# Essays on financial behaviour using transaction data

by
Fabian Gunzinger
Thesis submitted to the University of Warwick for the degree of Doctor of Philosophy
University of Warwick
Warwick Business School
September 2023

# Contents

# List of Tables

# List of Figures

# Acknowledgements

# Declarations

# Abstract

# Introduction

This is the intro to my thesis.

# Chapter 1

# Machine-learned behaviour traits

## 1.1 Introduction

Time preferences are important to understand consumer decision-making. Strongly discounting the future relative to the present has been linked to a higher likelihood for choosing unhealthy foods (Milkman et al. 2010, Read and Van Leeuwen 1998), overpaying on gym membership fees (DellaVigna and Malmendier 2006), a higher propensity for addictive behaviours and a higher BMI (Reimers et al. 2009, Urminsky and Bayer 2014), and a preference for jobs that are available immediately but are less enjoyable (Hesketh et al. 1998) or less well paid (Saunders and Fogarty 2001) over more enjoyable or better paid jobs that are available later on.

In this paper, we introduce a novel approach to elicit time preferences from mass financial transaction data. Lab and field studies typically elicit time preferences by making participants choose between a smaller-sooner and a larger-later effort or payoff. The core idea of our approach is to identify a series of consumption choices that mimic this trade-off, and then exploit these choices to estimate time preferences not only for the subset of users for whom we observe the choice in the data but for all users in the data. The three choice scenarios we identify are monthly or yearly payment for car insurance, purchasing a new mobile phone outright or paying in monthly installments, and swapping car and home insurer frequently or infrequently. Each of these chioces implicitly trades off a higher cost or effort in the present but reduced cost in the future with lower cost or effort in the present but higher-cost in the future. We identify individuals in our dataset that engage in each of these choice scenarios, and then predict their likelihood of making the patient choice based on features we extract from the data. We call that likelihood individual's propensity for patience. We then use this predictive model to predict the propensity for patience also for those users in our dataset who didn't engage in a particular choice scenario. Finally, we test whether predicted propensities from difference choice scenarios correlate positively, as we would expect if they all elicit the same underlying time preferences.

We find, however, that propensities based on differen choice scenarios are uncorrelated, suggesting that in our particular application, our approach is unsuccessful at eliciting underlying preferences. We discuss how noisy data, omitted controls, or invalid assumptions might explain this results.

Our approach lies on the intersection of two related strands in the literatures of estimating time preferences.[1] One uses field data to estimate time preferences based on consumption decisions of the smaller-sooner or larger-later type. For instance, the seminal paper in this literature, Hausman (1979), studies the trade-off between capital and operating cost when choosing between more or less energy efficient energy-using durables, whereas later studies look at purchases of fuel-efficient cars (Busse et al. 2013, Allcott and Wozny 2014). The second strand relies on high-frequency financial data. In particular, Kuchler and Pagel (2020), use data from a credit-card debt management app to measure present-biasedness based on whether consumption of immediately consumed goods like alcohol and meals consumed at a restaurant are higher in the week following payday. This approach to measuring time preferences is appealing as it dovetails nicely with present-bias models, which predict that such co-movement is high for present-biased people. Our approach lies on the intersection of these two strands as it relies on the same smaller-sooner or larger-later approach as the former literature but relies on high-frequency financial data.

The main difference of our approach to previous approaches and its main contribution is twofold: first, it can be used to estimate any behaviour trait given suitable data. Second, data that can be used to elicit the behaviour trait only needs to be available for a subset of individuals in the data. These features are valuable because they allow researchers to bridge lab or field experimental approaches, which allow for the elicitation of behaviour traits but can only be used at a relatively small scale, and the analysis of high-frequency data that becomes increasingly popular in social sciences, but from which it is usually harder to elicit behaviour traits. In addition, given that the approach relies on field rather than experimental data, it has high ecological validity. Finally, the approach can be used to measure time preferences in a context where income data is unavailable, and where the approach of Kuchler and Pagel (2020) is thus infeasible.

Our work also contributes to a rapidly growing literature of using financial-transaction data from banks or financial aggregator apps to understand consumer financial behaviour. As already mentioned, Kuchler and Pagel (2020) use data from a financial aggregator app to estimate time preferences. Similar data has been used to show that consumer spending varies across the pay cycle (Gelman et al. 2014, Olafsson and Pagel 2018), to test the consumer spending response to exogenous shocks (S. R. Baker 2018, Baugh et al. 2014),

[1]The literature on estimating time preferences is vast and diverse, spanning a large number of field and lab studies in different disciplines. For a more comprehensive discussions, see Frederick et al. (2002) and Cohen et al. (2020).

and to better understand the generational differences in financial platform usage patterns (Carlin et al. 2019). Some researchers use transaction-data directly provided by banks. Ganong and Noel (2019) show that consumer spending drops sharply after the predictable income drop from exhausting unemployment insurance benefits, Meyer and Pagel (2018) analyse how individuals reinvest realised capital gains and losses, and Muggleton et al. (2020) show that chaotic spending behaviour is a harbinger of financial distress.

The remainder of the paper is organised as follows: Section A.2 introduces the dataset used for our analysis and discusses data pre-processing and sample selection; Section A.3 introduces ie detail our empirical approach and estimates time-preferences; Section 2.6 discusses possible reasons for the lack of correlation between our different propensity for patience measures, general limitations of time preferences estimated with our approach, and opportunities for further research.

## 1.2 Economic models of intertemporal choice

This section briefly discusses the economic models that are the conceptual framework of our approach. We do not attempt to estimate specific parameters of these models, but they inform the way we conceptualise intertemporal decisions, and form the basis based upon which we classify certain choices as patient and others as impatient.

Within economics, the most widely used approach to model intertemporal choice and time preferences is time discounting. This section introduces the standard workhorse model, exponential discounting, and the most popular alternative among behavioural economists, quasi-hyperbolic discounting. The two other main classes of models of intertemporal choice and self-control within economics are models of coexisting multiple selves, which usually model the interaction between a myopic and a farsighted part of the self (R. H. Thaler and Shefrin 1981, Benhabib and Bisin 2005, Fudenberg and Levine 2006, Loewenstein and O'Donoghue 2004), and unitary self models, such as Gul and Pesendorfer (2001), who model agents as having preferences over choice sets and might choose to remove tempting options. Inside and outside of economics, there is a large number of alternative models, such as heuristic models like "subadditive discounting" Read (2001).

Models of time-discounting, introduced by Paul Samuelson (1937), assume that a person's intertemporal (or "total") utility, $U_t$, is the weighted sum of present and future (instantaneous) utilities, $u$, which are determined by the level of consumption, $c$, in each period. The weight is the discount function, $D(\tau)$, which indicates how much the person values utility $\tau$ periods from now relative to immediately experienced utility. Such a model can be written as:

$$U_t = \sum_{\tau=0}^{T-t} D(\tau)u(c_\tau). \tag{1.1}$$

Maximising utility thus requires optimally allocating consumption over time, usually given some constraints. In his seminal paper, Samuelson went one step further and specified the discount function as $D(t) = \delta^t$, where $\delta = \frac{1}{1+\rho}$ is the *discount factor*, and $\rho$ is the *discount rate*. We can thus rewrite the equation above as:

$$U_t = \sum_{\tau=0}^{T-t} \delta^\tau u(c_\tau). \tag{1.2}$$

This is the exponential discounting model. Because $\delta$ is generally assumed to be smaller than one, the model posits that while total utility depends on all future utilities, utility that is further in the future contributes exponentially less. A person who thinks in one-year intervals and is indifferent between $u(c_t)$ today and $u(c_{t+1}) = 1.2u(c_t)$ one year from now has a discount rate of 20 percent and a discount factor of 84 percent, implying that they value consumption one year from today at 84 percent of current consumption.[2]

As pointed out in Frederick et al. (2002), the model's simplicity and elegance come at the cost of strong underlying assumptions. Most importantly, the model assumes that all motives that could conceivably affect intertemporal choices – factors like impulse control, uncertainty, the expectation of changing tastes, inflation – can all be captured by the single parameter $\delta$, and are all equally relevant for all types of choices and over all horizons.[3] The model also has a consequential implication: it predicts that people's relative preferences for wellbeing at an earlier or a later date are independent of the time of choosing. For example, the person who was indifferent between $u(c_t)$ today and $1.2u(c_t)$ a year from now would also be indifferent between these two options if they were to occur $\tau$ periods in the future, such as five years from now and six years from now. While Samuelson himself was well aware that the model was unlikely to adequately capture human behaviour and pointed out in the original article that it was "completely arbitrary" to assume that a person's behaviour would be governed by a function like equation (1.2), the model rapidly became economists' standard framework to study intertemporal choice.

