

# Evaluation\*

Fabian Gunzinger

Neil Stewart

July 5, 2022

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Methods</b>	<b>2</b>
2.1	Dataset . . . . .	2
2.2	Data preprocessing . . . . .	2
2.3	Summary statistics . . . . .	4
2.4	Treatment . . . . .	5
2.5	Outcomes . . . . .	6
2.6	Covariates . . . . .	8
2.7	Estimation . . . . .	8
2.8	Code access . . . . .	8
<b>3</b>	<b>Results</b>	<b>9</b>
3.1	Static TWFE . . . . .	9
3.2	Dynamic TWFE . . . . .	9
3.3	Decomposing inflows and outflows . . . . .	13
3.4	Decomposing intensive and extensive margins . . . . .	14
3.5	Matching . . . . .	14
3.6	Alternative window lengths . . . . .	19
3.7	Subgroups . . . . .	19
3.8	Alternative outcome variables . . . . .	19
<b>4</b>	<b>Discussion</b>	<b>19</b>
<b>A</b>	<b>Money Dashboard application</b>	<b>21</b>
<b>B</b>	<b>Data</b>	<b>21</b>
B.1	Historical data availability . . . . .	21

---

\*Fabian Gunzinger, Warwick Business School, [fabian.gunzinger@warwick.ac.uk](mailto:fabian.gunzinger@warwick.ac.uk); Neil Stewart, Warwick Business School, [neil.stewart@wbs.ac.uk](mailto:neil.stewart@wbs.ac.uk).

## Abstract

Neat and succinct abstract right here...

## 1 Introduction

## 2 Methods

### 2.1 Dataset

- 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.
- 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 Data preprocessing

**Cleaning** MDB cleaning (mdb/mdb/clean.py)

- About 75 percent of transactions are not automatically categorised, and we drop transactions that have no automatic tag. Dropping untagged transactions means that we might underestimate spending and savings, but entures that we do not miscategorise transactions.
- Custom tag classifications.
- Drop transactions for which user ID, account ID, date, amount, and transaction description are identical. This will drop some genuine transactions, such as someone buying two identical cups of coffees at the same coffee shop on the same day. However, data inspection suggests that in most cases, we remove genuine duplicates.

**Sample selection** We select our sample so as to include users for whom we can be reasonably certain that we observe all relevant financial transactions, and do so for at least six months before and after they sign up to the app. In addition to that, we exclude users who might use the app for business purposes as well as pensioners, whose financial objectives might be different.

Table 1 lists the precise conditions we applied to implement these criteria and their effect on sample size. We remove the first and last month of data for all users because we are unlikely to

Table 1: Sample selection

	Users	User-months	Txns	Txns (m£)
Raw sample	271,856	7,948,520	662,112,975	124,573
Drop first and last month	265,760	7,406,482	643,851,490	121,098
Drop test users	264,698	7,338,506	637,863,549	119,436
At least 6 months of pre and post signup data	67,059	2,020,590	201,335,520	40,935
App signup after March 2017	30,160	860,619	85,085,026	17,569
At least one savings account	20,975	597,235	61,699,176	13,766
At least one current account	20,379	582,243	60,338,017	13,559
At least £5,000 of annual income	9,050	257,198	28,816,359	6,311
At least 10 txns each month	8,138	231,105	26,486,917	5,840
At least £200 of monthly spend	7,106	200,818	23,910,680	5,336
No more than 10 active accounts	6,474	178,293	20,126,663	3,977
Complete demographic information	5,454	152,387	17,263,970	3,375
Working age	5,343	148,932	16,977,919	3,246
Final sample	5,343	148,932	16,977,919	3,246

Notes: Number of users, user-months, transactions, and transaction volume in millions of British Pounds left in our sample after each sample selection step. Link to sample selection code: [🔗](#).

observe all transactions for these months. We also drop test users, since their objectives for app use might have been different from ordinary users.<sup>1</sup>

To ensure that we observe users for at least 12 months around app signup, we require 6 months of data before the signup month, and another five months after the signup month. Our main outcome variable is netflows into a user’s savings accounts. It is thus critical that we observe enough historical data for these savings accounts to ensure that we observe all transactions during our 12 month period of interest. This is complicated by the fact that we cannot see when an account was opened at the bank, but only when it was added to the app. While cases where a user adds an account to the app as soon as it was opened are unproblematic, users will often add accounts after they were opened, either because they have accounts that they opened before signing up to the app, or because they opened new accounts after signup but add them to the app with a delay. In such cases, it is critical that, once the account is added, we observe the complete historical data up to 6 months before signup or up to the month in which the account was opened, whichever happened later. To see why this is critical, imagine a scenario where a user opens an account 10 months before they sign up to the app, makes a monthly transfer to the account of £100, adds the account to the app on signup, but we observe only 3 months of historical data. In this case we would observe that the user saved £300 before signup and £600 after, and erroneously conclude that post signup savings were twice as high. The most extreme case we need to cover is that of a user opening a savings account more than six months before signup and adding the account to the app five months after signup, in which case we need to be sure to observe 12 months of historical data. As shown in Appendix B, all major banks started providing 12 months of historical data for current and savings accounts from April 2017 onwards, which is why we restrict our sample to users who signed up in or after that month.

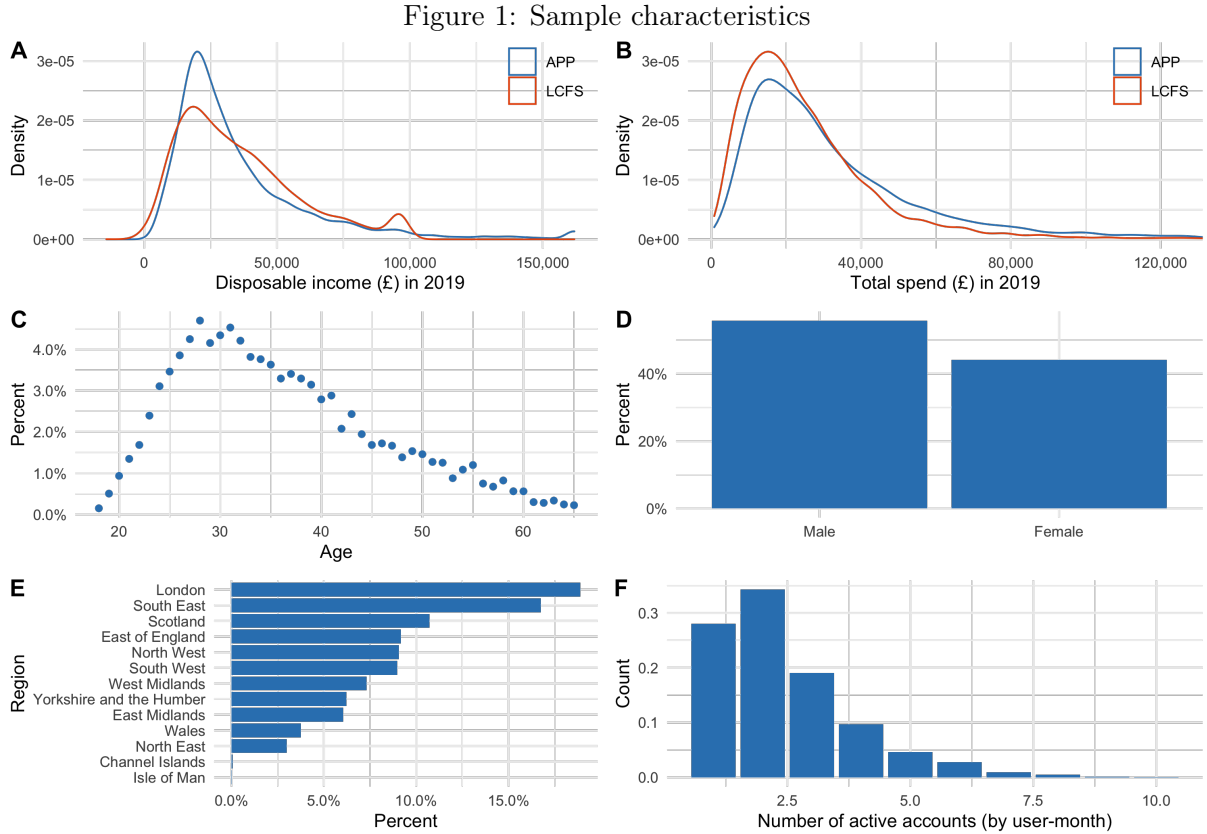
<sup>1</sup>We cannot identify test users precisely, but drop users who signed up prior to or during the first year the app was in operation.

