Evaluation

Fabian Gunzinger*
Warwick Business School

August 12, 2022

Contents

1 Introduction **3**

2 Methods **3**
- 2.1 Dataset . . . 3
- 2.2 Data preprocessing . . . 3
- 2.3 Summary statistics . . . 5
- 2.4 Treatment . . . 6
- 2.5 Outcomes . . . 7
- 2.6 Covariates . . . 9
- 2.7 Code access . . . 9

3 Results **9**
- 3.1 Baseline results . . . 10
- 3.2 Conditional parallel paths . . . 11
- 3.3 Relaxing anticipation assumption . . . 12
- 3.4 Unbalanced aggregation . . . 13
- 3.5 Robustness . . . 13

4 Old stuff **13**
- 4.1 Decomposing intensive and extensive margins . . . 15
- 4.2 Subgroups . . . 16

5 Discussion **16**

A Money Dashboard application **18**

B Additional specifications **18**
- B.1 Static TWFE . . . 18
- B.2 Additional TWFE specifications . . . 19
- B.3 Decomposing inflows and outflows . . . 21

*fabian.gunzinger@warwick.ac.uk

B.4 Twoway clustering 22
B.5 Longer event windows 22

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 pensioneers, 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
App signup after March 2017	81,309	1,981,147	181,596,348	34,993
At least 6 months of pre and post signup data	37,961	1,085,215	107,354,457	22,153
At least one savings account	26,412	753,505	78,015,160	17,252
At least one current account	25,675	734,952	76,340,243	16,994
At least £5,000 of annual income	11,379	323,729	36,380,564	7,895
At least 10 txns each month	10,237	291,337	33,455,079	7,305
At least £200 of monthly spend	8,945	253,547	30,189,093	6,667
No more than 10 active accounts	8,143	224,986	25,427,757	4,945
Complete demographic information	6,876	192,606	21,855,775	4,189
Working age	6,739	188,324	21,498,631	4,033
Final sample	6,739	188,324	21,498,631	4,033

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.[1]

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 perdiod 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 transfter to the account of £100, adds the account to the app on signup, but we onserve 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 **??**, 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.

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

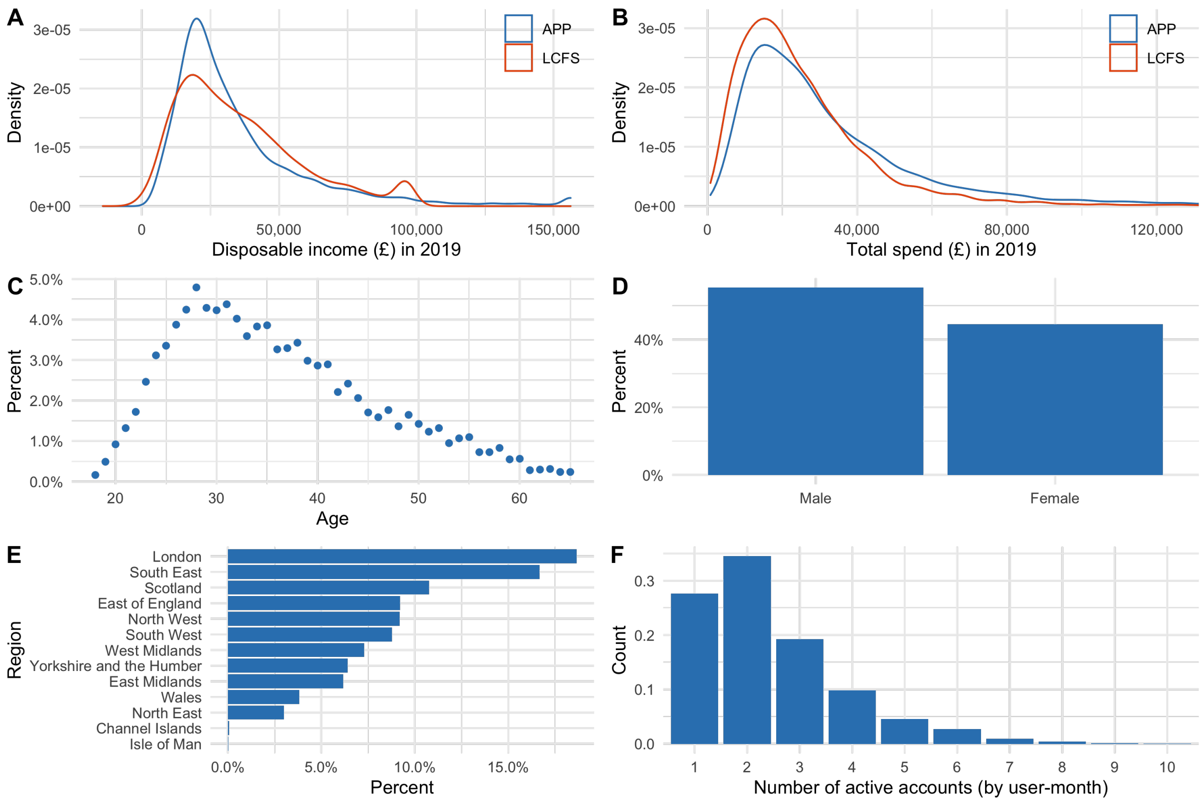
To ensure that we can be reasonably certain to observe users have added all their financial accounts to the app, we restrict our sample to users with at least one savings and current account, with an annual income of at least £5,000, and a minimum of 10 transactions and a spend of £200 every month. To 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.

Figure 1: Sample characteristics



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	503,356.1	34,778.6	441,309	474,770	499,754	528,755	575,009
ym	587.2	9.9	560	580	588	595	606
ymn	201,854.3	86.1	201,609	201,805	201,901	201,908	202,007
month	6.5	3.5	1	3	6	9	12
txns_count	113.5	60.4	10	71	102	143	335
txns_volume	18,110.9	25,559.3	227.1	6,036.4	10,291.1	18,875.1	180,795.9
month_income	2,931.6	2,238.4	417.3	1,510.9	2,243.2	3,567.6	13,016.6
inflows	748.7	2,498.0	0.0	0.0	0.0	400.0	19,353.4
outflows	738.6	2,388.1	0.0	0.0	0.0	370.0	17,933.6
netflows	15.2	2,894.2	−19,940.9	0.0	0.0	50.0	22,000.0
netflows_norm	−0.001	1.0	−7.5	0.0	0.0	0.02	7.7
inflows_norm	0.3	0.9	0.0	0.0	0.0	0.2	6.7
outflows_norm	0.3	0.9	0.0	0.0	0.0	0.2	6.7
has_pos_netflows	0.3	0.5	0	0	0	1	1
pos_netflows	413.9	2,078.5	0.0	0.0	0.0	50.0	22,000.0
user_reg_ym	590.8	5.4	580	587	590	595	601
t	0.4	0.5	0	0	0	1	1
tt	−3.6	11.2	−35	−10	−3	4	26
month_spend	2,742.6	2,613.1	200.1	1,198.0	1,974.8	3,288.0	16,936.3
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
dspend	859.4	748.0	0.0	365.6	653.6	1,109.4	4,236.4
dspend_count	39.8	27.9	0	21	35	52	465
dspend_mean	24.7	42.1	0.2	13.8	19.5	27.6	7,508.2
accounts_active	3.3	1.8	1	2	3	4	10
accounts_total	5.8	2.8	2	4	5	8	19

