Comments on "The Role of Regret in Prize-Linked Savings: Experimental Evidence from Kenya"

The paper studies the impact on savings of prize-linked savings (PLS) accounts. While the control group received matched contributions on their deposits, people in the two PLS arms entered a lottery to win a cash prize in proportion to the amount they saved that day. To test for regret aversion, one of the PLS arms received feedback on their lottery result, independent of whether they deposited – this is called the PLS-feedback group.

Individuals in the PLS group make more deposits, but do no save more. Based on the results from the PLS-feedback arm, the authors conclude that about 20% of the treatment effect on deposits can be attributed to regret aversion. Finally, they provide some evidence that PLS increases gambling.

The paper is very well written and interesting, but the experiment lacks power and the interpretation of the results is not always consistent. I outline this further below.

Main comments:

- There are roughly 100 respondents per treatment arm. Given that the observed impacts are rather small in the savings literature, it is not surprising that the study lacks power. The part on regret aversion is most novel and interesting, but the difference between the PLS-F and PLS-N conditions is not statistically significant. The important finding that 20% of the treatment effect can be attributed to regret aversion, is based on a substantial point estimate. I fully agree that the point estimate is substantial, but there are substantial point estimates in the balance table and in the heterogeneous effects as well, which are basically ignored:
    o Balance Table 2: there is significant imbalance for a couple of variables only, but for the most important variable – savings - the differences are substantial: the control save about 41% more than people in the PLS-N arm. From the appendix, it is clear that more details are available regarding savings at baseline, but those are not included in the baseline table (e.g. the use of ROSCAs, which - given their temporary nature - are a good comparison). It would be useful to present normalized differences across the groups, as in Imbens and Rubin (2015). That provides a scale-invariant measure of the size of the difference.
    o The authors put forward there is no heterogeneity across prior savings usage, problem gambling, risk attitudes and temporal discounting (pg 12). However, the Tables 12 and 13 show non-significant but important point estimates. The effect seems to be driven by people who do not save currently and people who are not risk averse? These traits should also be added to the balance table.

- The experiment has three important features: (i) it is limited in time (it runs for 60 days only), (ii) people were reminded about savings every day, (iii) withdrawals were allowed only once for the duration of the experiment (day 30). These features deviate from any standard account. For example, while most experiments focus on having at least two deposits in six months, the control deposit 14 times in a time span of 60 days… These features are the same across treatment arms, but it is still important to discuss what we learn about regular savings products (which is not limited in time, does not send out

reminders and where withdrawals may be costly, but are not restricted). Maybe nothing, but I would be explicit about that.

- Empirical strategy: While the use of SUR was pre-planned, this is unseen in the savings literature. A more standard approach is to estimate using OLS, but adjust the standard errors for multiple hypothesis testing. The results seem not very robust across specifications, so I would want to see those results for the main Table 3.

- The results in Table 5 are interesting, but potentially endogenous? Whether someone played in the previous time period is not exogenous and you restrict the sample to periods in which people did not play. The exercise is interesting, but should be interpreted with caution (are the results the same if you use the full sample and include a dummy for 'having played in the previous period')?

- Instead of focusing on endline outcomes (savings and expenditures), it may be worth spending more time on the heterogeneous effects (e.g. include a heterogeneous gambling effect!). First, the endline asked about the past 30 days, which implies there is an overlap with the experiment. What did you expect? People may save or spend less because their money is on the temporary account? For this exercise to be informative, the endline should have been conducted at least one month after the end of the experiment. This makes me a bit cautious about the gambling effects as well. People still had access to the PLF accounts, so 'same' gambling, implies more if they also deposit on the PLF accounts? Given the left-hand side is ordered, an ordered logit is more appropriate than a multinomial one, but for the ease of interpretation, I would just create a dummy indicating gambling increased (as you pre-planned).

- I am still a bit puzzled about the conclusions. People deposit more (but potentially for the wrong reasons), but savings – which should be the first order outcome variable – do not change. In addition, it may encourage people to gamble more. So, should we conclude this is not to be implemented? That would be an important conclusion.

Further comments:
- Some typos
    o Introduction : "They show that decision makers are (not a) reluctant to stop because their (not they) utility depends on the best foregone outcome from the entire history".
    o Pg 10: crtical
    o Conclusion: mdoel

- The participants play five different games to measure personal traits. These are used in the heterogeneous effect. There is no information about the games though, nor about what exactly they measure.

- Balance Table 2. There must be a mistake in the table: the control mean for *currently saves* is 56%.
    o (PLS-N – Control)= 5, so PLS-N==61%.
    o (PLS-F – Control)= -10, so PLS-F==46%
    o (PLS-N – PLS-F)= should thus be 15 and not -15?

- Figure 2: Is this not due to the savings reminder? We can see the peak for PLS-F, but do not know what happens in the other two arms.