To ensure that we can be reasonably certain to observe users have added all their financial accounts to the app, we restrict our sample to users with at least one savings and current account, with an annual income of at least £5,000, and a minimum of 10 transactions and a spend of £200 every month. To remove users who might use the app for business purposes, we drop users with more than 10 active accounts in any given month. Finally, we remove users for whom we cannot observe all demographic information we use as covariates in our analysis, and users who are not between the ages of 18 and 65, as their financial objectives are plausibly different.

**Data transformations:** To minimise the influence of outliers, we winsorise spend, income, and savings accounts flow variables at the 1 percent level or – if we winsorise on both ends of the distribution – at the 0.5 percent level.

## 2.3 Summary statistics

Figure 1 describes the sample.



Notes: Panels A and B show the distribution of disposable income and total spending in 2019, respectively, benchmarked against the 2018/19 wave of the ONS Living Cost and Food Survey (LCFS). The remaining panels show the data distributions of age, gender, region, and the number of active accounts.

Table 2 provides summary statistics.

Table 2: Summary statistics

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
user_id	504,257.1	34,391.8	442,639	475,685	500,741	529,551	575,009
ym	587.2	9.9	560	580	588	595	606
ymn	201,854.6	86.0	201,609	201,805	201,901	201,908	202,007
month	6.5	3.5	1	3	6	9	12
txns_count	113.3	60.4	10	70	102	143	332
txns_volume	18,361.4	26,224.6	256.2	6,073.4	10,381.0	18,934.2	185,464.6
month_income	2,967.2	2,294.2	417.3	1,521.1	2,256.5	3,612.6	13,490.9
inflows	761.1	2,545.1	0.0	0.0	0.0	400.0	19,732.5
outflows	753.1	2,436.4	0.0	0.0	0.0	370.0	18,250.4
netflows	17.0	2,961.7	-20,000.0	0.0	0.0	50.0	22,900.4
netflows_norm	-0.002	1.1	-7.6	0.0	0.0	0.02	7.8
inflows_norm	0.3	0.9	0.0	0.0	0.0	0.2	6.8
outflows_norm	0.3	0.9	0.0	0.0	0.0	0.2	6.9
has_pos_netflows	0.3	0.5	0	0	0	1	1
pos_netflows	422.7	2,143.4	0.0	0.0	0.0	50.0	22,900.4
user_reg_ym	590.9	5.3	581	587	591	595	601
t	0.4	0.5	0	0	0	1	1
tt	-3.7	11.1	-35	-10	-3	4	25
month_spend	2,751.0	2,625.2	200.1	1,202.6	1,987.0	3,293.2	17,050.0
age	36.7	10.1	18	29	35	43	65
is_female	0.4	0.5	0	0	0	1	1
is_urban	0.8	0.4	0	1	1	1	1
region_code	4.0	3.2	0	1	4	6	13
has_savings_account	1.0	0.0	1	1	1	1	1
has_current_account	1.0	0.0	1	1	1	1	1
generation_code	2.6	0.7	1	2	3	3	4
prop_credit	0.1	0.2	0.0	0.0	0.0	0.1	1.0
discret_spend	859.3	749.1	0.0	365.4	653.8	1,108.9	4,239.7
accounts_active	3.3	1.8	1	2	3	4	10
accounts_total	5.9	2.8	2	4	5	8	19

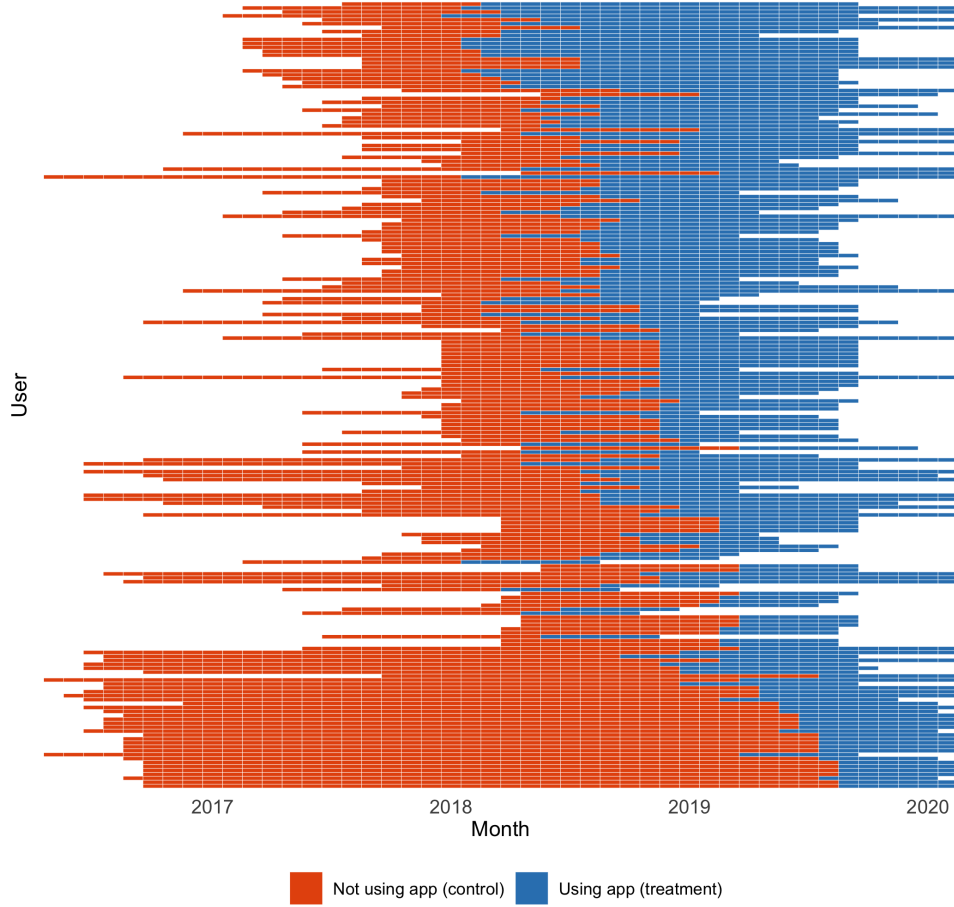
We use data from the 2018-2019 wave of the Office of National Statistics' Living Costs and Food Survey (LCFS).<sup>2</sup> Data covers the period between April 2018 and March 2019.