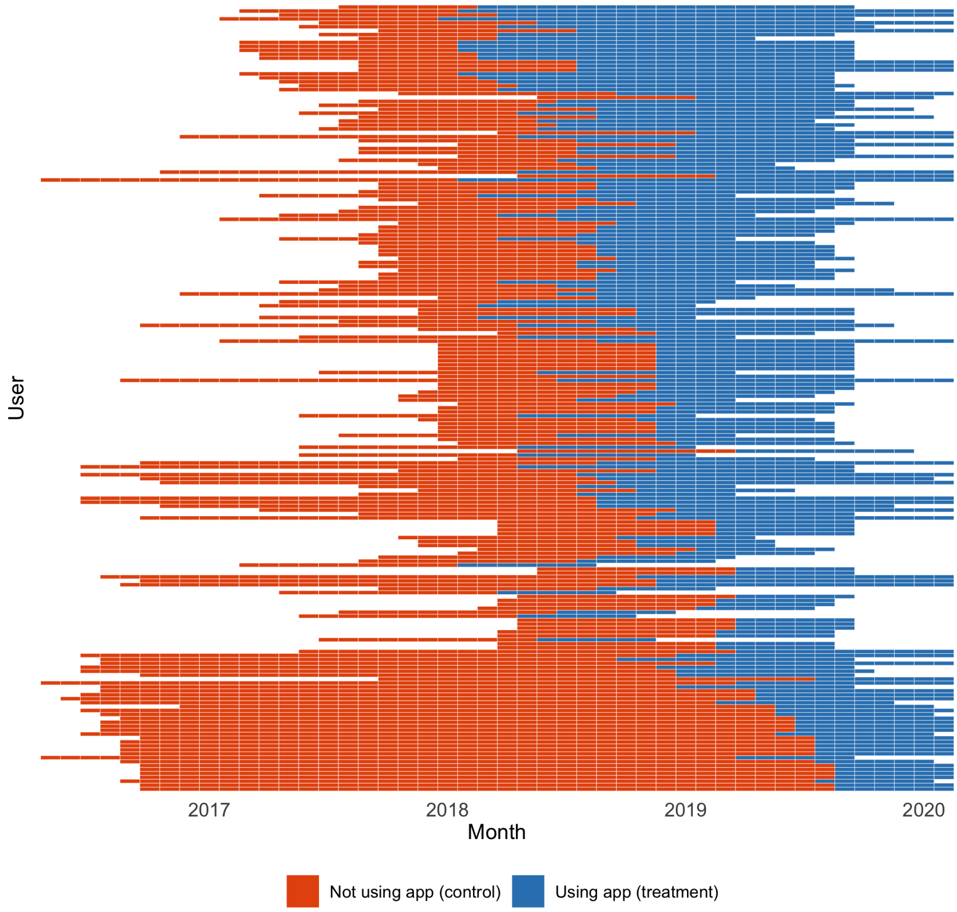
We use data from the 2018-2019 wave of the Office of National Statistics' Living Costs and Food Survey (LCFS).[2] Data covers the period between April 2018 and March 2019.

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.

[2]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.

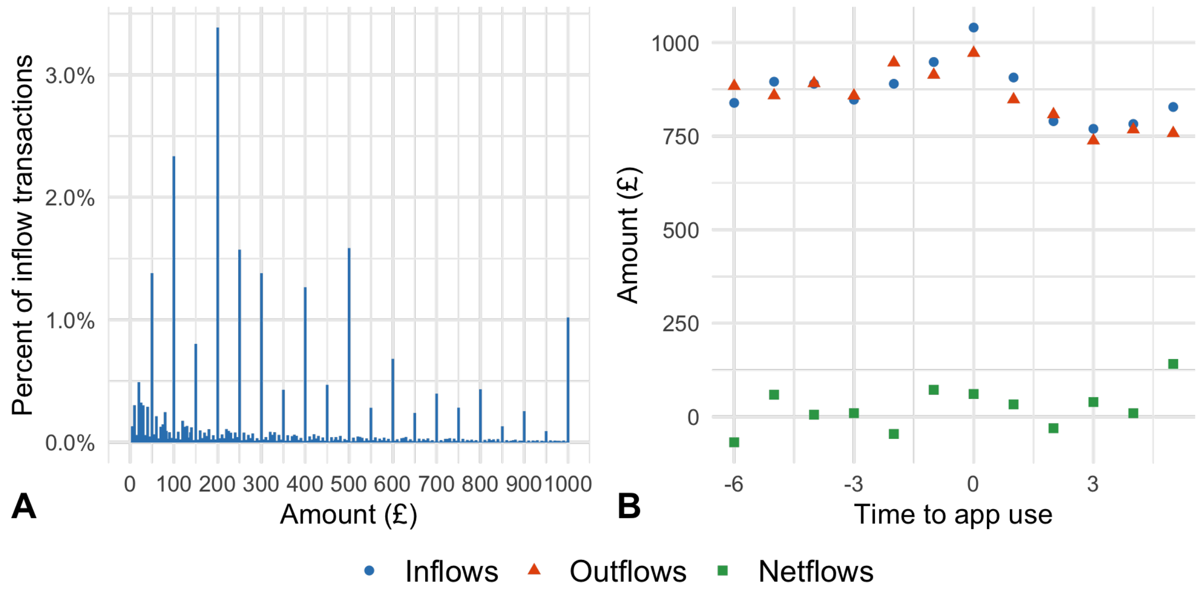
For a more nuanced understanding of how app use affects savings we also consider net-savings – total savings account inflows minus outflows – as a proportion of monthly income to see whether a willingness to save more might be offset by a (later) need to withdraw funds, and a dummy variable for whether 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 hows 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.[3] 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.[4]

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

[4]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.[5]

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
	Primary outcome	
	Covariates	
New loan dummy (*new_loan*)	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 (*unemp_benefits*)	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 (*month_income*)	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 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 clarify questions about subtleties they might have while reading the paper. The complete projects GitHub repo is at https://github.com/fabiangunzinger/mdb_eval.

3 Results

SA setup:

[5] 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.

- Units that are treated in period 1 are dropped from the sample, since (i) there exists no possible control group based on which to identify their treatment effects and (ii) they are not useful as a control group themselves. (wp footnote 3)

- If there exists no never-treated units (as in our setup), then last treated group is dropped since there exists no valid "not-yet-treated" comparison group for them.

- This means that if we restrict our data to a balanced panel (e.g. all data from 2018 and 2019), the groups first treated in Jan 2018 and Dec 2019 will be dropped from the sample.