Yet over the the past decades, a large empirical literature has documented that people do not behave as the exponential model predicts. For instance, R. Thaler (1981) and subsequent contributions have documented that discount rates are not constant but larger for short than for long horizons, and experiments with animals (mainly rats and pigeons), suggest that these animals discount rewards in a similar fashion Berns et al. (2007). Potentially as a result of this, Shapiro (2005) finds that people on food-stamps over-consume in the short-run. It is also well documented that the timing of decision does matter: people often experience preference reversals when choosing foods (Read and Van Leeuwen 1998, McClure et al. 2007, Milkman et al. 2010) and movies (Read, Loewenstein,

[2]From $u(c_t) = \frac{1.2}{1+\rho}u(c_t) \Leftrightarrow 1.2 = 1 + \rho$. The continuously compounded discount rate is $ln(1.2/1) \approx 0.18 = 18\%$.

[3]See Frederick et al. (2002), Urminsky and Zauberman (2015), and Read, McDonald, et al. (2018) for further discussion on these and additional assumptions.

et al. 1999, Milkman et al. 2009), and when they plan work (Augenblick et al. 2015) and exercise frequency (DellaVigna and Malmendier 2006).[4] The strength of the evidence against time-consistency led Laibson and Marzilli-Ericson (2019) to state that:

> The focus for intertemporal choice research is no longer whether the exponential discounted utility model is empirically accurate, but, instead, what models best explain the robust behavioral deviations we observe.

In economics, the most popular approach to capture these behavioural deviations are models of quasi-hyperbolic discounting, also called models of *present-bias*, because they explicitly model a preference for present over future consumption (Laibson 1994, Laibson 1997).[5] These models extend the exponential discounting model by specifying the discount function as

$$D(\tau) = \begin{cases} 1 & \text{if } t = 0 \\ \beta\delta^t & \text{if } t > 0, \end{cases} \tag{1.3}$$

so that the full model becomes:

$$U_t = u(c_t) + \beta \sum_{\tau=1}^{T-t} \delta^\tau u(c_\tau), \tag{1.4}$$

where $0 < \beta, \delta \leq 1$. The simple addition to the DU model is the parameter $\beta$, which additionally discounts all future periods, and hence captures the additional degree of impatience in the present relative to all future periods. It is this addition that can account for observed preference reversals.[6]

Quasi-hyperbolic models builds on earlier models of hyperbolic discounting (Strotz (1955), Ainslie (1975), Ainslie (2001), and Loewenstein and Prelec (1992), which posit that people have high discount rates for short horizons and increasingly smaller marginal discount rates for longer horizons. In a choice between a smaller reward sooner and a larger reward later it is this feature of the discount curve that shifts the choice from larger to smaller as the choice moves closer to the present, making a hyperbolic discounter patient when making long-run tradeoffs and impatient when making short-run tradeoffs.

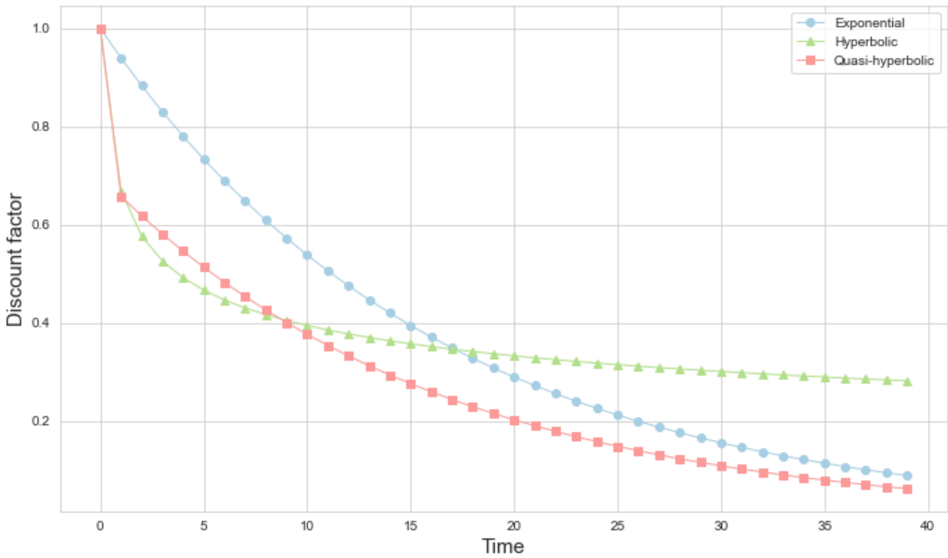
The main attraction of the quasi-hyperbolic model is that it also allows for different short-run and long-run discount rates, while being mathematically tractable. Figure 1.1 visualises the differences and similarities between stylised versions of the three different discount functions.

[4]For complete discussions of the evidence, see Frederick et al. (2002), Read, McDonald, et al. (2018), and Laibson and Marzilli-Ericson (2019).

[5]An important and influential extension to this literature that is beyond the scope of this essay is the difference between sophisticated agents, who are aware of their present bias, and naive agents, who are not (Strotz 1955, O'Donoghue and Rabin 1999b).

[6]Urminsky and Zauberman (2015) point out that an additional feature of adding another parameter is that it conceptually separates present-bias (captured by $\beta$ from the average discount rate ($\delta$).

Figure 1.1: Comparison of different discount functions




Notes: Comparison of exponential, hypterbolic, and quasi-hyperbolic discount functions. The functions are parametrise as follows: exponential with $\delta = 0.94$, hyperbolic with $\alpha = 4$ and $\gamma = 1$, and quasi-hyperbolic with $\delta = 0.94$ and $\beta = 0.7$.

Our estimates of individuals' propensity for patience do not correspond to any specific parameters discussed above, but build on the intuition that people often disproportionately discount the future relative to the benchmark of the exponential discounting model.

