

Deferring Wages and Labor Supply in Malawi

Analysis Plan Part 3: Addendum on Empirical Strategy

Lasse Brune, Eric Chyn and Jason Kerwin

March 1, 2017

Abstract:

This document is the third portion of the analysis plan for a randomized controlled trial (RCT) evaluation studying the effects of a deferred wage payment option on workers' investments, consumption, well-being and labor supply. Between January and May 2017, the Lujeri Tea Estates will allow randomly selected workers the choice of deferring a portion of their bi-weekly wages into a savings account to be paid out at the end of the agricultural season. The present document is an addendum about handling outliers in our empirical strategy.

Addendum to Empirical Strategy:

In Part 1 of our analysis plan, we described our main econometric specification for examining the treatment effects of the deferred wages product. We did not describe a plan for handling outliers, which are an important issue when considering survey data on income and expenditures.

We will handle outliers in the following way. First, all continuous survey variables will be winsorized (i.e. top- and/or bottom-coded) at the 99th percentile for variables with nonnegative support, at the 1st percentile for variables with nonpositive support, and at both 99th and 1st percentiles for variables with nonnegative and nonpositive support.¹ The exception are variables for which we specify a larger percentage of the data to be top- and bottom-coded (for example, in Part 2 of our analysis plan, we lay out a specific winsorization procedure for expenditure share variables). These winsorized versions of the variables will be used in our main regression analyses, for both outcomes and controls.

Second, we will conduct two additional sets of regressions as sensitivity analyses, to determine whether real treatment effects are being obscured by outliers (or by an outcome variable's thick tailed distribution more generally). Each will rely on our main econometric specifications from Part 1 of the analysis plan, and will use the winsorized versions of the outcome variables described above. For the first, for variables with no negative values, we will take the inverse hyperbolic sine transform of the outcome variable before running regressions. This function approximates a logarithm in functional form and is defined for variables that contain zeroes.

For the second additional set of regressions, we will use quantile regressions instead of ordinary least squares regressions. These will be median regressions wherever possible; if median regressions do not converge (which can happen e.g. when there is a mass point at zero) we will instead focus on higher quantiles. In that case our default will be to examine the middle of the distribution of the nonzero data. For instance, if 70% of the distribution is composed of zeroes, then we would look at the quantile midway between the 70th and 100th percentiles – the 85th percentile.

A Note on Data Availability

As of the posting of this addendum to our analysis plan, we have received none of the administrative outcome data on labor supply. Our survey team has conducted one day of data collection for the first high-frequency survey, which is our first source of survey outcomes. We have seen 31 of a planned total of 870 observations for that survey.

¹ We include zeroes whenever computing percentiles.