- Assumption 3 allows for known treatment anticipation. In our case, with units self-selecting into treatment, it will be useful to do robustness checks for sensible anticipation levels. (In our case, the relevant anticipation period is time since deciding to save more/spend less and time when signing up to app. This could realistically be up to 6 months). As CS point out in remark 1, increading delta makes the parallel trends assumption more restrictive, since it now needs to hold also for delta pre-treatment periods. Our results here are weird, though: if we think that during anticipation period peope already spend less, then this would bias baseline results downwards. Instead, we see the opposite. In the case with delta $>$ 0, this also means we can identify GT effects only up to time where last cohort effectively starts its treatment (max g - delta), and, as discussed before, we cannot estimate GT for that last group, either.

- Conditional parallel trends are important in contexts when (i) there are covariate specific trends in outcome paths and (ii) the distribution of covariates differs between groups. (E.g. people who sign up to job training differ from those who don't and job outcomes depend on these covariates). In my context, do covariates differ between those who sign up and those who don't? (can investigate - month income, num accounts, total spend, etc. pre and post signup), and does savings path depend on these variables (we'd certainly think so).

- Assumption 5 restricts pre-treatment trends (see discussion there and discuss in footnote).

3.1 Baseline results

Assumptions:

- A1: Absorbing treatment. This is unproblematic.

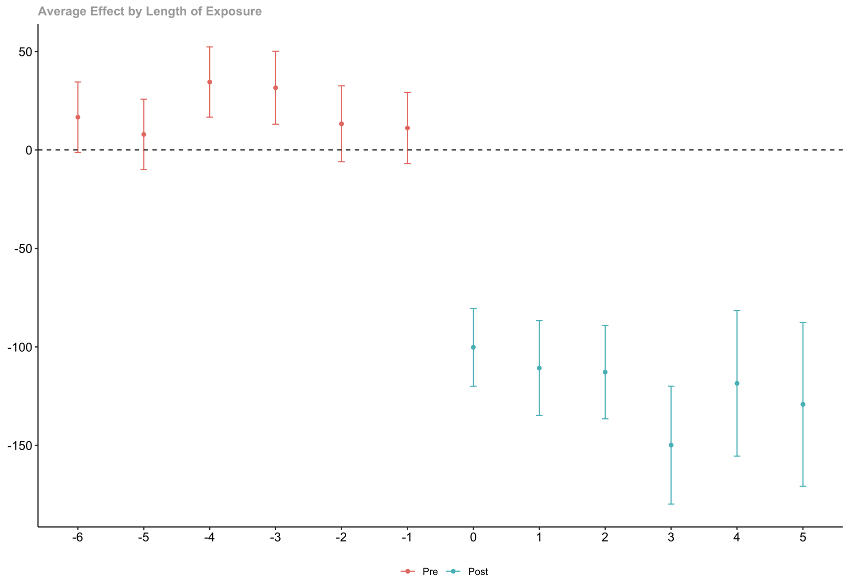
- A2: Random sampling (each unit is independently drawn from a larger population of interest). This is not self-evidently true. In a narrow sense, MDB users are different by virtue of having signed up. One way to think of a super population is to think about knowing of MDB's existence to some extent as random, in which case the super population consists of all people who would have signed up if they had heard of MDB. Another is for signup to be partially driven by need, and need to be partially random. The super population is than everyone with a potential need in the future.

- A3: Limited (known) treatment anticipation. In baseline, we assume that there is no anticipation. We provide robustness checks with different anticipation periods.

- A5: Unconditional parallel trends based on not-yet treated groups.
- A6: Overlap condition: positive fraction of population starts treatment in each period, and propensity scores are bounded away from 1 for all groups and times.
- We aggregate GT effects using a panel balanced in event times, with all units being greated for at least 5 periods. This avoids that the event study parameters are influenced by compositional issues (SA show in section 3.1.1 that without this restriction, aggregated treatment effects also are affected by different composition of groups and different group weights). This comes at the cost of fewer groups used for calculations. And in appendix we compare the results without this restriction.

Allows for: Heterogeneous treatment across time and units

Figure 4: New results



Notes: ...

3.2 Conditional parallel paths

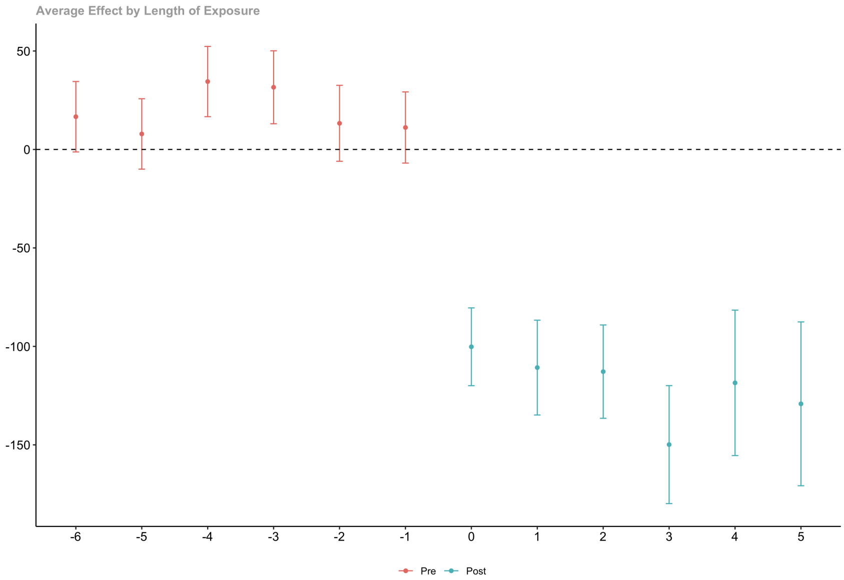
Assumption:

- Above, we assume that parallel paths hold for all units at all times.
- Here, we assume they hold for all units at all times with the same characteristics. (SA are the first to show that this works.)

Notes:

- We wouldn't expect conditional parallel trends. If we think that discretionary spend is constant fraction of income or total spend, then this implies parallel trends in the variables we control for.
- Same with number of accounts: average spend from account is constant, then dspend has parallel trends in the number of accounts.