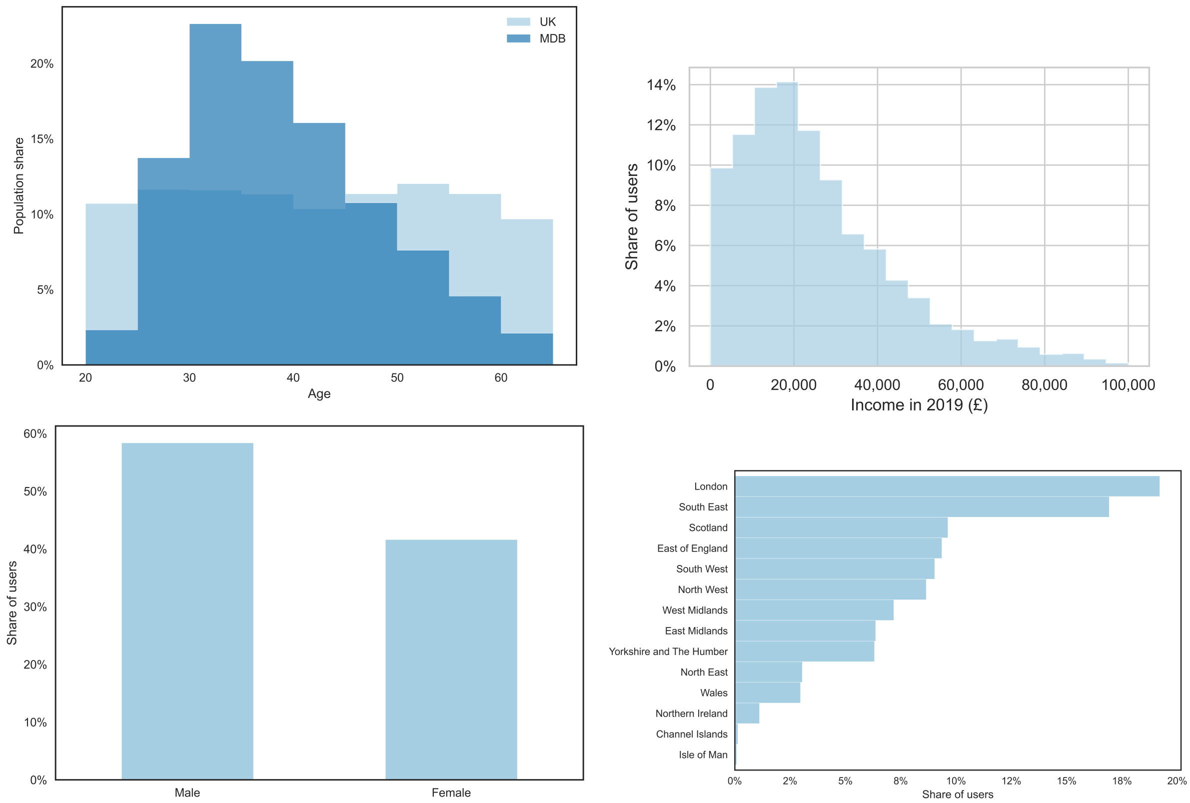
## 1.3 Data

### 1.3.1 Dataset description

Data is provided by Money Dashboard (MDB), a UK-based financial management app that allows its users to add accounts from all their banks to obtain an integrated view of their finances. Our dataset contains information on more than 500 million transactions made between 2012 and June 2020 by more than 250,000 users. For each transaction, we can see the amount, date, and description of the transaction, as well as transaction *tags*, classifications added by MDB that indicate the type of the transaction (e.g. 'groceries', 'insurance'). We also have basic information on each user (e.g. year of birth, postcode sector) as well information about each bank account (e.g. type of account, date added).

The main advantages of the data for the study of consumer financial behaviour are its high (transaction-level) frequency, that it is automatically collected and updated and thus less prone to errors and unaffected by biases that bedevil survey measures, and that it offers a view of consumers' entire financial life across all their accounts, rather than just a view of their accounts held at a single bank (provided they added all their accounts to MDB).

Figure 1.2: Demographic characteristics of Money Dashboard users




Notes: Top left panel shows distribution of ages in decade buckets in the MDB data relative to the UK, top right panel shows 2019 after-tax income distribution of MDB users, bottom left shows proportion of men and women among MDB users, bottom right panel shows distribution of regions.

The main limitation is the non-representativeness of the sample relative to the population as a whole. Financial management apps are known to be used disproportionally by men, younger people, and people of higher socioeconomic status (Carlin et al. 2019, MAS 2014), a fact that is reflected in our data. The four panels in Figure 1.2 provide an overview of demographic characteristics of our sample. It makes clear that Money Dashboard users are not a representative sample of the UK population: they are predominantly males in their thirties who live in London or the South East and are relatively well off (the income distribution is shifted to the right relative to the UK as a whole). To calculate incomes, we broadly follow Hacioglu et al. (2020) in defining total income as the sum of earnings, pension income, benefits, and other income. Also, as pointed out in Gelman et al. (2014), a willingness to share financial information with a third party might not only select on demographic characteristics, but also for an increased need for financial management, or, one could argue, for a higher degree of financial sophistication. However, while non-representativeness could partially be addressed by re-weighting the sample, as was done in Bourquin et al. (2020), it is not of much consequence for our purpose here, since our ability to infer behaviour traits from transaction data is not dependent on having a representative sample of people.

### 1.3.2 Data preprocessing

To reliably identify user behaviour from the data requires a number of preprocessing steps to eliminate noisy, duplicated, or unclassifiable transaction.

One major challenge in this context is to identify transfers. This is important because transfers can significantly affect how we perceive a users' spending and savings behaviour. Not all identifiable transfers are correctly tagged by MDB's automatic tag. Thus, in addition to relying on said tag, we identify additional transfers in two ways: we use regular expressions to tag as transfers for which the description is indicative of a transfer (e.g. "tfr" or "xfer"), and we identify transaction pairs in the data that are mirror images of one another. Following Bourquin et al. (2020) we identify two transactions as a pair when they share the same user, are of the same amount and higher than £50, are no more than 4 days apart, of the opposite sign, and if neither transaction is already part of another pair.

As discussed above, each transaction is automatically tagged. MDB uses three types of tags: an automatic tag set by their machine learning algorithms, a manual tag tag can optionally be set by the user, and a user precedence tag, which equals the automatic tag if a user doesn't manually tag a transaction and the manual tag if they do. We correct the user precedence tag in those cases where id doesn't behave as just described. Untagged transactions account for about 18 percent of transactions in the raw data. Because we cannot determine whether these are purchases or transfers, we drop them from our sample.

This means we probably underestimate income and expenditure, but avoid misclassifying transfers as either of those categories.

Duplicated accounts can occur in the data for a number of reasons, such as when a user registers twice and adds the same account both times, or when a couple both add their joint account. MDB provides a list of accounts that they have identified as duplicates, and we drop all such duplicated accounts, while leaving the original account in the data.

Finally, MDB categorises user's bank accounts into current, credit card, savings, and other. We exclude savings and other accounts whenever we calculate spending-related measures.

### 1.3.3 Sample selection

The goal of our sample selection procedure is to only retain users for whom we observe all expenditures and incomes, and whom we observe long enough so that we can reliably perform the calculations for our analysis. We start with a 10 percent sample of the full dataset and apply a series of selection steps. Table A.1 shows how each criteria affects sample size.

Table 1.1: Sample selection

| | Users | User-months | Txns | Txns (m£) |
|---|---|---|---|---|
| Raw sample | 25,321 | 570,959 | 55,254,410 | 7,434 |
| Annual income of at least £10k | 8,886 | 186,685 | 19,751,162 | 2,700 |
| Income in 2/3 of all observed months | 8,886 | 186,685 | 19,751,162 | 2,700 |
| At least one savings account | 5,299 | 121,210 | 13,365,032 | 1,949 |
| At least 6 months of data | 4,683 | 119,755 | 13,238,027 | 1,929 |
| Monthly debits of at least £200 | 4,383 | 112,278 | 12,581,384 | 1,838 |
| Five or more monthly current account txns | 4,099 | 105,693 | 11,783,774 | 1,712 |
| Minimum level of spend diversity | 4,095 | 105,541 | 11,771,533 | 1,710 |
| Complete demographic information | 3,423 | 91,644 | 10,205,122 | 1,476 |
| Final sample | 3,423 | 91,644 | 10,205,122 | 1,476 |

Notes: Sample selection table based on a 10 percent sample of the complete data. Columns indicate the number of users, accounts, transactions and the total transaction value remaining in the data after each selection step.

We select only users whom we observe for at least six months to ensure that we have enough data to produce features for our predictive model. We then apply a series of steps to filer out users are unlikely to have linked all their accounts to MDB: we require users to have at least one current account, at least 5 monthly debits worth at least £200, have income payments in two out of every three months, and an annual income of between £5,000 and £100,000. Next, we drop users who use more than 10 accounts in a single year and have debits of more than £100,000 per month, since these are likely to be users who use MDB for business purposes. Finally, we drop users who are not yet 18 years old or are older than 65 years.

