Evaluation*

Fabian Gunzinger
Warwick Business School

Neil Stewart
Warwick Business School

May 12, 2022

Contents

*This research was supported by Economic and Social Research Council grant number ES/V004867/1. WBS ethics code: E-414-01-20.

Abstract

Neat and succinct abstract right here...

1 Introduction

2 Methods

While we were unable to pre-register the analysis because we have had access to and been working with the Money Dashoard data for months, we proceeded in the same spirit: we first wrote a draft of the paper in the form of a pre-analysis plan, following Olken (2015), then tested the entire code base – data pre-processing, balance checks, main analysis, and extensions – with a 1 percent sample, and finally ran the entire analysis.

2.1 Dataset

Limitations:

- We have more data for users that signed up later. So average user in the study is not the average MDB user. If time of signup is mainly driven by financial savyness, then study sample is closer to overall population than MDB sample (if we rank groups as early joiners > late joiners > never joiners in terms of financial sophistication). If, however, signup reflects something like openness to newness, then it's not necessarily correlated with financial savyness. Either way, we might ignore it for now. We could test whether behaviour differs between early or late adopters, but that doesn't seem important enough.

2.2 Sample selection

To assess the impact of MDB use on users' financial behaviour we need to observe their relevant financial history for a sufficiently long period of time prior to and after they started using the app. For our purpose here, "relevant financial history" includes the complete set of spending transactions and all savings account inflows and outflows, and "sufficiently long period" is a period of 6 months prior to and after signup, with the month of signup being the first month of the latter period.[1]

Table 1 provides an overview of the precise conditions we applied to implement these criteria and their effect on the sample size. The set of functions that implement each condition can be found on path to github.

2.3 Treatment

Figure 1 shows treatment assignment for a raw sample of random subset of 200 users. Compare to treatplot of final sample and talk about important implications...

Notes: Screenshot from the top of the Money Dashboard website, at moneydashboard.com, accessed on 29 April 2022.

Table 1: Sample selection

	Users	User-months	Txns	Txns (m£)
Raw sample	27,329	803,012	67,088,876	12,722
Annual income of at least £5,000	9,338	256,891	24,582,995	4,463
At least one savings account	5,579	167,705	16,952,445	3,249
At least 12 months of data	4,884	163,602	16,657,929	3,207
Monthly spend of at least £200	2,506	82,860	9,303,691	1,823
Final sample	2,506	82,860	9,303,691	1,823