Figure 5: New results

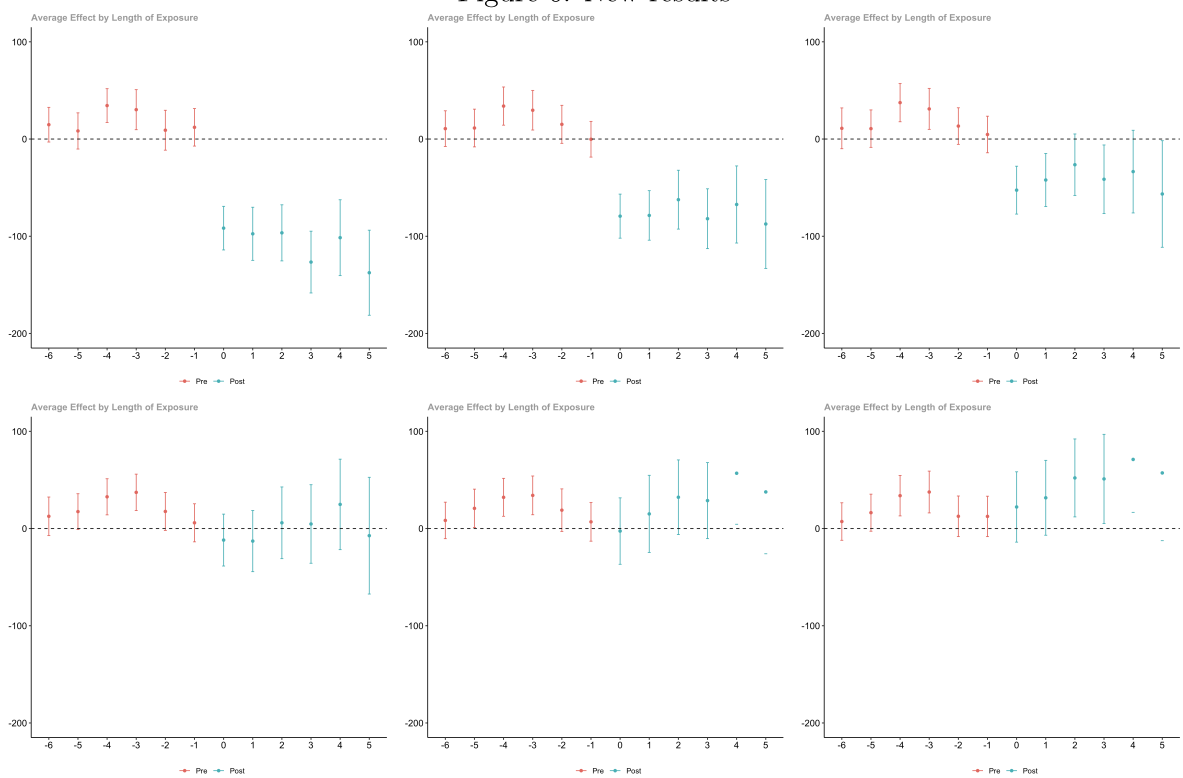


Notes: ...

3.3 Relaxing anticipation assumption

See callaway2021difference-appendix for discussion.

Figure 6: New results



Notes: ...

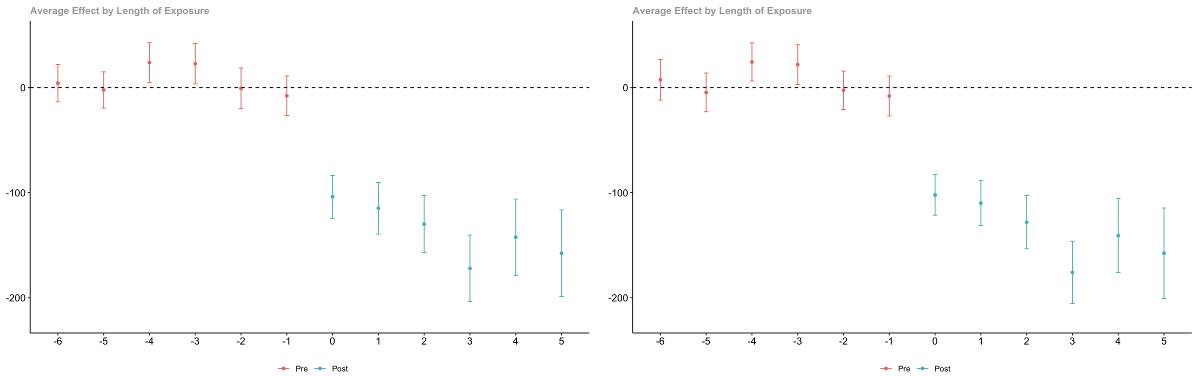
3.4 Unbalanced aggregation

The baseline specification relies on a panel balanced in event time and thus only includes groups that have been exposed to treatment for at least 5 periods. Figure 7 reproduces the baseline results in the left panel and compares it with results based on the full sample.

As discussed in Callaway and Sant'Anna (2021) (section 3.1.1), these two approaches entail a trade-off. When using the full sample, the aggregated parameters are a function of the weighted average treatment effects for each group e periods after treatment (which is what we want) as well as compositional changes due to different groups being included for different periods e and different weights attached to these groups. While parameters aggregated using a panel balanced in event time do not suffer from compositional and weighting changes, but are calculated based on a smaller number of groups.

As expected, using the full data reduces the size of the confidence intervals. But the results are otherwise very similar.

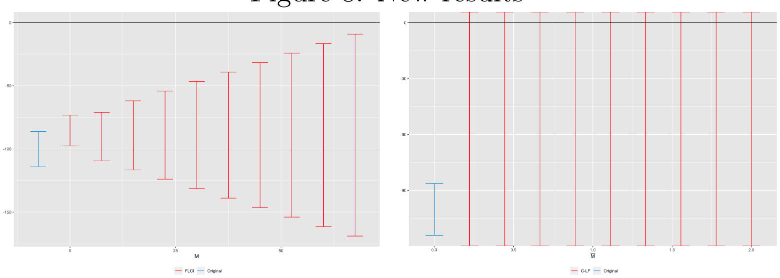
Figure 7: Comparison of balanced and unbalanced aggregations



Notes: ...

3.5 Robustness

Figure 8: New results



Notes: ...

4 Old stuff

We use an event study design of the following form:

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

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

We normalise our treatment effects by expressing them relative to reference period $s = -1$, which is necessary to avoid perfect multicollinearity between the relative period indicators (which sum to 1 for each user-month observation) and the user-level fixed effect $alpha_i$ (which also equals 1). We follow common practice and normalise relative to the last period before treatment exposure because this serves as a natural benchmark for post-treatment behaviour.

To avoid multicollinearity issues and uniquely identify the lead and lag effects, we omit relative period indicator $s = -1$,

and bin the endpoints of our observation window so that $D_{it}^{-6} = 1$ for all periods six or more periods prior to signup and $D_{it}^{5} = 1$ for all months five or more months after the month of signup. 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.

Normalising:

This expresses all estimates relative to the omitted period, which – given that it is the last month before signup – serves as a natural benchmark.

Binning: Binning endpoints assumes that coefficients remain constant outside the observation window.

Both of these approaches are standard in the literature, as discussed in Sun and Abraham (2021) and Schmidheiny and Siegloch (2019).

Assumptions:

- Homogenous treatment paths across cohorts.
- Strict exogeneity of treatment assumption doesn't hold due to self-selection into app use. Consequence: can't estimate causal effect of app use.