## 1.4 Estimating time-preferences

Our approach to elicit time-preferences from transaction data is based on three steps: first, we identify choice scenarios in the data that reveal a user's time preferences, the subset of users for whom we observe that scenario, and the choice of each user in that scenario; second, we fit a supervised machine-learning model to predict the choices made by the formerly identified subset of users; and third, we use that model to predict what choice all users would have made had we observed that scenario. In effect, the last step provides us with an estimate of how patiently a user would have behaved in a particular choice scenario had they engaged in it. We refer to this choice as a user's propensity for patience in a particular scenario. We repeat these three steps for each choice scenario. In the final step, we test whether the propensities for patience from the different scenarios are positively correlated – as we would expect them to be if they all measured the same underlying time-preferences – and can be combined into a single score.

### 1.4.1 Identifying choice scenarios

For each of the three choice scenarios, this section discusses the rational, the approach used to classify users into different types, and the limitations of that approach.

#### Car insurance payments

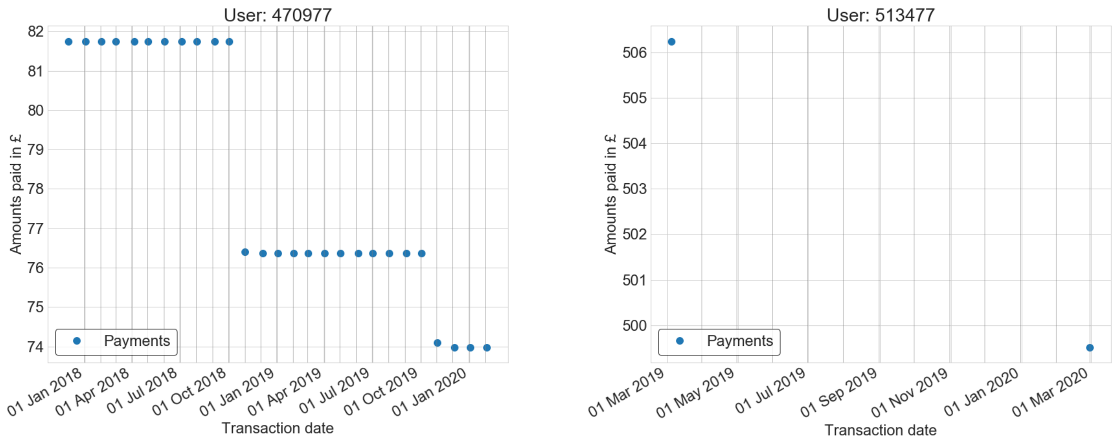
In the UK, car owners can choose to pay car insurance either one a year or in monthly instalments, which come with an additional interest charge of about 20 percent per year (Money Supermarket 2020). Paying in instalments thus leaves an individual with more disposable income in the short term but less in one year's time. Because a 20 percent APR will generally exceed the return from any alternative use of the money, paying yearly is the rational option to choose for people who can afford to do so.[7]

We identify as a car insurance payer anyone who has at least one transaction classified as "vehicle insurance" by MDB. We then identify monthly payments as payments of a fixed amount that are part of a series of at least five consecutive payments with an identical amount. We identify yearly payments as those payments that are higher than £250 and are not part of such a series. To identify monthly payments, we truncate pence because inspection of the data suggests that payments can vary slightly month to month. We choose a series length of 5 because data inspection also showed that the amount occasionally jumps to a slightly higher level after five or six payments. For yearly

[7] For a thorough discussion of what factors should influence the smaller-sooner vs. larger-later choice, see Read, McDonald, et al. (2018).

payments, the £250 threshold is based on data provided by Money Supermarket (2020). We then classify users as monthly payers if they have at least one series of monthly payments but never make a yearly payment, and as yearly payers if the reverse is true. Users who make both types of payments are excluded. Figure 1.3 provides an example of a typical monthly payer (left panel) and a typical yearly payer (right panel).

Figure 1.3: Monthly and yearly car insurance payment patterns




Notes: Illustrative case studies for a users paying for their car insurance monthly (left) and annualy (right). Gray vertical lines indicate calendar months, blue dots car insurance payments. Figures are based on actual user data.

One limitation of our classification is that we don't account for liquidity and credit constraints to allow for the possibility that some users who need a car might want to pay yearly but cannot afford to. Future research could capture such constraints in a number of different ways, such as by taking into account a user's current account balance, their savings, or whether their current balance allows them to cover non-discretionary spending for a certain amount of time.

### New mobile phone payments

UK mobile phone providers often give customers the option to either pay for a new headset upfront or in monthly instalments. Because paying upfront is usually more expensive in the present but cheaper overall (in the future), while paying in instalments is cheaper in the present but more expensive overall, we can exploit this decision to elicit the degree of a user's present-biasedness.

To differentiate between upfront and instalment payments for new headsets we rely on a contractual feature of O2, one of the largest UK mobile providers, which takes headset instalment payments separately from airtime payments but on the same day for all customers on their Refresh plan (O2 Mobile 2020). Payments are also exclusively taken via debit direct, so that they should fall on the same date each month (baring weekends or bank holidays). Similarly to our strategy for identifying monthly car insurance users, we identify users who pay for a new mobile phone in instalments as users who make at least six payments of the same amount to O2, under the additional restriction that they also make at least one additional payment on each payment day. We identify individuals who pay upfront for a new phone as those who make an unusually large payment of at least £100. We define 'unusually large' as a payment that does not happen more than once per year, and that is 3-times larger than the average mobile payment in a rolling window that includes 10 payments on either side of the payment. Together, these two criteria ensure that we do not accidentally capture individuals with generally high phone bills or individuals who had temporarily high bills. O2's Refresh plans were introduced in April 2013 (O2 Mobile 2013), but different mobile phone tariffs for upfront and instalment payments were abandoned in August 2018 (O2 Mobile 2018). We thus restrict the data used to identify instalment and upfront payers to that time period. As with car insurance, we drop users who make both instalment and upfront payments.

Similar to the car insurance scenario above, some users might face liquidity or credit constraints when buying a new headset and thus not have the choice to pay upfront. Future research could account for this by ways similar to the ones discussed above. Comparing payment methods of low and high income users in our data suggests, however, that financial constraints are not a major driver of sorting into monthly and upfront payers: in a one-percent sample of our data, users with income below £20k account for 46 percent of upfront users and only 30 percent of instalment users.

**Swapping car and home insurer**

Research shows that home and car insurance customers who do not shop around every year for a cheaper insurer pay a premium of about £30 and £60 per year, respectively (Zuke 2017). Shopping for a new insurer takes effort and is thus costly in the short-term. For insurance customers, the choice whether to invest some time now in order to save money over the course of the next year is thus another case where a relatively small reward in the present (saving time) is traded off against a larger reward in the future (saving on insurance premiums). The trade-off differs from the ones in the above two scenarios in that it doesn't involve two financial rewards, but is conceptually analogous in that an investment in the present leads to a reward in the future.

We classify users for whom we do observe car or home insurance payments as swappers if they change either their car or home insurer every three years.

The main limitation of our classification is that some individuals might decide that the time it takes them to shop around for a new insurer is worth more than the annual savings and thus consciously decide to not swap their insurer. One way future research could test for this is to see whether high income individuals are less likely to swap.

### 1.4.2 Model fitting

To model users's choices we derive a set of behavioural features from the data, fit a range of candidate models, and then fine-tune the best performing model to arrive at our final model.

#### Features

The set of behavioural fparagraphFeatures:eatures we use in our models comprises expenditure shares for 750+ identified merchants, expenditure shares for 170+ expenditure categories, the average number of weekly grocery shops, the percent of purchases paid for by credit-card, the percent of transactions tagged manually in the MDB app, and spending entropy over categories. Spending entropy is calculated as

$$H = -\sum p_i log_2(p_i), \tag{1.5}$$

where $p_i$ is the probability that an individual makes a purchase in spending category $i$. The measure can broadly be interpreted as the degree of randomness in how individuals spend their money: the next purchase of an individual who only ever spends money on a single category is perfectly predictable and their H-score would be 0, while predicting the next purchase of a person who spends on all categories with equal probabilities would be akin to drawing category-labelled balls out of an urn, and their H-score would be higher than zero.