Figure 1: Treatment assignment plot for raw sample

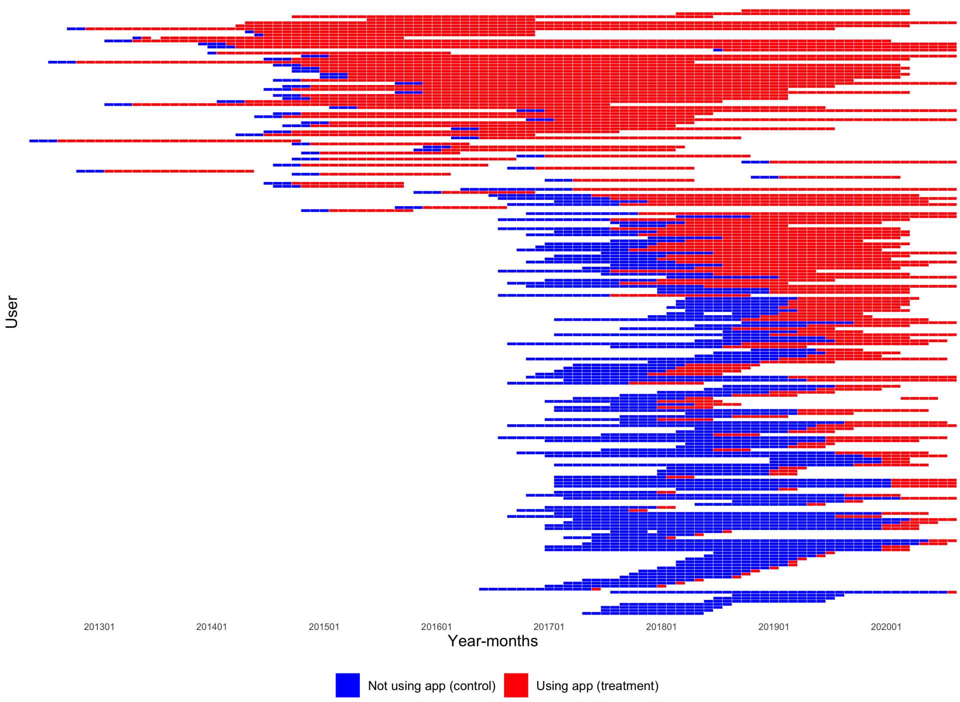


Table 2: Outcome description

Outcome (name in dataset)	Definition	Rationale
	Primary outcome	
Net savings (*sa_netflows_norm*)	Inflows into minus outflows out of all of a user's savings accounts divided by monthly income. To capture only "user-generated" flows, we exclude interest and "save the change" transactions, as well as transactions of less than £5 in absolute value. Monthly income and raw inflows and outflows are winsorised at the 1 percent level.	We focus on net inflows to capture effective savings.
	Secondary outcomes	
Positive net savings dummy (*has_pos_sa_netflows*)	Dummy equal to 1 if there were positive net savings (as defined above).	Captures extensive margin of savings (change in number of deposits).
Positive net savings (*pos_sa_netflows*)	Equal to net savings if there were positive net savings.	Captures intensive margin of savings (change in deposit amount).
...add from text below		

2.4 Outcomes

For a more nuanced understanding of how app use affects savings we also consider net-savings – total savings account inflows minus outflows – as a proportion of monthly income to see whether a willingness to save more might be offset by a (later) need to withdraw funds, and a dummy variable for whether a user has any savings account inflows in a given month to see whether the app helps users save at all. To investigate possible channels, we consider total spend, highly discretionary spend, banking charges, the total amount of borrowing, as well as payday borrowing, all as proportion of monthly income.

Adjusting for multiple hypothesis tests We think of our secondary outcomes as exploratory and do not make any adjustments for multiple hypothesis testing.[2] An alternative approach, based on Anderson (2008), would be to group outcomes into "savings", "spending", "borrowing", and "fees", and consider them as different dimensions of a latent variable of interest which we might call "financial management skills". We do not do that for two reasons: first and foremost, because we think it is natural to think of the amount saved as the ultimate outcome and of other outcomes as providing a more nuanced understanding of savings behaviour or as suggesting possible channels through which app use affects savings. Thinking of savings as the main goal is also reflected in Money Dashboard's main promise, which is to help users spend less and save more, as shown in Figure 2. Second, as pointed out in Carlin et al. (2017), incurring overdraft fees is not an unambiguous sign of a financial mistake, as the opportunity to go into

[1] In Appendix B.4 we show results with different window lengths. The results are unchanged.

[2] For a recent game-theoretically motivated discussion of when and how to correct for multiple hypothesis testing, see Viviano et al. (2021).

overdraft confers a benefit to the consumer.[3]

2.5 Covariates

Description of covariates.

2.6 Difference-in-difference

Control group design:

- We only have data for a self-selected group of people who choose to use Money Dashboard. By virtue of signing up to an app that helps them manage their money, these users are different from those who don't sign up. As a result, we are unable to answer the question of whether use of Money Dashboard helps the average person in the population as a whole save more.[4] Instead, we are answering the question whether Money Dashboard succeeds in helping its *users* save more.

- Money Dashboard can access up to three years of historic data for each account a user links to their account.

- Each user for whom we have sufficient data thus serves as both a treatment unit and a potential control unit.

- We use a difference-in-differences design to estimate the effect of app use. Because we do not have a separate control group, we use the per-signup data of Money Dashboard users as control periods and use matching to find comparable control user for each tretment user.