## 2.4 Treatment

A user changes treatment status from untreated to treated when they start using the app. Figure 2 shows the treatment history for 200 randomly selected users.

<sup>2</sup>We accessed the data via the UK Data Service at the following url: <https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8686>.

Figure 2: Treatment assignment plot



Notes: Each horizontal line shows the observed pre and post signup periods in blue and red, respectively, for one of 200 randomly selected users. The faint vertical white lines indicate month borders, whitespace indicates periods in which we do not observe the user. To the left of the observed period, this is because the app cannot access data before that point when the user signs up; to the right, because they have stopped using the app.

## 2.5 Outcomes

Savings... see Table 3 for details.

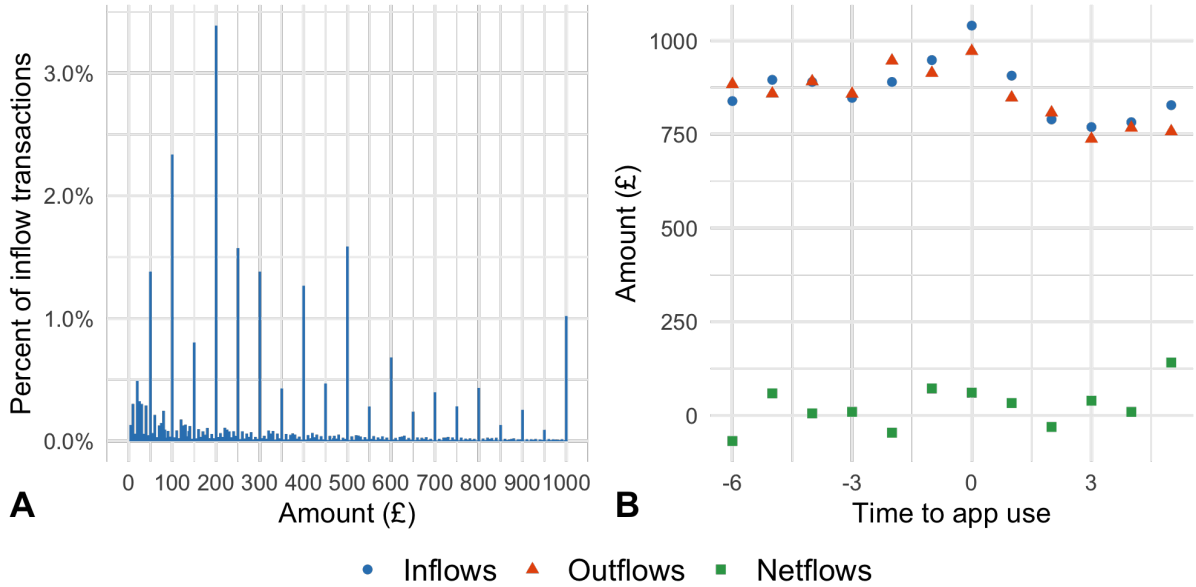
For a more nuanced understanding of how app use affects savings we also consider net-savings – total savings account inflows minus outflows – as a proportion of monthly income to see whether a willingness to save more might be offset by a (later) need to withdraw funds, and a dummy variable for whether 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.

Net savings (*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.

Positive net savings dummy (*has\_pos\_netflows*) Dummy equal to 1 if there were positive net savings (as defined above). Captures extensive margin of savings (change in number of months with positive net deposits)

Positive net savings (*pos\_netflows*) Equal to net savings if there were positive net savings. Captures intensive margin of savings (change in deposit amount in months with positive net deposits)

Figure 3: Savings patterns



Notes: Panel A shows distribution of savings account inflow amounts, making clear that most transactions are the kinds of round amounts we would expect savings transactions to be. The data is truncated at £1000. Panel B shows inflows, outflows, and netflows into savings accounts for six months before and five months after app use.

**Adjusting for multiple hypothesis tests** We think of our secondary outcomes as exploratory and do not make any adjustments for multiple hypothesis testing.<sup>3</sup> 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 11. 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 overdraft confers a benefit to the consumer.<sup>4</sup>

<sup>3</sup>For a recent game-theoretically motivated discussion of when and how to correct for multiple hypothesis testing, see Viviano et al. (2021).

<sup>4</sup>For further discussions on fees, see Jørring (2020) and Stango and Zinman (2009).

## 2.6 Covariates

We control for baseline behaviour, events, and personal characteristics that, to various degrees, capture a person’s need, capacity, motivation, and awareness to save. Table 3 lists all covariates used together with their definition and the rationale for including them. For all variables, we include contemporaneous values as well as lags for up to 6 periods. In addition, we control for the previous six months of savings to capture time-invariant unobserved drivers of savings behaviour (in specifications without fixed effects) as well as a possible signal for a higher or lower need for future savings.

Following VanderWeele (2019) we include covariates that affect either outcomes or the propensity for treatment or both, exclude from this set of variables those that are instruments (affect the outcome only through their effect on treatment propensity) and add to it proxies for unobserved variables that are a common cause of both outcomes and treatment propensity.<sup>5</sup>

The table below describes the construction and rationale for including of all variables used. The code used to construct the variables is available on [GitHub](#).


Table 3: Covariates

Variable (name in dataset)	Definition	Rationale
<b>Primary outcome</b>		
<b>Covariates</b>		
New loan dummy ( <i>new_loan</i> )	Dummy variable equal to 1 if user takes out a new loan. Calculated positive inflows of funds tagged as “loan”.	Might increase (additional funds) or decrease (need to repay) propensity to save in month of takeout and lower propensity to save in the future due to need to repay.
Unemployment benefits dummy ( <i>unemp_benefits</i> )	Dummy variable equal to 1 if user has inflow of funds tagged as “job seeker benefits”.	Might lower a user’s ability to save but increase their need for a money management app.
Monthly income ( <i>month_income</i> )	Average monthly income in a calendar year, calculated as the sum of all credits tagged income payments in said year divided by 12.	Income may alter the need and ability to save and correlate with cognitive characteristics that alter a person’s propensity to use a money management app.

## 2.7 Estimation

We choose a window from -6 to 5 for two reasons: because the longer a window we consider, the less plausible our ceteris paribus assumption is, and the longer the window, the smaller our sample.