We standardise all numeric features before feeding them into our machine learning algorithms. Because including demographic variables such as age group, gender, or income does not substantially improve the fit of our models and so we do not include them in our final specification.

#### Fitting candidate models

To estimate users's propensity for patience we need classifying models that return probabilities. We fit a number of different such models (e.g. logistic and ridge classifiers). We put 20 percent of our data aside for testing later on, and then use 5-fold cross-validation to evaluate the fit of each model. We measure model performance using the $F_1$ score, which combines a model's precision and recall scores. Precision and recall are defined, respectively, as

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}, \tag{1.6}$$

and

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}. \tag{1.7}$$

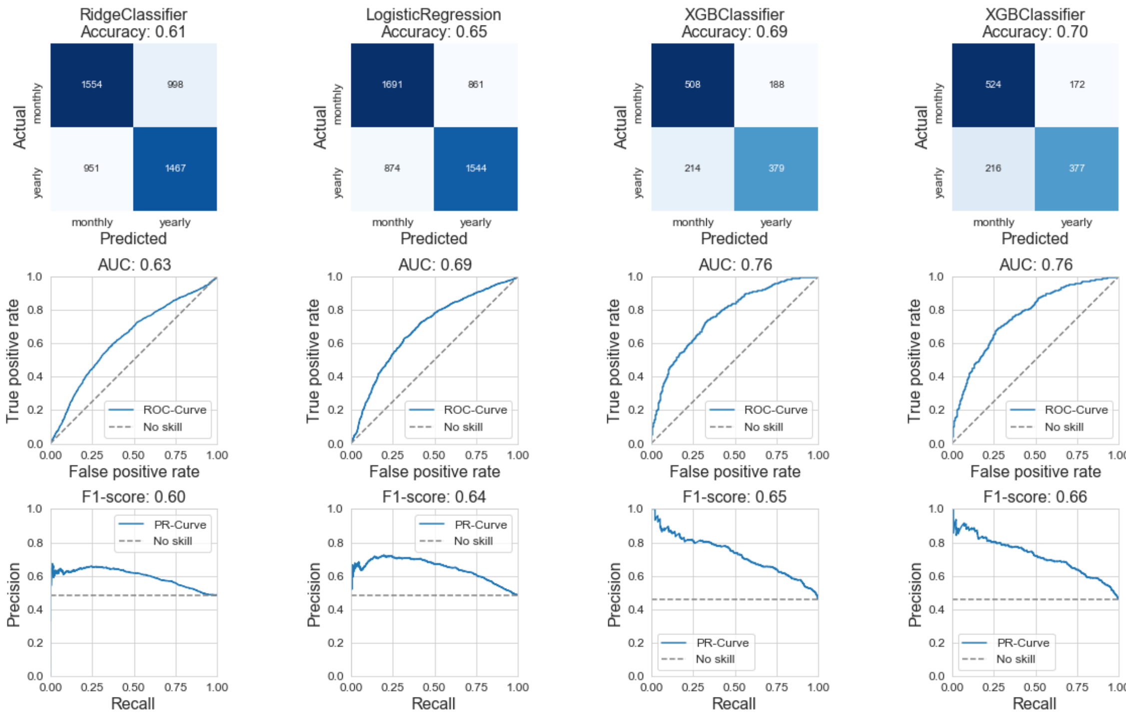
The two concepts originate in information retrieval, and capture two intuitive ideas: precision is the proportion of all true positives out of all returned positives, while recall is the proportion of all true positives that have been returned as such out of all existing true positives. A model might have achieve high precision if it is conservative and minimised the number of false positives, or high recall if it returned a high number of positives. The $F_1$ balances these extremes. It is defined as:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{1.8}$$

We do not use the more common accuracy score because the choices observed in our three choice scenarios tend to exhibit strong class imbalance: most users make the same one out of two possible choices. In such a scenario it would be possible for a model to achieve high accuracy by simply predicting the more frequent class in all cases. In contrast, to achieve a high $F_1$ score, a model needs to have both high precision and high recall.

The first three columns in Figure 1.4 shows the performance of a Ridge classifier, logistic regression, and XGBoost model for the car insurance data. The first row shows confusion matrices, the second row AUC curves, and the third row F1-scores, our main model assessment score.

Figure 1.4: Candidate models comparison




Notes: Confusion matrix, ROC curve, and Precision-Recall curve for four candidate models for car insurance scenario. Top left cell in each confusion matrix provides the number of users that pay monthly for their car insurance that are accurately classified by the model. The dummy classifier is a skill-free model that classifies all users as monthly payers.

We can see that the XGBoost model outperforms the ridge and logistic classifier across all scores. It is thus this model that we fine-tune to build our prediction model for the car insurance data. While not shown, we perform the same model comparisons for the mobile phone payment and insurance swap scenarios. In both cases, the XGBoost model performs best, as well.

#### Model fine-tuning

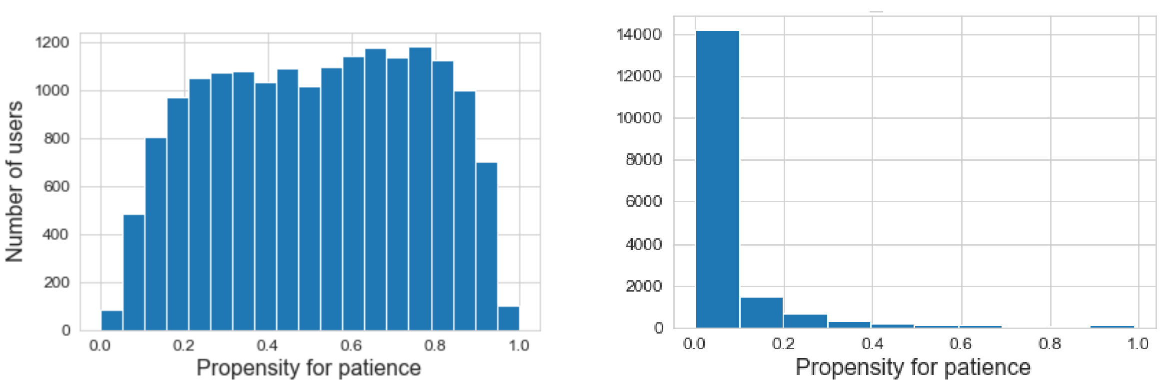
The XGBoost model outperforms all other models for all three scenarios. We optimise its performance by tuning some of its hyperparameters using grid search and 10-fold cross validation. Parameters we optimise for are the maximal tree depth (2, 3, 4, 5, 6), the learning rate (0.01, 0.05, 0.1, 0.3, 0.5), the tree complexity parameter $\gamma$ (0, 0.25, 1), and the regularisation term $\lambda$ (0, 1, 10). The final model has a maximal tree depth of 4, a learning rate of 0.1, a complexity parameter $\gamma$ of 1, and a regularisation term $\lambda$ of 0.

Column four in Figure 1.4 shows the XGBoost model performance after fine-tuning and shows that we managed to perform performance only slightly.

### 1.4.3 Predicting time preferences

The fine-tuned models from the previous subsections allow us to predict, for each choice scenario, the behaviour of users for whom we observe a choice in that scenario in the data. For each such user, we know whether they made the patient or impatient choice, and for each such user, the model provides us with the probability of the user making the patient choice, which we interpret as a user's propensity for being patient in that scenario. To obtain such a propensity for all users, we simply rerun the model on the entire sample of users. The resulting probability for each user gives us their propensity for making the patient choice if they had faced the choice of a given choice scenario (e.g. their probability of paying for car insurance yearly had they bought car insurance).