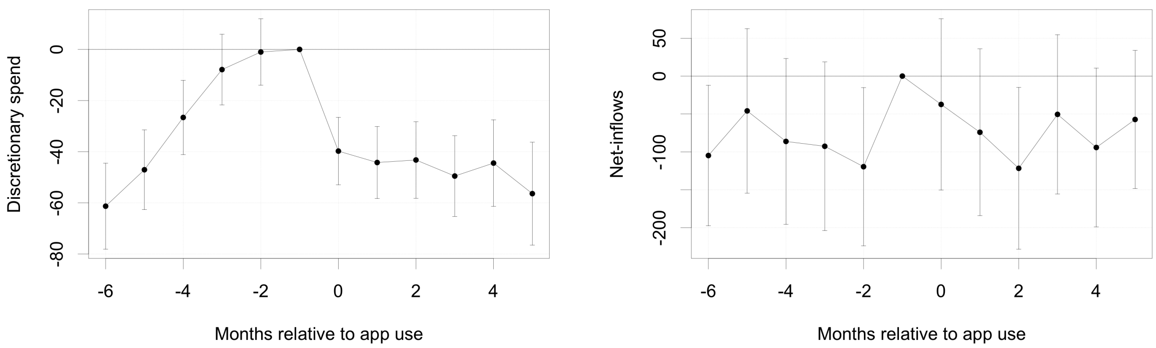
Notes:

- Sun and Abraham (2021) show that event study design is fine even in the presence of no untreated units, as long as dynamic trajectory of effect is the same for all cohorts.
-

Time fixed effects:

- Using ym vs m.

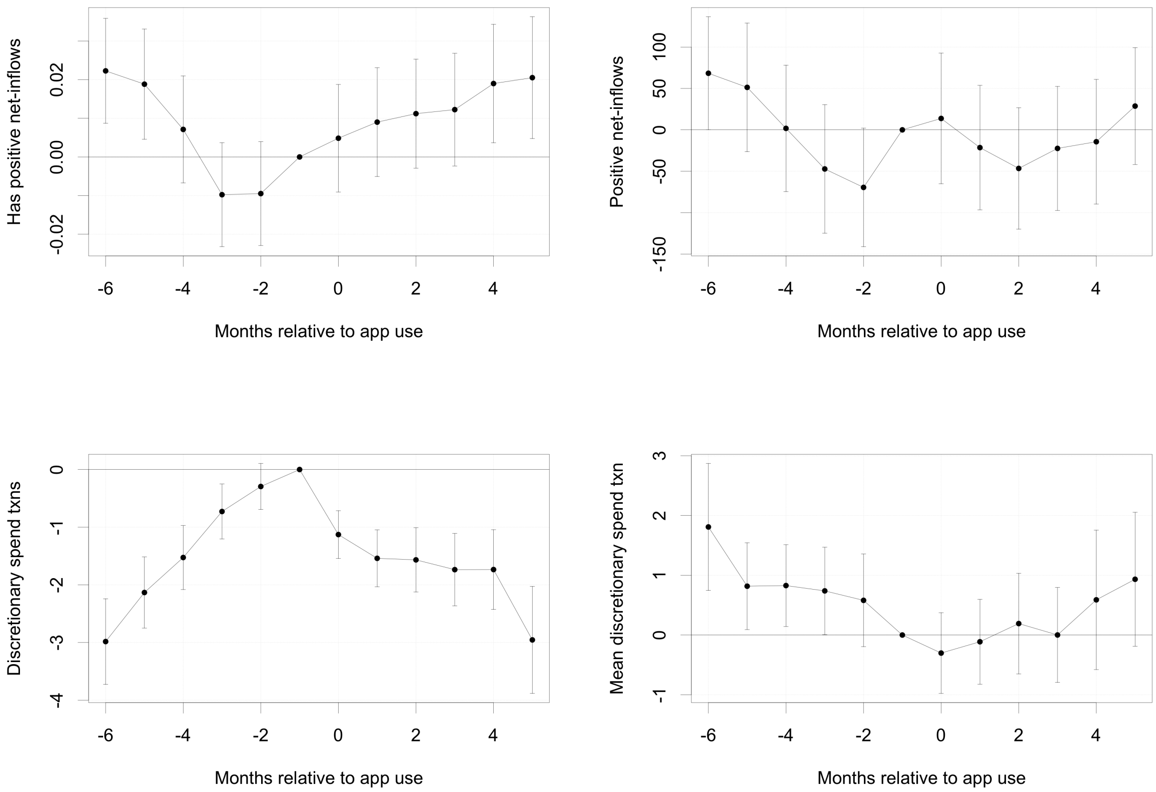
Figure 9: Main results



Notes: Figures show estimates of leads and lags frome estimating model (1) with distretionary spend (left) and netflows into savings accounts (right) as the dependent variable. Standard errors are clustered by user.

4.1 Decomposing intensive and extensive margins

Figure 10: Decomposition of intensive and extensive margins



Notes: ...

Table 4: Intensive and extensive margins

Dependent Variables:	Has positive net-inflows	Positive net-inflows	Discretionary spend txns	Mean discretionary spend txn
Model:	(1)	(2)	(3)	(4)
Variables				
Months relative to app use = -6	0.02***	68.33**	-2.98***	1.81***
	[0.01; 0.04]	[0.10; 136.56]	[-3.73; -2.24]	[0.75; 2.87]
Months relative to app use = -5	0.02***	51.29	-2.13***	0.82**
	[0.00; 0.03]	[-26.47; 129.04]	[-2.75; -1.52]	[0.09; 1.54]
Months relative to app use = -4	0.01	1.77	-1.53***	0.83**
	[-0.01; 0.02]	[-74.62; 78.15]	[-2.08; -0.97]	[0.14; 1.51]
Months relative to app use = -3	-0.01	-47.23	-0.73***	0.74**
	[-0.02; 0.00]	[-124.87; 30.40]	[-1.20; -0.25]	[0.01; 1.47]
Months relative to app use = -2	-0.01	-69.51*	-0.29	0.58
	[-0.02; 0.00]	[-141.13; 2.11]	[-0.69; 0.10]	[-0.19; 1.35]
Months relative to app use = 0	0.00	13.75	-1.13***	-0.30
	[-0.01; 0.02]	[-65.21; 92.72]	[-1.54; -0.71]	[-0.98; 0.37]
Months relative to app use = 1	0.01	-21.45	-1.54***	-0.11
	[-0.01; 0.02]	[-96.59; 53.68]	[-2.03; -1.05]	[-0.82; 0.60]
Months relative to app use = 2	0.01	-46.60	-1.57***	0.19
	[-0.00; 0.03]	[-119.91; 26.70]	[-2.13; -1.01]	[-0.65; 1.03]
Months relative to app use = 3	0.01*	-22.42	-1.74***	0.00
	[-0.00; 0.03]	[-97.31; 52.48]	[-2.37; -1.11]	[-0.79; 0.79]
Months relative to app use = 4	0.02**	-14.37	-1.74***	0.59
	[0.00; 0.03]	[-89.61; 60.87]	[-2.43; -1.04]	[-0.58; 1.76]
Months relative to app use = 5	0.02**	28.69	-2.96***	0.93
	[0.00; 0.04]	[-41.84; 99.21]	[-3.88; -2.03]	[-0.19; 2.05]
Month income	0.00***	0.06***	0.00***	-0.00***
	[0.00; 0.00]	[0.04; 0.08]	[0.00; 0.00]	[-0.00; -0.00]
Month spend	-0.00***	0.02***	0.00***	0.00***
	[-0.00; -0.00]	[0.01; 0.03]	[0.00; 0.00]	[0.00; 0.00]
Active accounts	0.09***	184.51***	3.41***	-0.87***
	[0.09; 0.10]	[170.36; 198.66]	[3.21; 3.62]	[-1.21; -0.53]
Fixed-effects				
User ID	Yes	Yes	Yes	Yes
Year-month	Yes	Yes	Yes	Yes
Fit statistics				
Observations	188,324	188,324	188,324	187,306
R^2	0.28475	0.11551	0.68049	0.26390
Within R^2	0.06323	0.01277	0.17182	0.01562

Clustered (User ID) co-variance matrix, 95% confidence intervals in brackets
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

4.2 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.[6]

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)

5 Discussion

Limitations:

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

[6]Based on age ranges provides 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.