## 2.8 Code access

We provide links to code that creates key elements of the paper such as variable definitions and sample selection directly in the relevant places in the paper so they can be accessed conveniently. The links are indicated with the GitHub logo, . The hope is that this helps the curious reader

<sup>5</sup>VanderWeele (2019) calls this the “modified disjunctive cause criterion” for covariate selection, as it includes the set of variables that are causally related to either outcomes, or treatment propensity, or both, but modified to account for potential bias by excluding instruments and including proxies of unobserved causes of both outcomes and treatment.



clarify questions about subtleties they might have while reading the paper. The complete projects GitHub repo is at [https://github.com/fabiangunzinger/mdb\\_eval](https://github.com/fabiangunzinger/mdb_eval).

### 3 Results

#### 3.1 Static TWFE

$$y_{it} = \alpha_i + \lambda_t + \beta D_{it} + \gamma X_{it} + \epsilon_{it} \quad (1)$$

Notes:

- Assumption: there are no confounding effects (either time-varying, individual varying, or individual-time varying), so treatment assignment is as good as random.
- With controls, we assume that there are no confounding variables other than the ones we control for.

Table 4: Static results

Dependent Variables: Model:	(1)	(2)	Net-inflows (3)	(4)	(5)	Discretionary spend (6)	(7)	(8)
<i>Variables</i>								
App use	49.96*** [19.11; 80.82]	14.21 [-31.50; 59.92]	36.60** [6.08; 67.12]	0.22 [-49.16; 49.61]	71.37*** [65.39; 77.36]	0.84 [-13.45; 15.13]	53.08*** [43.96; 62.19]	-22.61*** [-33.62; -11.6]
Month income	0.07*** [0.06; 0.08]	0.07*** [0.05; 0.09]	0.08*** [0.05; 0.10]	0.08*** [0.05; 0.11]	0.02*** [0.02; 0.02]	0.02*** [0.02; 0.03]	0.01*** [0.00; 0.02]	0.01*** [0.00; 0.01]
Month spend	-0.12*** [-0.13; -0.11]	-0.12*** [-0.14; -0.10]	-0.16*** [-0.18; -0.14]	-0.16*** [-0.19; -0.13]	0.16*** [0.16; 0.16]	0.16*** [0.15; 0.17]	0.12*** [0.11; 0.12]	0.12*** [0.11; 0.12]
Active accounts	38.80*** [30.04; 47.56]	37.49*** [23.46; 51.53]	70.26*** [53.08; 87.44]	68.78*** [46.71; 90.85]	46.84*** [45.14; 48.54]	42.56*** [39.57; 45.55]	86.86*** [81.78; 91.95]	75.18*** [70.63; 79.72]
Intercept	-6.05 [-41.06; 28.95]				171.35*** [164.56; 178.14]			
<i>Fixed-effects</i>								
Year-month		Yes		Yes		Yes		Yes
User ID			Yes	Yes			Yes	Yes
<i>Fit statistics</i>								
Observations	148,932	148,932	148,932	148,932	148,932	148,932	148,932	148,932
R <sup>2</sup>	0.00879	0.00932	0.03981	0.04033	0.41708	0.42534	0.66066	0.66973
Within R <sup>2</sup>		0.00869	0.00933	0.00922		0.41021	0.28163	0.23660

Signif. Codes: \*\*\*, 0.01, \*\*, 0.05, \*, 0.1

#### 3.2 Dynamic TWFE

$$y_{it} = \alpha_i + \lambda_t + \sum_{\substack{s=-6 \\ s \neq -1}}^5 \beta_s D_{its} + \gamma X_{it} + \epsilon_{it}, \quad (2)$$

where  $y_{it}$  is the outcome for individual  $i$  at time  $t$ ,  $\alpha_i$  and  $\lambda_t$  are individual and year-month fixed effects, respectively, and  $X_{it}$  is a vector of individual and time varying controls.  $D_{its}$  equals 1 if, in period  $t$ , individual  $i$  is  $s$  months away from signing up to the app. The set of  $\beta_s$  coefficients measure the effect of treatment  $s$  periods away from treatment, which is what we are interested in.

We omit the relative period indicator for period  $s = -1$  because we need to omit one relative period indicator to avoid perfect collinearity among the period indicators, and we choose the last

pre-treatment period because it serves as a natural benchmark against which to compare the outcomes in other periods.<sup>6</sup>

Sun and Abraham (2021) define an event study design as a staggered adoption design where units are treated at different times and where there may or may not be never treated units. In our case, we have no never treated units, and treatment is absorbing in that once a unit is treated they will also be treated in all subsequent units.<sup>7</sup>

Setup:

- We observe  $N + 1$  units for  $T + 1$  periods and, for each  $i \in \{0, \dots, N\}$  and  $t \in \{0, \dots, T\}$  observe outcome  $y_{it}$  and treatment status  $D_{it} \in \{0, 1\}$ , where  $D_{it}$  equals 1 if unit  $i$  is treated in period  $t$  and 0 otherwise.
- We can uniquely characterise treatment paths by the time period of initial treatment, denoted as  $E_i = \min\{t : D_{it} = 1\}$ .
- We can group units into cohorts  $e \in \{0, \dots, T\}$ , where units in cohort  $e$  were all first treated at time  $e$ , so that  $\{i : E_i = e\}$ .
- We define  $y_{it}^e$  as the potential outcome in period  $t$  if unit  $i$  was first treated in period  $e$ .

Assumptions:

- Observations  $\{y_{it}, D_{it}\}_{t=0}^T$  are independent.
- A1: parallel trends: difference in baseline outcomes over time do not differ between treatment cohorts. Not obviously violated in our context. Early adopters might differ from late adopters, but difference might plausibly be constant over time. - We don't have never treated units, so Ashenfelter dip scenario is not a problem, even though we seem to observe something like this in discretionary spending graph (increase in disc spend before signup)
- A2: no anticipatory behaviour. Plausibly violated if people are motivated to save more and start doing so even before app use. Can test for whether there is a peak prior to signup. Because our units have private knowledge about future of treatment path (their intention to reduce spending and save more and sign up to an app), this might be violated. The trajectories of discret spend and net inflows are conflicting on this, though, suggesting that they increase discret spend in runup to app use (which might provide motivation to eventually sign up) but also might have increased net savings slightly.
- A3: treatment effect homogeneity across all cohorts and all relative periods. (Note: treatment effects can be dynamic, but need to be the same across cohorts). We could test for this.

---

<sup>6</sup>As Sun and Abraham (2021) point out, there are two sources of perfect multicollinearity when estimating a fully dynamic model (i.e. one including all possible lags). The first results from all relative period indicators summing to 1 in each period, so that the entire set of relative period dummies across all time periods is perfectly multicollinear. We deal with this by excluding the indicator for  $s = -1$ . The second issue arises from the fact that for initial treatment period  $E_i$ ,  $t = s + E_i$ . We deal with this issue by "trimming" our sample to be balanced in relative periods by only using data from relative periodl  $\{-6, 5\}$ . Both of these approaches are standard in the empirical literature.

<sup>7</sup>We cannot rule out that some users who stopped using the app and closed their account rejoined later on, in which case they would appear in our dataset as a new user. However, we can plausibly assume that such cases are rare.