Figure 1.5: Propensity for patience estimates



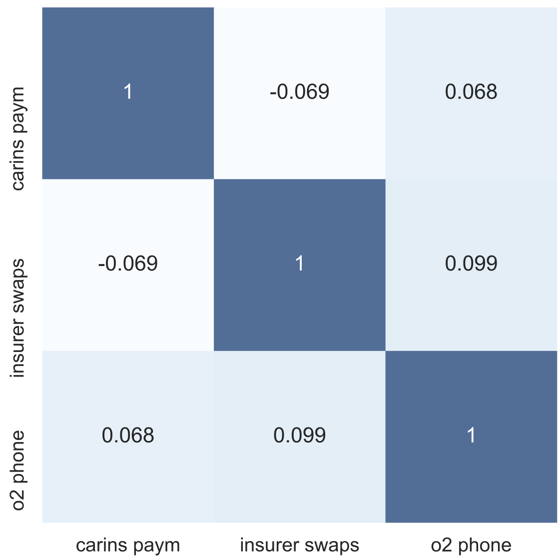

Notes: Estimated propensity for patience scores for based on car insurance payments (left) and mobile phone payemnts (right).

Figure 1.5 shows the distribution of estimated propensities for the car insurance mode of payment (left) and mobile phone purchase (right) scenarios. We can see that the distribution of estimated propensities for the car insurance payment resembles a uniform distribution, implying the model cannot clearly distinguish between yearly and monthly payers (if it could, we would see a bimodal distribution with peaks around zero and one). Conversely, the distribution of estimated propensities for mobile phones shows that the model classifies most users as monthly payers.

### 1.4.4 Combining propensities

The final step is to combine the propensities from each scenario into a single metric that captures a user's estimated propensity for patience. If estimated propensities from each scenario capture a user's stable underlying time preferences we would expect those propensities to positively correlate, and combining them into a single metric by taking some for of average measure would seem unproblematic. Figure 1.6 shows, however, that the correlation among the three propensities is virtually zero.

Figure 1.6: Correlations of scenario-based propensities for patience




Notes: Correlation matrix of propensity for patience scores based on car insurance payments, insurer swaps, and mobile phone purchases.

There are three broad possible reasons for that: noisy data, omitted controls, and that the assumptions of our approach are invalid.

Noisy data is a problem because identifying choice scenarios and classifying users's choice is not straightforward. Figure 1.3 shows two examples of "well-behaved" car insurance payment patterns that make it easy to classify the user on the left as a monthly and the user on the right as a yearly payer. Not all data is like that: often cases are harder to classify both because users make multiple transactions of very different amounts in a given period (i.e. paying mobile phone bills for multiple phones), and Money Dashboard's algorithms to tag transactions are imperfect (i.e. some payments tagged as mobile phone payments might be incorrect). As a result, not only do we loose observations because we cannot identify all cases where a user engages in a particular scenario, but we also misclassify the choice of some users in some scenarios (i.e. we might identify someone as having paid upfront for a new mobile headset even though they really paid monthly or maybe did not even buy a new headset). We have tried to limit such cases as well as possible, but eliminating them entirely is not possible.

In addition to challenging data, the fact that we did not control for certain constraints on users's choices might also influence results. As discussed above, it is reasonable to assume that some users are forced to pay monthly because they do not have the means to pay yearly, so that their decision is not driven by their level of patience. It is further conceivable that some of those users would patiently take the time to look for better insurance deals every year and change their insurer frequently, leading us to classify them as patient in that scenario and leading to a low correlation among their scenario-based propensity measures.

Finally, the fundamental assumption our approach relies on is that a set of stable underlying time-preferences guide users's intertemporal choices across different situations, and that the influence of those preferences on each choice is strong enough relative to all other influences that we can pick it up in our data. The assumption can be violated in two ways: either because there are no stable underlying preferences, or because, for some reason, we are unable to pick them up in our data.

A lack of stable underlying preferences is consistent with a large body of behavioural decision research, which suggests that when making a choice, people construct, rather than reveal, their preferences by choosing one of a number of preference-generating methods they have at their disposal. They choose that method based on context factors (e.g. the framing of the problem), their multiple and potentially conflicting goals (e.g. minimise effort and regret), their mental processing capabilities, and their values and beliefs (Payne et al. 1992). If preferences are constructed using a highly context-dependent method rather than revealed when a user makes a choice, then we have, of course, no hope of eliciting such preferences.

Yet even if stable preferences did exist and influence users's choices, it is conceivable that, compared to other factors, their influence was relatively weak and produces too small a signal for our approach to pick up. One way to get a sense of how much of users's decisions is driven by stable and context-dependent factors would be to compare within and between variation if, in addition to the large number of users on the cross-sectional dimension, we had scenarios for which we observed a large number of decisions over time for each user. Finally, it is possible that the discounted utility framework on which our approach is based is a poor model of individual choice, leading our approach to produce biased estimates. Stewart et al. (2019) show that – far from measuring stable parameters – risk-preference estimates from lab experiments based on either expected utility or prospect theory models are systematically biased and sensitive to the precise experimental setup. They conclude that estimating risk preferences is futile because not only do such estimates suffer from model bias (because people don't make choices according to either of the two standard models), but the bias varies depending on the choice set used to elicit the estimate. Our results might suffer from similar problems in the context of time-preferences and field-experiments: if users make intertemporal trade-offs based on a different model than exponential discounting and its variants, and if the choice set of the decision (e.g. what mobile headsets are on offer) strongly influence our elicited propensities – if, in other words, our scenario-specific propensities are heavily biased – this could explain the lack of correlation among them.

### 1.4.5 Validation

If we had been able to obtain a single propensity measure for each user, we would have validated them using two approaches: we would have compared our propensities to estimates of present-biasedness based on income-consumption co-movement as calculated by Kuchler and Pagel (2020), and we would test how our estimates correlate with variables that have been found to correlate with present-biasedness, such as age and salary level (Lawrance 1991, Warner and Pleeter 2001, Reimers et al. 2009), smoking (Reimers et al. 2009), and drug consumption (Kirby et al. 1999).

### 1.4.6 Predicting financial outcomes

There exists a large literature that links high discount rates to lower savings (Laibson 1997, Hurst 2004, Falk et al. 2018) lower financial literacy (Meier and Sprenger 2013), higher credit card debt (Meier and Sprenger 2010, Harrison et al. 2002), a lower likelihood to paying bills in full (Chabris et al. 2008), and less home equity positions (Johnson et al. 2011). Furthermore, the closely related literature on the link of self-control and financial outcomes finds a positive correlation between self-control and levels of wealth (Ameriks et al. 2007, Biljanovska and Palligkinis 2018) and between low levels of self-control and financial literacy and non-payment of consumer credit and higher self-reported excessive financial burdens of debt (Gathergood 2012). Interestingly, the findings in Sutter et al. (2013) suggest that the link between time-preference and life outcomes forms early: of the more than 600 10-18-year-olds in their study, those with higher experimentally elicited discount rates are more likely to smoke, drink, have a higher BMI, save less, and display worse conduct at school.

Hence, an interesting next step based on a single propensity for patience measure would have been to test whether patience is correlated with financial outcomes such as the frequency and amounts of emergency and long-term savings, monthly discretionary spend, and payment of penalty fees. Chabris et al. (2008) asked 555 participants to make a set of 27 smaller-sooner vs. larger-later choices in the lab, and then correlated elicited discount rates with six self-reported financial behaviours: gambling, paying late fees on credit card, paying credit card bills in full, savings, and wealth. Perhaps surprisingly given the above discussion, they only find a relationship between discount rates and full payment of credit card bills but no relationship with any other behaviour. Testing whether patience based on our proposed approach predicts financial outcomes would be a way to validate the results in Chabris et al. (2008) with automatically collected data for a much larger sample of individuals.

## 1.5 Discussion

We have introduced a novel approach to estimate time-preferences from large-scale transaction data. While the specific application in this paper has not succeeded in eliciting time preferences from Money Dashboard users, we believe that attempting to refine the approach and apply it in other contexts is worthwhile and promising, because the benefit of succeeding is large, and because there are concrete ways in which our approach here could be improved.