- To do this, we use the matching estimator for panel data proposed by Imai, Kim, and Wang (2021). Following paper, we conduct the following steps:

- For each treated observation, we find a set of control observations with that share the same treatment history for a period of L periods before the treatment and F periods after the treatment. In our baseline specification, we rely on a year's worth of data around the treatment period and set $L = 6$ and $F = 0, 1, 2, 3, 4, 5$.

- Identification assumption is that potential outcomes only depend on treatment status of the past L periods. In general, this means that if treatment has a cumulative effect over time, the full effect is reached after L periods. In our context, this means that any effect on savings behaviour from usign the app is fully realised after L periods. (I think this means that if we look at the treatment effect for F periods forward, the effect should not become stronger after F = L).

- tbc once implemented.

[3] For further discussions on fees, see Jørring (2020) and Stango and Zinman (2009).

[4] One way to get closer to that answer is to re-weight our sample on observable demographic variables so as to match the UK population as a whole. But our sample differs from the population as a whole both is ways that are observable (demographic variables) and unobservable (self-awareness that they need help managing their money, cognitive resources to engage with the app, motivation to do so). Re-weighting would only help us deal with the first of these.

- Selection of covariates: all variables that simultaneously affect treatment and outcomes. No need to control for fixed effects: these capture unobserved time-invariant factors that make an individual sign up to MDB and affect its spending habits. Given that these are time invariant, and that all users eventually sign up, there is no difference between control and treatment units in these factors. Month of year: should probably include, as can affect p of signup and spending behaviour.

Is estimate causal?

- King and Zeng (2006) show that there are four sources of bias (ommitted variable, post-treatment, interpolation, extrapolation).
- Discuss each in turn to argue that effect is causal (for our population of interest, which are people signing up to MDB).

3 Results

Main results: Figure and associated table akin to sakaguchi2022default Figure 2 and Table 6.

3.1 Descriptive analysis

3.2 Difference-in-difference analysis

3.3 Subgroups

To analyse which groups benefit most from adopting Money Dashboard, we split our sample by gender, generation, income quartiles, and pre-adoption savings behaviour.

We define generations as follows: boomers were born between 1946 and 1964, Gen X between 1965 and 1980, Millennials between 1981 and 1996, and Gen Z after 1997.[5]

Subgroup analysis: same Fig an Tab as in main analysis, but with line for each subgroup. One figure for each of: gender, generations, income terciles, per-adoption average savings tercile (inspired by Carlin et al. (2017), see Fig 5 and Table 4).

See also section 6 in Gargano and Rossi (2021)

4 Discussion

[5]Based on age ranges provides by Beresford Research.

References

Abadie, Alberto and Jérémy L'Hour (2021). "A penalized synthetic control estimator for disaggregated data". In: *Journal of the American Statistical Association* 116.536, pp. 1817–1834.

Anderson, Michael L (2008). "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". In: *Journal of the American statistical Association* 103.484, pp. 1481–1495.

Carlin, Bruce, Arna Olafsson, and Michaela Pagel (2017). "Fintech adoption across generations: Financial fitness in the information age". Tech. rep. National Bureau of Economic Research.

Gargano, Antonio and Alberto G Rossi (2021). "Goal Setting and Saving in the FinTech Era". In:

Imai, Kosuke and In Song Kim (2021). "On the use of two-way fixed effects regression models for causal inference with panel data". In: *Political Analysis* 29.3, pp. 405–415.

Imai, Kosuke, In Song Kim, and Erik H Wang (2021). "Matching Methods for Causal Inference with Time-Series Cross-Sectional Data". In: *American Journal of Political Science*.

Jørring, Adam (2020). "Financial sophistication and consumer spending". Tech. rep. Working Paper.

King, Gary and Langche Zeng (2006). "The dangers of extreme counterfactuals". In: *Political analysis* 14.2, pp. 131–159.