Notes:

- Comparison: pre vs post signup within each individual.
- Assumption: there are no time-varying unobserved effects that affect both y and D (formally:  $E[Du] = 0$ , since u is by definition correlated with y.
- 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, exposure to something that motivates saving or change in financial behaviour).
- See Imai and Kim (2021) for problems with twfe

Table 5: Dynamic results

Dependent Variables: Model:	(1)	(2)	Net-inflows (3)	(4)	(5)	Discretionary spend (6)	(7)
<i>Variables</i>							
Months to/since app use = <-6	-110.84*** [-194.27; -27.42]	-103.66* [-208.63; 1.30]	-98.69** [-196.09; -1.30]	-89.62* [-195.15; 15.91]	-125.40*** [-141.55; -109.25]	-78.16*** [-90.83; -65.49]	-118.89*** [-133.11; -104.67]
Months to/since app use = -5	-35.38 [-147.30; 76.55]	-33.48 [-104.30; 37.33]	-32.47 [-157.19; 92.26]	-29.27 [-101.94; 43.41]	-75.75*** [-97.42; -54.08]	-54.58*** [-75.39; -33.77]	-70.99*** [-86.97; -54.99]
Months to/since app use = -4	-84.56 [-196.42; 27.31]	-72.77 [-210.15; 64.60]	-83.59 [-209.56; 42.39]	-70.73 [-199.69; 58.23]	-57.64*** [-79.30; -35.98]	-35.26*** [-51.02; -19.50]	-55.29*** [-70.63; -39.95]
Months to/since app use = -3	-71.07 [-182.90; 40.75]	-62.18 [-170.79; 46.43]	-70.38 [-199.96; 59.20]	-60.24 [-173.35; 52.87]	-30.29*** [-51.94; -8.63]	-14.47 [-32.82; 3.87]	-28.69*** [-43.89; -13.49]
Months to/since app use = -2	-122.75** [-234.56; -10.93]	-119.55** [-225.23; -13.87]	-122.03** [-243.52; -0.54]	-118.23** [-216.49; -19.96]	-4.43 [-26.08; 17.22]	-1.01 [-18.48; 16.47]	-2.29 [-17.65; 13.07]
Months to/since app use = 0	-21.42 [-133.23; 90.40]	-11.68 [-111.40; 88.05]	-26.70 [-154.64; 101.24]	-15.70 [-113.59; 82.19]	-45.13*** [-66.78; -23.48]	-31.21*** [-47.29; -15.14]	-51.69*** [-66.74; -36.64]
Months to/since app use = 1	-58.47 [-170.30; 53.36]	-62.50 [-182.33; 57.34]	-62.12 [-189.27; 65.03]	-63.71 [-176.44; 49.01]	-40.35*** [-62.01; -18.70]	-34.83*** [-53.64; -16.02]	-44.49*** [-60.03; -28.95]
Months to/since app use = 2	-122.16** [-234.00; -10.31]	-128.48** [-241.72; -15.24]	-124.56** [-245.46; -3.66]	-128.78** [-240.72; -16.84]	-27.21*** [-48.87; -5.56]	-34.63** [-61.44; -7.83]	-30.19*** [-46.11; -14.27]
Months to/since app use = 3	-55.31 [-167.18; 56.55]	-66.69 [-184.08; 50.71]	-57.53 [-176.91; 61.85]	-66.80 [-174.31; 40.72]	-27.30** [-48.96; -5.64]	-43.86*** [-64.72; -22.99]	-30.19*** [-46.04; -14.34]
Months to/since app use = 4	-85.94 [-197.82; 25.94]	-97.30 [-219.28; 24.68]	-87.90 [-207.50; 31.69]	-96.92 [-219.67; 25.83]	-23.89** [-45.55; -2.22]	-37.71*** [-60.17; -15.24]	-26.99*** [-42.85; -11.13]
Months to/since app use = 5	46.03 [-65.87; 157.93]	33.02 [-100.42; 166.46]	44.40 [-76.21; 165.01]	34.50 [-95.93; 164.93]	-25.47** [-47.13; -3.80]	-36.66*** [-56.31; -17.02]	-28.89*** [-45.14; -12.64]
Months to/since app use = >5	-39.39 [-125.16; 46.38]	-63.30 [-171.70; 45.11]	-40.34 [-138.79; 58.11]	-62.71 [-176.22; 50.79]	-13.55 [-30.16; 3.05]	-41.94*** [-59.32; -24.57]	-22.49*** [-37.23; 2.24]
Month income	0.07*** [0.06; 0.08]	0.07*** [0.05; 0.09]	0.08*** [0.05; 0.10]	0.08*** [0.05; 0.11]	0.02*** [0.02; 0.02]	0.02*** [0.02; 0.03]	0.03*** [0.01; 0.05]
Month spend	-0.12*** [-0.13; -0.11]	-0.12*** [-0.14; -0.10]	-0.16*** [-0.18; -0.14]	-0.16*** [-0.19; -0.13]	0.16*** [0.16; 0.16]	0.16*** [0.15; 0.17]	0.12*** [0.11; 0.13]
Active accounts	36.80*** [27.80; 45.80]	36.50*** [22.39; 50.60]	67.24*** [49.37; 85.11]	67.15*** [45.29; 89.01]	42.73*** [40.98; 44.47]	41.25*** [38.24; 44.26]	78.79*** [73.58; 84.00]
Intercept	93.70** [6.98; 180.42]				276.72*** [259.93; 293.52]		
<i>Fixed-effects</i>							
Year-month		Yes		Yes		Yes	Yes
User ID			Yes	Yes			Yes
<i>Fit statistics</i>							
Observations	148,932	148,932	148,932	148,932	148,932	148,932	148,932
R <sup>2</sup>	0.00892	0.00944	0.03993	0.04044	0.41920	0.42596	0.60000
Within R <sup>2</sup>		0.00881	0.00945	0.00933		0.41085	0.28000