The main appeal of the approach is that, if successful, it would allow researchers to estimate time-preferences for a large set of individuals based on the smaller-sooner and larger-later payoff approach that researchers commonly use in lab experiments. More broadly, the general approach we introduce – identify scenarios that identify choices used to estimate behaviour traits, train a prediction model on individuals in the data for whom that scenario is observed, and use that model to predict the trait for all individuals in the data – could, in principle, be used for other behaviour traits, too. Large-scale observational data of the type used in this study has become available only in recent years and is increasingly used across the social sciences Buyalskaya et al. (2021). Once challenge with such data is that they do not usually provide the same detailed information about behaviour traits that can be gained from observing a small set of people in a lab. Our approach has the potential to bridge that divide by allowing researchers to augment large-scale observational data with behaviour traits elicited from the data itself.

There are three ways in which our work could be refined in future research. First, as discussed at the end of Section A.3, identifying choice scenarios is made more challenging because of imperfect tagging of transactions in the data, and because we do not have direct access to the algorithms used for tagging. Using the approach on data with more complete and reliable tagging or in collaboration with the data provider, so that anomalies in the data can be identified and handled, could eliminate that source of risk. Second, the prediction models we used could be enhanced both with additional features and by the use of more sophisticated models and further tuning. In terms of features, we have already discussed that, in our context, adding variables that more fully capture users' financial constraints might have been useful. Furthermore, we have relied on a relatively straightforward set of features, that could probably be augmented. In terms of models, we have used a number of relatively simple "off-the-shelf" models. Using more advanced models could help provide more accurate predictions. Finally, as discussed in Section A.3, it is possible that people the way people actually make intertemporal choices is not well captured, even approximately, by the discounting models discussed in Section 1.2, on which our elicitation approach has built. If this is true, then either eliciting time-preferences based on a different framework, or focusing on eliciting different behaviour traits might prove more successful.

Finally, even if our approach should be successful, it would come with certain limitations. In its specific application to estimating time preferences, our approach has the same three main limitations as other field studies that try to estimate time preferences using a version of the smaller sooner or larger later approach. First, and most fundamentally, the approach does not account for how money – which is what implicitly is being traded off in the mobile phone purchase decision – is translated into utils – which is what time discounting models are based on. The issue here is threefold:[8] how money translates into consumption (it could be invested), when consumption is taking place (consumption could be smoothed over many periods), and how consumption translates into utility (utility functions are generally assumed to be concave).[9] Second, field studies of this kind are unable to distinguish between time preferences and other confounding factors that influence intertemporal choices, such as users beliefs about uncertainty, inflation, changing preferences, and the influence of habit formation, anticipatory utility, and visceral influences on decision-making. They also cannot control for situational factors that might influence choice, such as attention, beliefs, and knowledge. Finally, the approach relies on the assumption that people's discount rate for money is the same as for other goods, which, as discussed in Read, McDonald, et al. (2018) and Urminsky and Zauberman (2015), may well not be the case.

Ultimately, the usefulness of our approach in measuring time preferences depends on how it is being used and interpreted. Given the limitations just discussed, our estimates are unlikely to be a precise measure of peoples' discount rates or of their time preferences in a narrow sense, as captured by the parameters in canonical models of intertemporal choice. But if interpreted as a broad measure of impatience as it manifests itself in a real-life setting with all its confounding influences, such estimates can be very useful. In

[8]See Frederick et al. (2002), Read, McDonald, et al. (2018), and Cohen et al. (2020) for a more comprehensive discussion.

[9]As pointed out in Read, McDonald, et al. (2018), assuming that people do not engage in intertemporal arbitrate requires assuming that they use "narrow bracketing" and think about the choice at hand in isolation.

a research setting, they can, for instance, serve to further our understanding both of what drives impatience (e.g. whether it is circumstantial or innate) as well as the behaviour it leads to (e.g. higher credit card debts). In commercial applications, our estimates could, for instance, help financial providers identify users that make costly financial mistakes (e.g. high credit-card fees) because of impatience and design interventions that make being patient easier (or automate the process so no active decision is required on the part of the user).

# Chapter 2

# Spending entropy predicts savings behaviour

## 2.1 Introduction

This paper documents variations in spending profiles and payments into emergency savings for a large set of users of a financial management app and shows that spending profiles predict emergency savings.

We define emergency savings as inflows into savings accounts. These savings will be made up of savings for particular goals – a new car, a holiday, a wedding – and savings directed towards building up a buffer for financial emergencies. Because in our data we cannot distinguish between these two cases, we refer to all of them as emergency savings.[1] These short-term savings are distinct from long-term savings aimed to build up funds for retirement, either through individually owned pension and investment vehicles or employer-linked pension schemes. Both kinds of saving are important for financial well-being, yet while there is a large literature on pension savings, little is known about how people save for the short-term.[2]

Studying the determinants of emergency savings is important because around a quarter of adults in the UK and the US are unable to cover irregular expenses like car and medical

[1] MDB allows users to create custom tags and some users use them to indicate the intended use for their savings transactions (e.g. "wedding", "holidays"). But only a very small number of transactions have such tags, and we do not pursue this further.

[2] Well-documented behavioural biases that help explain undersaving for pensions are, among others, present bias (Laibson 1997, Laibson and Marzilli-Ericson 2019), inertia (Madrian and Shea 2001), over-extrapolation (Choi et al. 2009), and limited self-control and willpower (R. H. Thaler and Shefrin 1981, Benhabib and Bisin 2005, Fudenberg and Levine 2006, Loewenstein and O'Donoghue 2004, Gul and Pesendorfer 2001). One danger of viewing low savings mainly as a result of behavioural biases is that while these biases likely do play some role and designing environments and tools to help correct them are thus part of the solution, it is at least conceivable that this is an area where the focus on behaviour-level solutions distracts from an effort to find more effective society-level solutions, a danger inherent in behavioural science research convincingly highlighted in Chater and Loewenstein (2022): if the main problem is that many people are unable to earn enough to save, then the effectiveness of helping them manage their low incomes more effectively pales in comparison with efforts to help them earn more.

bills: In the UK, 25 percent of adults would be unable to cover an unexpected bill of £300 (Philipps et al. 2021), while in the US, about 30 percent would be unable to cover a $400 bill (Governors of the Federal Reserve System 2022). Similarly, resarch in the UK has shown that having £1000 in savings reduces by more than half a household's chances of falling into debt that leads to financial problems (Philipps et al. 2021).

Studying spending profiles is of interest because:

- Our understanding of how people spend their money is based on survey data.
- Large-scale transaction-level data offers the possibility to study spending behaviour based on real-time data that are automatically collected for a large number of users. Such data has only become available ver recently and have not, thus far, been used to investigate systematically how people spend their money.
- Research in psychology suggests that disorder is maladaptive and associated with a range of negative outcomes such as impaired executive function (Vernon-Feagans et al. 2016), lower cognitive inhibition (Mittal et al. 2015), and activation of anxiety-related neural circuits (Hirsh et al. 2012). In the study of human behaviour, more chaotic behaviour has been found to predict the a higher number of visits to and higher spend in supermarkets (Guidotti et al. 2015), higher calorie intake (Skatova et al. 2019) and financial distress (Muggleton et al. 2020).

We hypothesise that less predictable spending patterns are associated with a lower probability for making payments into emergency savings accounts. Possible channels:

- Disorder (personal life or environment): leads to more impulsive shopping behaviour and makes forgetting to save more likely.
- Scarcity: life challenges focus attention away from deliberate shopping, causing hore impulse purchases, and make fortetting to save more likely.

What we do:

- Systematically documenting emergency savings patterns.
- Systematically documenting variation in spending profiles.
- Showing that unpredictability in spending profiles is associated with lower emergency savings.