Olken, Benjamin A (2015). "Promises and perils of pre-analysis plans". In: *Journal of Economic Perspectives* 29.3, pp. 61–80.

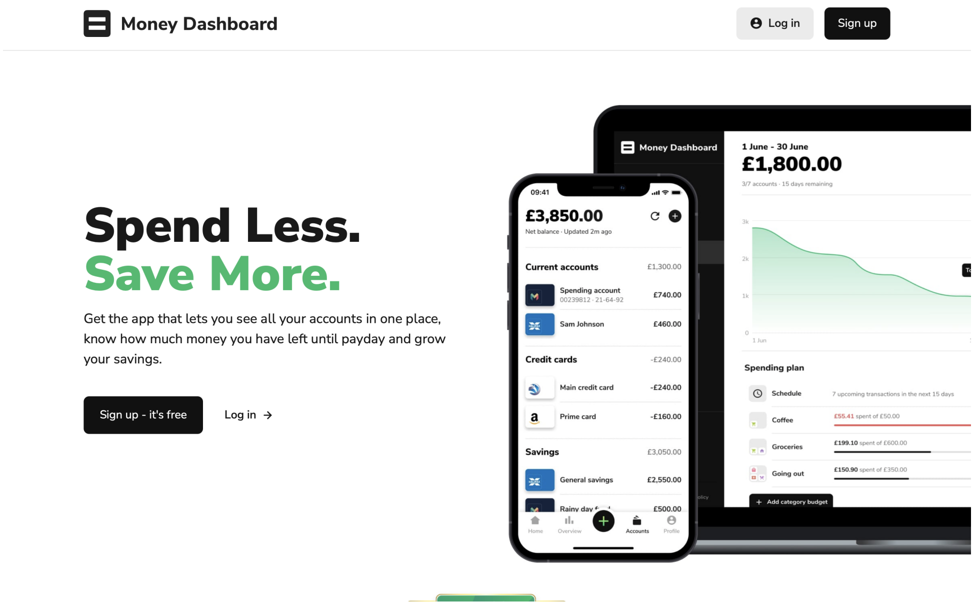
Stango, Victor and Jonathan Zinman (2009). "What do consumers really pay on their checking and credit card accounts? Explicit, implicit, and avoidable costs". In: *American Economic Review* 99.2, pp. 424–29.

Sun, Liyang and Sarah Abraham (2021). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". In: *Journal of Econometrics* 225.2, pp. 175–199.

Viviano, Davide, Kaspar Wuthrich, and Paul Niehaus (2021). "(When) should you adjust inferences for multiple hypothesis testing?" In: *arXiv preprint arXiv:2104.13367*.

A Money Dashboard application

Figure 2: Money Dashboard website screenshot



Notes: Screenshot from the top of the Money Dashboard website, at moneydashboard.com, accessed on 29 April 2022.

B Alternative specifications

B.1 Static TWFE

$$y_{it} = \alpha_i + \lambda_t + \beta T_{it} + \gamma X_{it} + \epsilon_{it} \tag{1}$$

- Comparison: pre vs post signup within each individual.
- Assumption: there are no time-varying unobserved effects that affect both y and T.
- Discussion: there is something that made the individual sign up in the first place, and it might well be an individual level shock that we don't observe (unexpected large expense, loss of job, etc.)
- See Imai and Kim (2021) for problems with twfe

B.2 Dynamic TWFE

$$y_{it} = \alpha_i + \lambda_t + \sum_{s=-6}^{5} \beta_s T_{its} + \gamma X_{it} + \epsilon_{it} \tag{2}$$

- See Sun and Abraham (2021) for problems with that and compare to their approach as implemented in fixest.

B.3 Alternative window lengths

Show results for 12 months on either end.

B.4 Alternative matching method

B.5 Post treatment periods as control

B.6 Synthetic controls

See Abadie and L'Hour (2021) for how to use synthetic controls for disaggregated data.

B.7 Classic two-way fixed-effects model