Signif. Codes: \*\*\*, 0.01, \*\*, 0.05, \*, 0.1

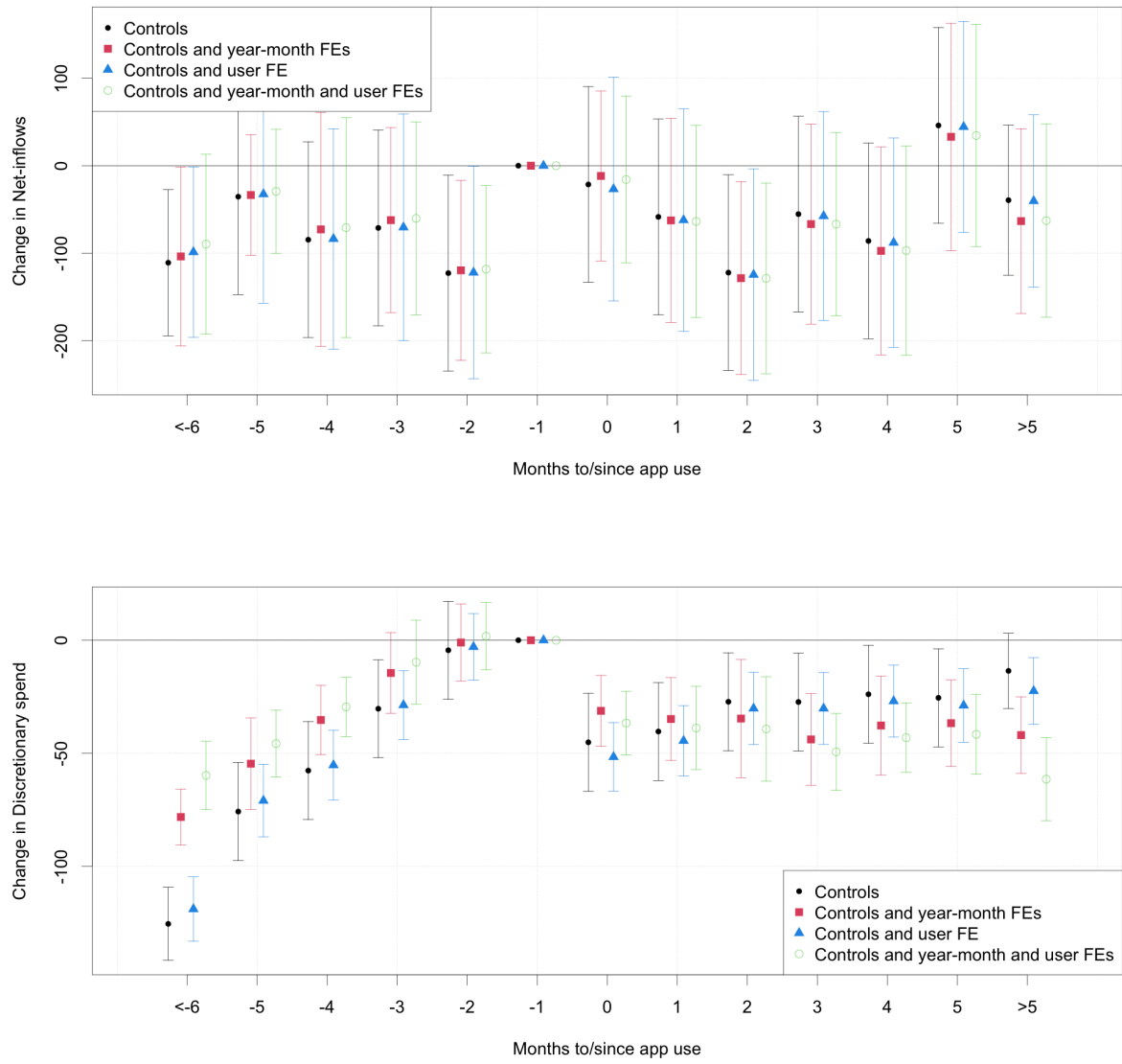


Figure 4: Dynamic results

### 3.3 Decomposing inflows and outflows

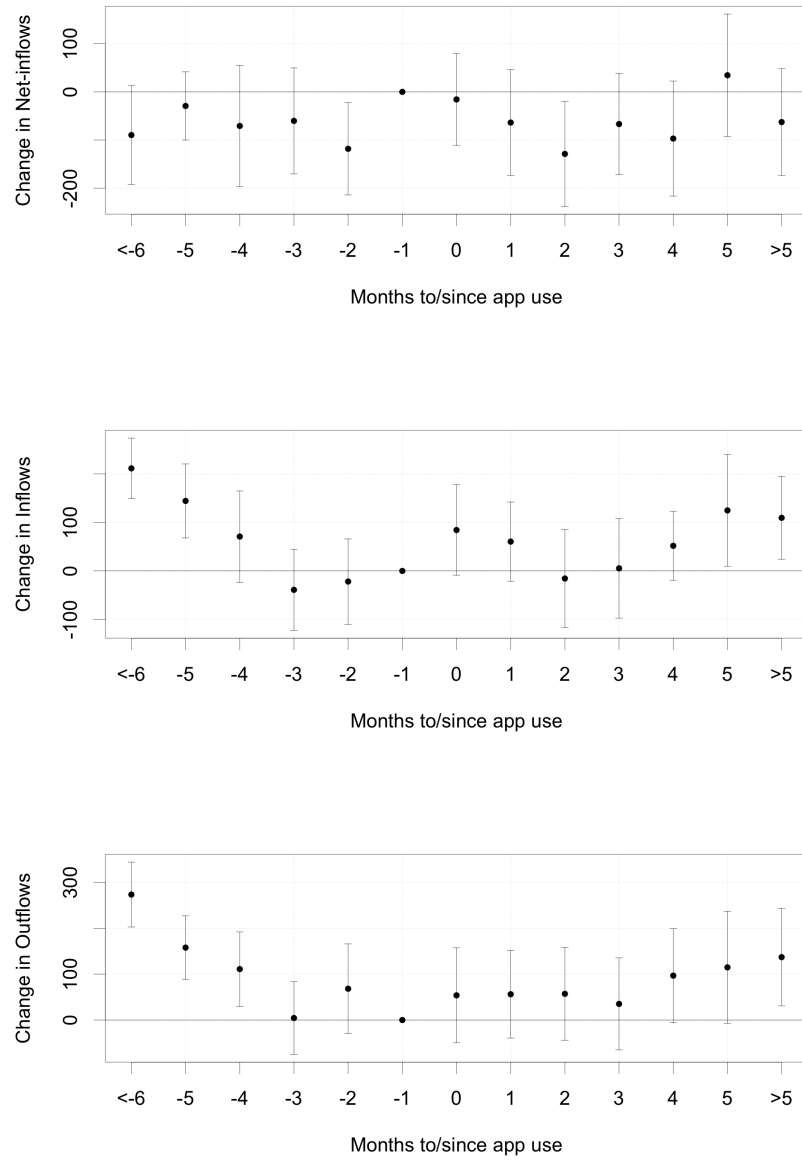


Figure 5: Decomposing inflows and outflows

### 3.4 Decomposing intensive and extensive margins

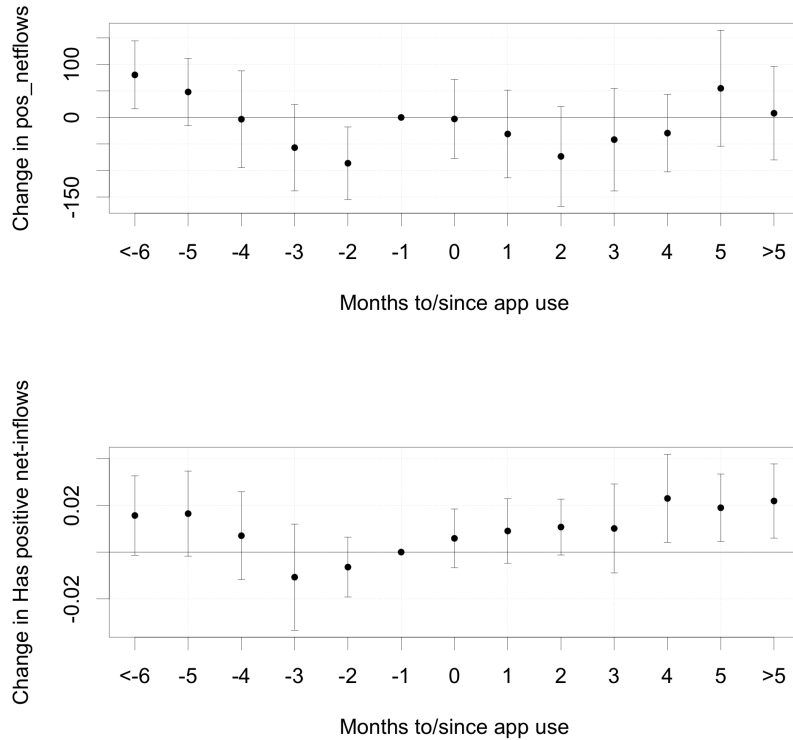


Figure 6: Decomposing intensive and extensive margin

### 3.5 Matching

Control group design:

- We only have data for a self-selected group of people who choose to use the app. This has a couple consequences:
- 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 app use helps the average person in the population as a whole save more.<sup>8</sup>
- Even among people who do eventually sign up to the app, the decision when to do so is unlikely to be random – *something* makes them sign up at the particular point in time they do and not before or after. If we think of this factor as “motivation to save more”, then said motivation is inextricably linked with the decision to sign up so that we cannot differentiate between the two.