Callaway, Brantly and Pedro HC Sant'Anna (2021). "Difference-in-differences with multiple time periods". In: *Journal of Econometrics* 225.2, pp. 200–230.

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:

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

Schmidheiny, Kurt and Sebastian Siegloch (2019). "On event study designs and distributed-lag models: Equivalence, generalization and practical implications". In:

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.

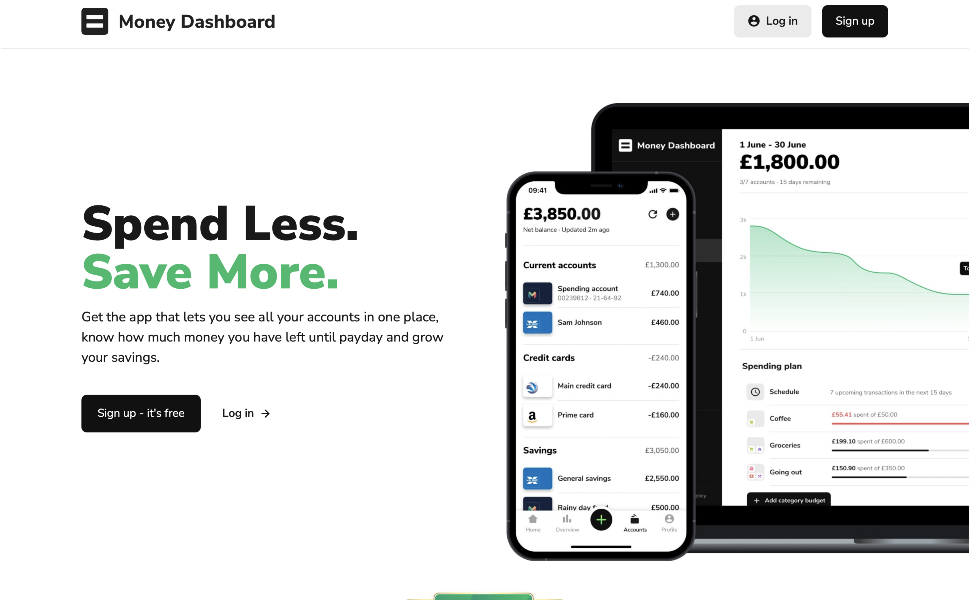
Sun, Liyang and Sarah Abraham (2021). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects". In: *Journal of Econometrics* 225.2, pp. 175–199.

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, accessed on 29 April 2022.

B Additional specifications

B.1 Static TWFE

$$y_{it} = \alpha_i + \lambda_t + \beta D_{it} + \gamma X_{it} + \epsilon_{it} \tag{2}$$

Table 5: Static results

Dependent Variables:	Net-inflows				Discretionary spend			
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables								
App use	37.93***	27.14**	5.60	-0.08	69.09***	49.57***	1.05	-26.59***
	[11.16; 64.71]	[0.73; 53.55]	[-26.57; 37.78]	[-50.70; 50.55]	[63.78; 74.40]	[41.46; 57.69]	[-12.92; 15.03]	[-38.73; -14.4
Month income	0.07***	0.07***	0.07***	0.08***	0.02***	0.01***	0.02***	0.01***
	[0.06; 0.08]	[0.05; 0.10]	[0.05; 0.09]	[0.05; 0.10]	[0.02; 0.02]	[0.01; 0.02]	[0.02; 0.03]	[0.00; 0.02]
Month spend	-0.12***	-0.16***	-0.12***	-0.16***	0.16***	0.12***	0.16***	0.12***
	[-0.13; -0.11]	[-0.19; -0.14]	[-0.14; -0.10]	[-0.19; -0.14]	[0.16; 0.16]	[0.11; 0.12]	[0.15; 0.17]	[0.11; 0.12]
Active accounts	40.31***	75.54***	38.87***	73.25***	45.89***	85.71***	42.11***	74.92***
	[32.66; 47.95]	[60.63; 90.45]	[25.50; 52.23]	[57.69; 88.81]	[44.37; 47.40]	[81.21; 90.20]	[39.61; 44.60]	[70.47; 79.38
Intercept	-7.11				172.15***			
	[-37.67; 23.45]				[166.09; 178.21]			
Fixed-effects								
User ID		Yes		Yes		Yes		Yes
Year-month			Yes	Yes			Yes	Yes
Fit statistics								
Observations	188,324	188,324	188,324	188,324	188,324	188,324	188,324	188,324
R^2	0.00888	0.04113	0.00940	0.04163	0.41618	0.66171	0.42425	0.67060
Within R^2		0.01014	0.00877	0.01000		0.28046	0.40985	0.23849

