we will examine an index of credit market access that combines information from key metrics constructed from credit bureau reports. As before, the index is constructed as the (equally weighted) average of the metrics defined below, after they are standardized to have zero mean and standard deviation one. We plan to include the following components:

i. Has a credit or store card.
Definition: the borrower has any credit or store card reported in the credit report.

ii. Has credit line.
Definition: the borrower has any credit line in the credit report.

iii. Total amount of credit lines.
Definition: the number of credit lines in the credit report.

iv. Has an installment loan from another lender

v. Has credit score.
Definition: flag equal to one if the borrower has a credit score assigned.

vi. Credit score conditional on having one.

Definition: the credit score reported in the credit report. If no score, we use only the other index components to construct the index.

Monthly measures: the snapshot provided in each monthly report.
12-month measures: final snapshot available at the end of the 12-month period.
Dealing with missing information: we are generally going to interpret missing information as the lack of credit access: in other words, a person without a report is considered to have no credit lines, no cards, no credit score. If information is not provided consistently for each month, we will always try to use the most updated information before the event. We will consider both winsorized and non-winsorized version for the variables that are continuous.

(D) Borrowing index: we will examine an index assessing the "outside" borrowing activity of each borrower in our sample, using information from key metrics constructed using credit bureau reports. As before, the index is constructed as the (equally weighted) average of the metrics defined below, after they are standardized to have zero mean and standard deviation one. We plan to include the following components:

i. Has non-zero debt with other lenders.
   Definition: the borrower has non-zero debt with any other lender reporting in the credit report.

ii. Line utilization.
   Definition: share of the credit lines available that are drawn down at the time considered.

iii. Total amount owed on installment loans from other lenders.

iv. Total amount owed on revolving lines of credit.

Monthly measures: snapshot in each monthly report.
12-month measures: average across the available months in the 12-month period.
Dealing with missing information: we are generally going to interpret missing information as the lack of borrowing: in other words, a person without a report is assumed to not have any borrowing activity. If information is not provided consistently for each month, we will always try to use the most updated information before the event. We will consider both winsorized and non-winsorized versions for the variables that are continuous.

(E) Repayment index or indices: we examine an index assessing the driver's repayment behavior on other loans. This index combines information from key metrics constructed using the credit bureau reports. As before, the index is constructed as the (equally weighted) average of the metrics defined below, after they are standardized to have zero mean and standard deviation one. Our plan is to create two versions of the index, one for secured borrowing and one for unsecured borrowing. For each version, we plan to include the following components:

i. Any account that is past due

Definition: Indicator for any credit line being 30+ days past due

ii. Number of accounts that are past due
Definition: Count of the credit lines with payments 30+ days past due

iii. Any account in default
Definition: Indicator for any credit line being 90+ days past due

iv. Number of accounts in default:
Definition: Count of number of credit line being 90+ days past due

v. Account balances past due
Definition: Sum of accounts balances for credit lines that are 30+ past due

vi. Account balances in default:
Definition: Sum of accounts balances for credit lines that are 90+ days past due

Our primary index here will be based only on outside loans outstanding at time of treatment assignment.

Possible secondary indices: if we find treatment effects on outside borrowing, we will also consider indices based on all outside loans observed during the experimental period.

Monthly index or indices: will use measures from the snapshot provided for that month.

12-month index or indices: average measure across the available months in the 12-month period.

Dealing with missing information: we are generally going to interpret missing information as the lack of borrowing: in other words, a person without a report is assumed to not have any borrowing activity. If information is not provided consistently for each month, we will always try to use the most updated information before the event. We will consider both winsorized and non-winsorized version for these variables.

Process for dealing with outliers

All of the data here will be shared with us by the lender, and we are still working with the lender to understand all the nuances of the various data-generating and -reporting processes. E.g., if a borrower pays significantly more than contractual installment in a particular month, the lender reports the overpayment as a negative payment from the client (i.e. a credit to the client) in subsequent month. As such, we are still unsure exactly whether and how outliers should be dropped or recoded, and how outlier-dealing rules might vary across outcomes. Our plan is to provide quantitative description of each outcome measure's distribution, to detail how and why we are dealing with any outliers, and to present results with and without any such dealings.

Additional tests for heterogeneous treatment effects

We expect it will be interesting, given prior work on strategic default on collateralized loans (mostly on the housing market), to estimate HTEs based on measures of the borrower's equity position on the vehicle. We expect to get detailed vehicle information that should be useful for this (make/model/year). Another likely source of cross-sectional variation in vehicle equity is the months remaining to maturity at baseline, since the loans follow typical amortization schedule. Vehicle accidents post-origination but pre-experiment (we might relax the pre-experiment condition if we don't find treatment effects on accidents) are another potential source of variation, although we are still working to understand their frequency and likely effects on vehicle equity (e.g., given insurance coverage, accidents need not have large negative effects).