<sup>8</sup>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 in 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.

- Hence: due to the first point above, we cannot estimate an ATE (effect of app use on the average citizen), and due to the second point we also cannot estimate a pure ATT (effect of app use on users). Instead, our estimated effect of app use captures the effect of being motivated to save more and using the app to do so.<sup>9</sup>
- This is true for both our matched DiD and our TWFE design. While these two approaches use a different counterfactual to estimate the effect of app use (behaviour of a matched control in the case of DiD and extrapolating within-user pre-signup behaviour in the case of TWFE), neither can help us with the fact that the decision to sign up is likely correlated with the time-varying unobserved effect “motivation to save more”.

DiD:

- 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 treatment 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 using 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  $F = L$ ).

DiD identification assumptions:

- No spillover effects: the potential outcome of unit  $i$  at time  $t$  is independent of the treatment status of other units. This is violated if a user’s partner or friends also use the app and, through sharing their experiences or motivations, influence the user’s savings behaviour.
- Carryover effects no longer than  $L$  periods: a user’s potential outcome in period  $t$  is independent of treatment status in periods more than  $L$  periods ago. Given that we are dealing with an absorbing treatment, this is not a very strong assumption in our context, and we choose  $L$  based on what we think is an informative number of periods to observe pre-app use behaviour.<sup>10</sup>

---

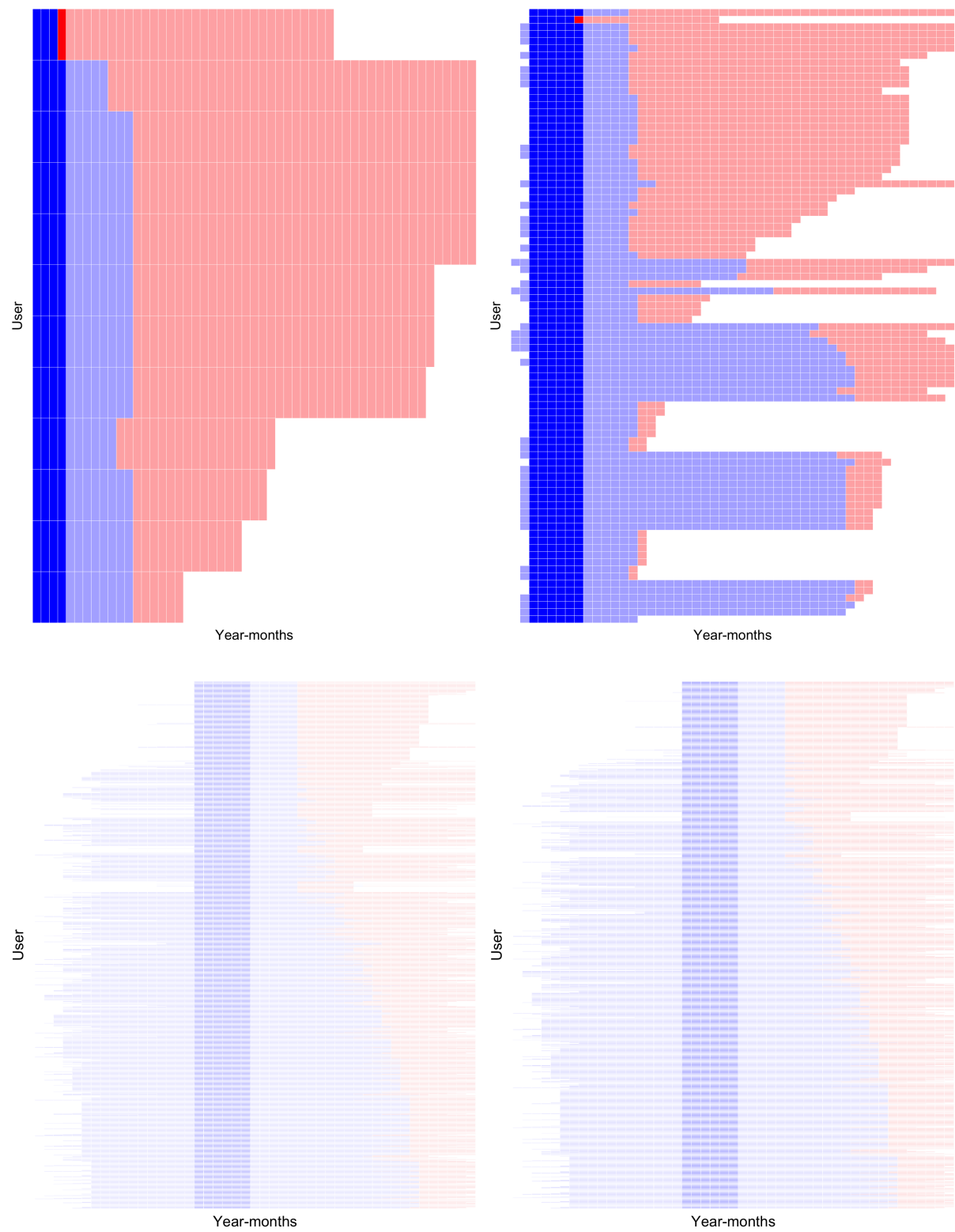
<sup>9</sup>One way to get a step closer to ATT would be to find a variable that correlates with “motivation to save” and control for it / match on it.

<sup>10</sup>An absorbing treatment is one that cannot be reversed, and hence we only change from untreated to treated once.

- Parallel trends: the spending trajectory between treated and control units would have continued to be parallel if the treated unit hadn't started using the app. This is violated whenever an intended change in savings behaviour also provided the impetus for the user to start using the app, which is likely to occur frequently. To the extent this is the case, we have an omitted variable "motivation to save more", which both changes the user's savings behaviour and their treatment status. Because of this, what we are measuring is not a pure ATT of app use – the effect of app use on savings over and above the change precipitated by a change motivation – but the effect of app use for users motivated to save more.