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

B.2 Additional TWFE specifications

Figure 12: Alternative model specifications

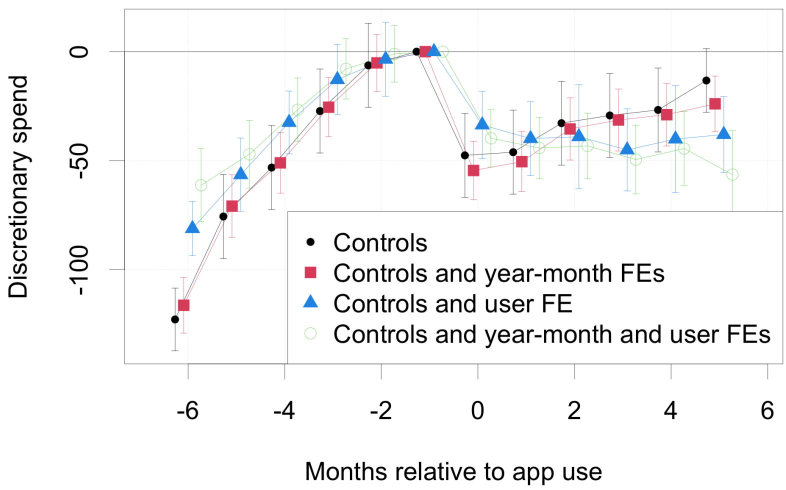


Table 6: Alternative model specifications

Dependent Variable:	Discretionary spend			
Model:	(1)	(2)	(3)	(4)
Variables				
Months relative to app use = -6	-122.89***	-116.38***	-81.19***	-61.31***
	[-137.26; -108.52]	[-129.14; -103.61]	[-93.92; -68.47]	[-78.13; -44.48]
Months relative to app use = -5	-75.65***	-70.88***	-56.47***	-47.06***
	[-94.93; -56.37]	[-85.23; -56.54]	[-73.72; -39.23]	[-62.64; -31.47]
Months relative to app use = -4	-53.18***	-51.02***	-32.49***	-26.61***
	[-72.45; -33.91]	[-64.94; -37.10]	[-47.26; -17.72]	[-41.13; -12.09]
Months relative to app use = -3	-27.25***	-25.50***	-12.83	-7.89
	[-46.51; -7.98]	[-39.08; -11.91]	[-29.37; 3.72]	[-21.71; 5.94]
Months relative to app use = -2	-6.28	-5.15	-3.49	-1.01
	[-25.55; 12.98]	[-18.28; 7.98]	[-20.95; 13.98]	[-13.00; 11.97]
Months relative to app use = 0	-47.60***	-54.52***	-33.64***	-39.76***
	[-66.86; -28.34]	[-67.92; -41.12]	[-49.51; -17.76]	[-52.94; -26.58]
Months relative to app use = 1	-46.13***	-50.48***	-39.90***	-44.23***
	[-65.40; -26.86]	[-64.31; -36.65]	[-57.33; -22.47]	[-58.30; -30.15]
Months relative to app use = 2	-32.80***	-35.47***	-39.02***	-43.24***
	[-52.07; -13.53]	[-49.73; -21.20]	[-63.52; -14.51]	[-58.23; -28.24]
Months relative to app use = 3	-29.28***	-31.36***	-45.09***	-49.52***
	[-48.56; -10.01]	[-45.57; -17.15]	[-64.44; -25.75]	[-65.30; -33.74]
Months relative to app use = 4	-26.75***	-28.93***	-40.10***	-44.45***
	[-46.02; -7.47]	[-43.25; -14.60]	[-65.31; -14.89]	[-61.37; -27.53]
Months relative to app use = 5	-13.23*	-23.94***	-38.01***	-56.39***
	[-27.83; 1.37]	[-36.72; -11.16]	[-55.92; -20.11]	[-76.52; -36.27]
Month income	0.02***	0.01***	0.02***	0.01***
	[0.02; 0.03]	[0.01; 0.02]	[0.02; 0.03]	[0.00; 0.02]
Month spend	0.16***	0.12***	0.16***	0.12***
	[0.16; 0.16]	[0.11; 0.12]	[0.15; 0.17]	[0.11; 0.12]
Active accounts	41.94***	77.85***	40.80***	72.39***
	[40.39; 43.49]	[73.30; 82.41]	[38.37; 43.22]	[67.89; 76.88]
Intercept	274.97***			
	[260.03; 289.91]			
Fixed-effects				
User ID		Yes		Yes
Year-month			Yes	Yes
Fit statistics				
Observations	188,324	188,324	188,324	188,324
R^2	0.41828	0.66337	0.42491	0.67097
Within R^2		0.28398	0.41052	0.23935

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

Figure 13: Alternative model specifications

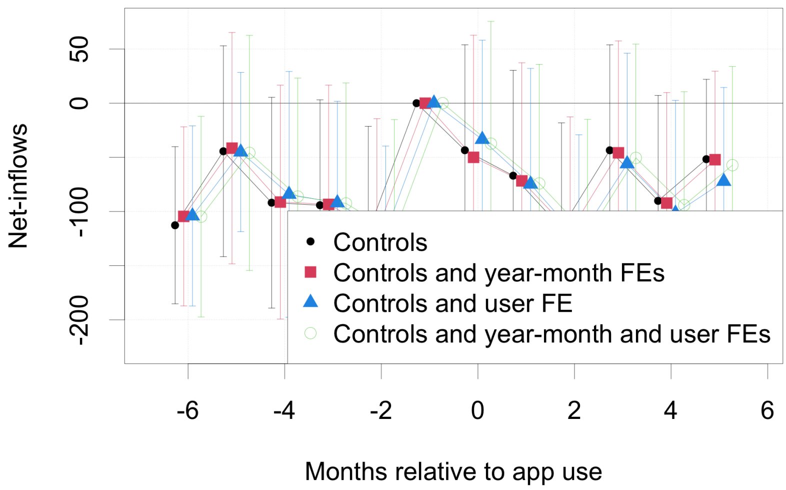


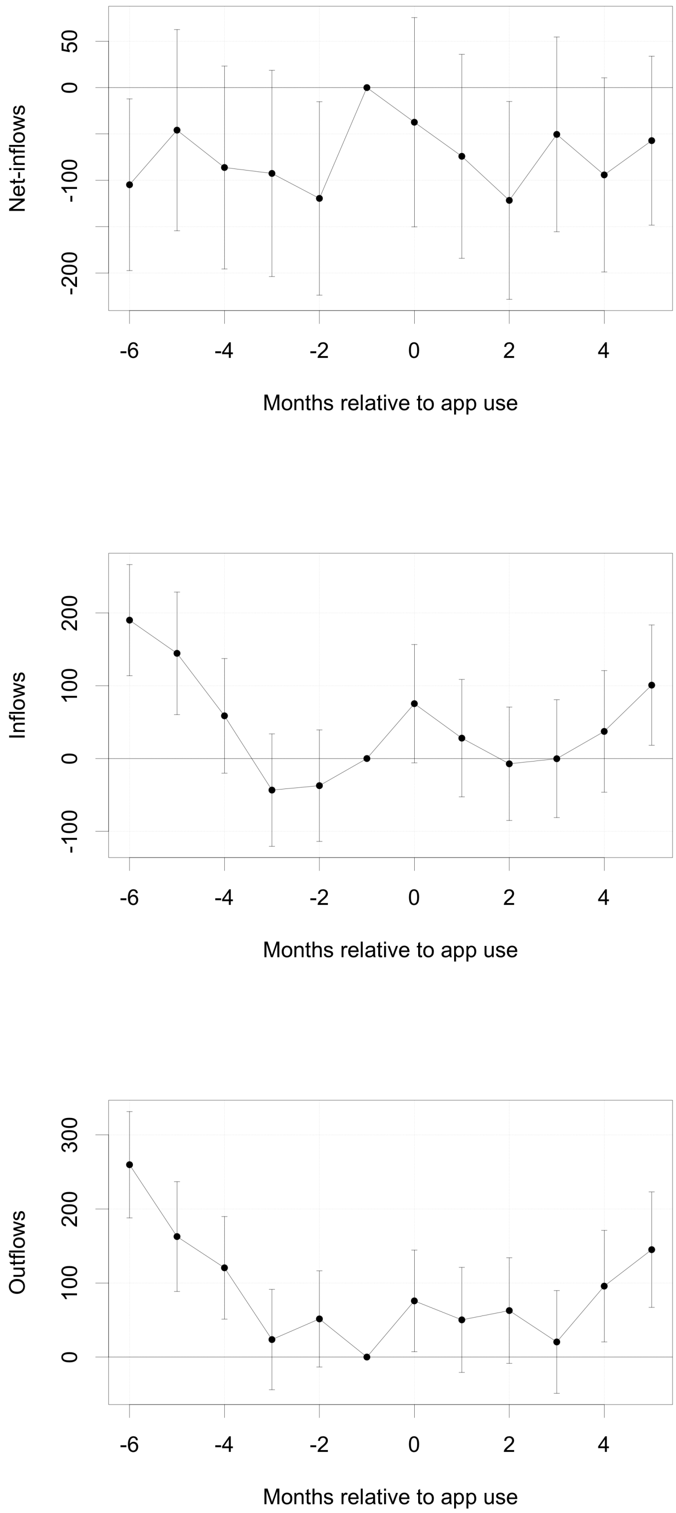
Table 7: Alternative model specifications