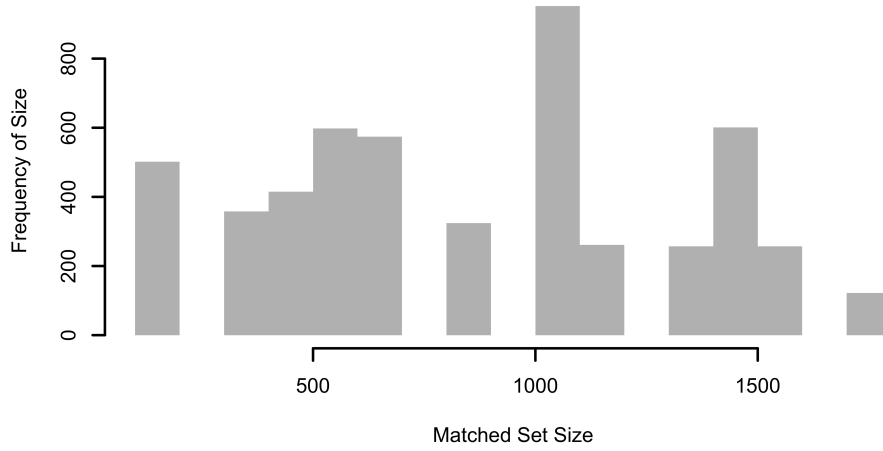


Figure 7: Match set examples



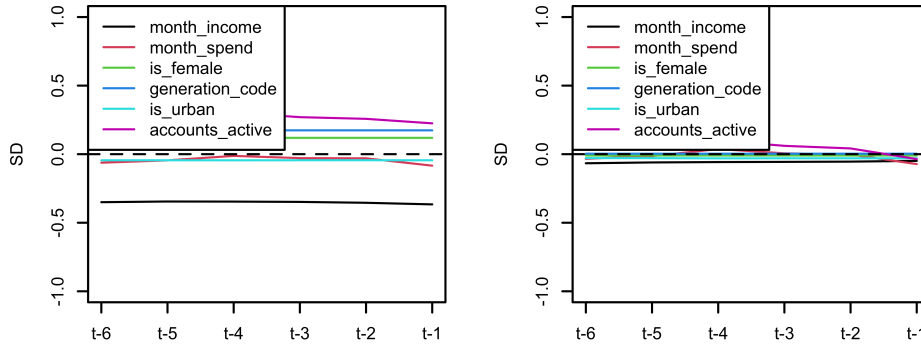
Notes:

Figure 8: Distribution of size of matched control units



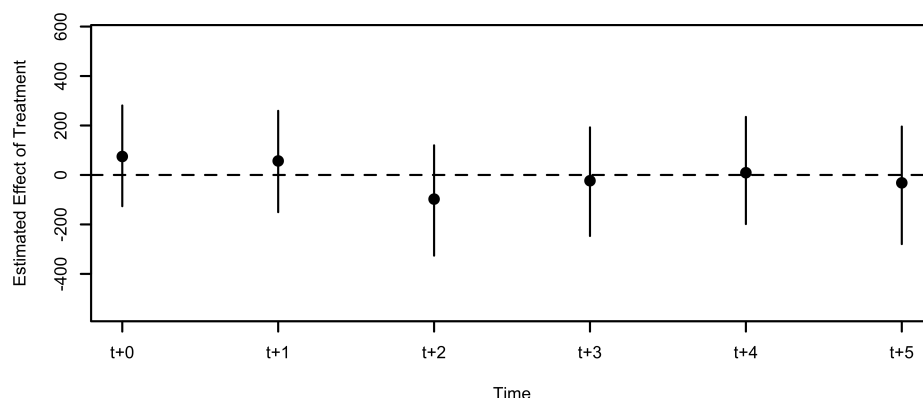
Notes: In the first step of the matching procedure, each user gets assigned a set of potential control users that share the same treatment history for a specified number of periods before the user signs up to the app (6 months, in our baseline specification), but that do not sign up themselves for a specified number of periods after the treatment user has signed up (another 6 months, in our baseline specification). The figure shows the distribution of the sizes of these sets of potential control users. The pink vertical bar on the left shows the count of users for whom no control users could be found.

Figure 9: Covariance balance



Notes: Average covariate standard deviation between treatment and control units for each pre-treatment period using the entire set of potential controls on the left and, on the right, the refined set of controls, which, in our baseline specification, consists of the nearest neighbour match based on the propensity score.

Figure 10: Matching estimates



Notes:

Is estimate causal?

- King and Zeng (2006) show that there are four sources of bias (omitted 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).

### 3.6 Alternative window lengths

### 3.7 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.<sup>11</sup>

Subgroup analysis: same Fig and Tab as in main analysis, but with line for each subgroup. One figure for each of: gender, generations, income terciles, pre-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)

### 3.8 Alternative outcome variables

Look at netflows scaled by income

## 4 Discussion

Limitations:

- Can't say whether increase in savings was achieved by going into debt elsewhere

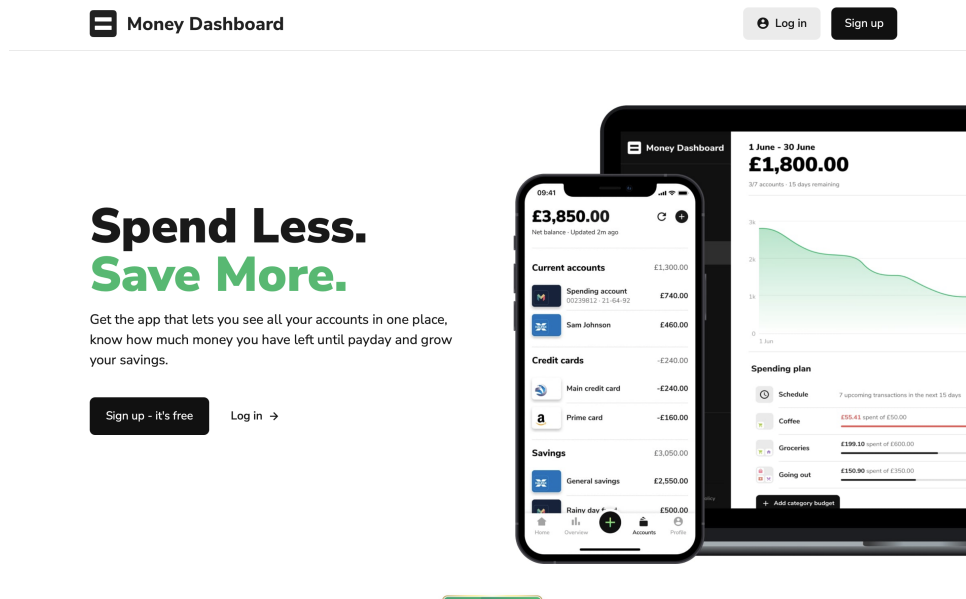
<sup>11</sup>Based on age ranges provided by Beresford Research.

## References

- 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.
- 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.
- VanderWeele, Tyler J (2019). “Principles of confounder selection”. In: *European journal of epidemiology* 34.3, pp. 211–219.
- 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 11: Money Dashboard website screenshot



Notes: Screenshot from the top of the Money Dashboard website, at [moneydashboard.com](https://moneydashboard.com), accessed on 29 April 2022.

## B Data

### B.1 Historical data availability

Translate jupyter notebook fig to R.