Dependent Variable:	Net-inflows			
Model:	(1)	(2)	(3)	(4)
Variables				
Months relative to app use = -6	-112.73***	-104.53**	-104.14**	-104.79**
	[-185.30; -40.15]	[-187.23; -21.82]	[-189.54; -18.75]	[-197.45; -12.13]
Months relative to app use = -5	-44.39	-41.53	-45.12	-45.90
	[-141.77; 52.99]	[-148.42; 65.37]	[-120.66; 30.41]	[-154.47; 62.67]
Months relative to app use = -4	-91.89*	-91.29*	-84.19	-86.20
	[-189.22; 5.44]	[-199.28; 16.69]	[-200.74; 32.37]	[-195.61; 23.21]
Months relative to app use = -3	-94.20*	-93.45*	-92.07*	-92.58
	[-191.50; 3.10]	[-203.59; 16.70]	[-188.41; 4.27]	[-203.84; 18.67]
Months relative to app use = -2	-118.68**	-118.36**	-119.19***	-119.53**
	[-215.98; -21.39]	[-222.44; -14.29]	[-200.97; -37.41]	[-223.87; -15.19]
Months relative to app use = 0	-43.44	-50.00	-33.48	-37.36
	[-140.74; 53.85]	[-162.78; 62.77]	[-127.72; 60.75]	[-150.33; 75.61]
Months relative to app use = 1	-66.94	-71.74	-74.79	-74.11
	[-164.24; 30.37]	[-180.80; 37.33]	[-184.62; 35.05]	[-184.14; 35.92]
Months relative to app use = 2	-115.51**	-118.47**	-125.25**	-121.66**
	[-212.82; -18.19]	[-224.24; -12.70]	[-224.00; -26.50]	[-228.41; -14.90]
Months relative to app use = 3	-43.50	-45.81	-56.03	-50.48
	[-140.83; 53.83]	[-149.17; 57.54]	[-161.02; 48.97]	[-155.51; 54.56]
Months relative to app use = 4	-90.18*	-92.22*	-101.71*	-94.18*
	[-187.52; 7.17]	[-194.27; 9.84]	[-208.92; 5.49]	[-198.96; 10.59]
Months relative to app use = 5	-51.70	-52.18	-72.25	-57.21
	[-125.43; 22.03]	[-133.90; 29.54]	[-161.36; 16.86]	[-148.40; 33.97]
Month income	0.07***	0.07***	0.07***	0.08***
	[0.06; 0.08]	[0.05; 0.10]	[0.05; 0.09]	[0.05; 0.10]
Month spend	-0.12***	-0.16***	-0.12***	-0.16***
	[-0.13; -0.11]	[-0.19; -0.14]	[-0.14; -0.10]	[-0.19; -0.14]
Active accounts	38.47***	72.54***	37.00***	72.05***
	[30.62; 46.31]	[56.98; 88.10]	[24.59; 51.40]	[56.26; 87.84]
Intercept	95.97**			
	[20.52; 171.42]			
Fixed-effects				
User ID		Yes		Yes
Year-month			Yes	Yes
Fit statistics				
Observations	188,324	188,324	188,324	188,324
R^2	0.00897	0.04121	0.00948	0.04170
Within R^2		0.01022	0.00885	0.01007

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

B.3 Decomposing inflows and outflows

Figure 14: Decomposing inflows and outflows



Notes: ...

Table 8: Decomposing inflows and outflows

Dependent Variables:	Net-inflows	Inflows	Outflows
Model:	(1)	(2)	(3)
Variables			
Months relative to app use = -6	-104.79**	190.11***	259.80***
	[-197.45; -12.13]	[113.66; 266.56]	[187.96; 331.64]
Months relative to app use = -5	-45.90	144.54***	162.75***
	[-154.47; 62.67]	[60.37; 228.70]	[88.62; 236.88]
Months relative to app use = -4	-86.20	58.67	120.54***
	[-195.61; 23.21]	[-20.02; 137.37]	[51.23; 189.85]
Months relative to app use = -3	-92.58	-43.36	23.62
	[-203.84; 18.67]	[-120.62; 33.89]	[-44.16; 91.39]
Months relative to app use = -2	-119.53**	-37.28	51.52
	[-223.87; -15.19]	[-113.85; 39.30]	[-13.38; 116.42]
Months relative to app use = 0	-37.36	75.39*	75.86**
	[-150.33; 75.61]	[-5.93; 156.72]	[7.16; 144.56]
Months relative to app use = 1	-74.11	28.01	50.24
	[-184.14; 35.92]	[-52.74; 108.76]	[-20.73; 121.21]
Months relative to app use = 2	-121.66**	-7.26	62.77*
	[-228.41; -14.90]	[-85.19; 70.68]	[-8.57; 134.11]
Months relative to app use = 3	-50.48	-0.26	20.39
	[-155.51; 54.56]	[-81.18; 80.66]	[-48.99; 89.77]
Months relative to app use = 4	-94.18*	37.26	95.80**
	[-198.96; 10.59]	[-46.32; 120.84]	[20.47; 171.13]
Months relative to app use = 5	-57.21	100.83**	145.05***
	[-148.40; 33.97]	[18.23; 183.44]	[67.11; 223.00]
Month income	0.08***	0.06***	-0.01
	[0.05; 0.10]	[0.04; 0.09]	[-0.03; 0.02]
Month spend	-0.16***	0.10***	0.25***
	[-0.19; -0.14]	[0.08; 0.12]	[0.23; 0.26]
Active accounts	72.05***	306.25***	243.69***
	[56.26; 87.84]	[286.09; 326.40]	[224.27; 263.11]
Fixed-effects			
User ID	Yes	Yes	Yes
Year-month	Yes	Yes	Yes
Fit statistics			
Observations	188,324	188,324	188,324
R^2	0.04170	0.24635	0.31097
Within R^2	0.01007	0.03612	0.07191

Clustered (User ID) co-variance matrix, 95% confidence intervals in brackets
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

B.4 Twoway clustering

Figures compares main results, which are clustered at the user id-level, with specification that clusters at the user and year-month level.

B.5 Longer event windows

Figures show main results for longer effect windows (additional periods have fewer observations and thus larger confidence intervals)