Contribution to literatures:

- Understanding emergency savings behaivour (nest, aspen reports), (Sabat and Gallagher 2019) for sources on short-term savings literature, Colby and Chapman (2013) for lit on savings goals. See Colby and Chapman (2013) has useful literature review on short-term savings and suggests that subgoals can increase willingness to forego short-amounts in the present because they move the reference point in a prospect-theory framework. Philipps et al. (2021) present results from an employer-linked initiative that offers employees to have a portion of their salary automatically transferred into a savings pot. Policy literature: (CAN 2019, CFPB 2017, MPS 2018). Older literature:Savings lit: Lunt and Livingstone (1991) and Oaten and Cheng (2007)

- Understanding effect of behavioural entropy - eliciting useful personality characteristics from large-scale data.
- Use of high-frequency transaction data (itself a sub-literature of use of newly available large-scale datasets).

## 2.2 Methods

### 2.2.1 Dataset description

We use data from Money Dashboard (MDB), a financial management app that allows its users to link accounts from different banks to obtain an integrated view of their finances.[3] The dataset contains more than 500 million transactions made between 2012 and June 2020 by about 250,000 users, and provides information such as date, amount, and description about the transaction as well as account and user-level information.

The main advantages of the data for the study of consumer financial behaviour are its high frequency, that it is automatically collected and updated and thus less prone to errors and unaffected by biases that bedevil survey measures, and that it offers a view of consumers' entire financial life across all their accounts, rather than just a view of their accounts held at a single bank, provided they added all their accounts to MDB. The main limitation is the non-representativeness of the sample relative to the population as a whole. Financial management apps are known to be used disproportionally by men, younger people, and people of higher socioeconomic status (Carlin et al. 2019). Also, as pointed out in Gelman et al. (2014), a willingness to share financial information with a third party might not only select on demographic characteristics, but also for an increased need for financial management or a higher degree of financial sophistication. Because our analysis does not rely on representativeness, we do not address this.[4]

**Data issues** Bourquin et al. (2020) argue that because some of the accounts in the data will be joint accounts, units of observations should be tought of as "households" rather

[3] https://www.moneydashboard.com.

[4] For an example of how re-weighing can be used to mitigate the non-representative issue, see Bourquin et al. (2020).

than "users". We do not agree that this is the most prudent approach. The validity of thinking of units as households depends on the proportion of users in the data who add joint accounts and on the proportion of transactions – out of a user's total number of transactions – additionally observed as a result. Given that the sample is skewed towards younger individuals we think it is unlikely that a majority of them has added joint accounts. Furthermore, it seems reasonable to assume that in most cases, joint accounts are mainly used for common household expenditures similar that are similar to those of a single user (albeit in higher amounts), and are thus unlikely to alter the observed spending profile much. Thus, we think of units of observations as individuals, not households.

Some accounts might be business accounts. Using versions of the algorightms used by Bourquin et al. (2020) to identify such accounts showed, however, that such accounts only make up a tiny percentage of overall accounts and would not influence our results. We thus do not exclude them.

### 2.2.2 Preprocessing and sample selection

We restrict our sample to users for whom we can observe a regular income, can be reasonably sure that they have added all their bank account to MDB, and for whom we observe at least six months of data. Table A.1 summarises the sample selection steps we applied to a 1 percent sample of the raw data, associated data losses, and the size of our final sample.

Table 2.1: Sample selection

| | Users | User-months | Txns | Txns (m£) |
|---|---|---|---|---|
| Raw sample | 271,856 | 7,948,520 | 662,112,975 | 124,573 |
| Drop test users | 270,782 | 7,878,398 | 656,047,534 | 122,887 |
| App signup after March 2017 | 88,368 | 2,320,421 | 202,580,838 | 38,816 |
| At least one savings account | 50,226 | 1,334,328 | 125,841,337 | 26,645 |
| At least one current account | 48,794 | 1,303,164 | 123,468,715 | 26,263 |
| At least £5,000 of annual income | 20,647 | 541,746 | 55,857,451 | 11,760 |
| At least 10 txns each month | 14,229 | 369,944 | 40,662,904 | 8,529 |
| At least £200 of monthly spend | 10,438 | 272,228 | 31,529,498 | 6,837 |
| No more than 10 active accounts | 9,788 | 248,975 | 27,589,696 | 5,426 |
| Complete demographic information | 7,720 | 202,633 | 22,671,753 | 4,378 |
| Working age | 7,568 | 197,951 | 22,279,279 | 4,213 |
| Final sample | 7,568 | 197,951 | 22,279,279 | 4,213 |

### 2.2.3 Dependent variables

Identifying savings transactions: We classify as payments into savings accounts all savings account credits of £5 or more that are not identified as interest payments or automated "save the change" transfers (similarly for debits).[5]

Dummy for savings txn in current month. Motivation: MPS (2018) finds that saving habit is often more important than amount saved.

### 2.2.4 Spending profiles

We define a user's spending profile as the distribution of the number of spending transactionas across different spend categories. To summarise these distributions, we calculate spending entropy, based on the formula proposed by Shannon (1948), who defines entropy as $H = -\sum p_i log(p_i)$, which sums, for all possible events, the product of the probability of an event $i$ occuring with the logarithm of that probability. [6] The base the logarithm is often chosen to be 2, though other choices are possible.[7] Entropy is a cornerstone of information theory, where it measures the amount of information contained in an event. In the behavioural sciences, behavioural entropy has recently been shown to predict the frequency of grocery visits and the per-capita spend per visit (Guidotti et al. 2015), the amount of calories consumed (Skatova et al. 2019), and the propensity for financial distress (Muggleton et al. 2020). In our context, we define the entropy of a user's spending profile at a particular point in time as:[8]

$$H = -\sum p_c log(p_c), \tag{2.1}$$

where $p_c$ is the probability that an individual makes a purchase in spending category $c$, and $log$ is the base 2 logarithm. For simpler interpretation of our regression coefficients below, we standardise entropy scores to have a mean of 0 and a standard deviation of 1.

Higher entropy means that transactions are more equal across different spending categories, which makes it hard to predict the next transaction, whereas low entropy profiles have the bulk of transactions in a few dominant categories (such as groceries and transportation) and have relatively few transactions in other categories.[9]

We calculate entropy based on three sets of spend categories. The first measure is

[5]While standing order transactions are unlikely to be related to entropy in the short-run, we do not exclude such transactions since, best we can tell, the only account for a small fraction of total transactions.

[6]Shannon entropy is customarily denoted as $H$ following Shannon's own naming after Ludwig Boltzman's 1872 H-theorem in statistical mechanics, to which it is analagous.

[7]The choice of the base for the logarithm varies by application and determines the units of $I(E)$. Base 2 means that information is expressed in bits. The natural logarithm, another popular choice, expresses information in *nats*.

[8]We omit individual and time subscripts to keep notation simpler.

[9]For further discussion on how to interpret Equation 2.1, see Appendix .1.

based on 9 spending categories used by Muggleton et al. (2020).[10] The second measure is based on our own, more fine-grained, categorisation into 48 different categories.[11] The third measure is based on merchant names, as labelled by Money Dashboard.

We also handle categories with zero transaction counts in two different ways. To calculate what we call "unsmoothed" entropy scores, we calculate the $p_c$s in Equation 2.1 as simple frequentist probabilities

$$p_c = \frac{f_c}{\sum_{c=1}^{N} f_c}, \tag{2.2}$$

where $f_c$ is the number of transactions in spend category $c$ (the frequency with which $c$ occurs) and $N$ the number of categories. To avoid taking the log of zero for categories with zero transactions, the sum in Equation 2.1 is taken over categories with positive transaction counts only.[12] To calculate "smoothed" entropy scores, we apply additive smoothing to calculate propabilities as

$$p_c^s = \frac{f_c + 1}{\sum_{c=1}^{N} f_c + N}. \tag{2.3}$$

That is: we add 1 to each category's transaction count and divide by the sum of the total number of transactions and the number of categories. Because categories with a zero transaction count will have a numerator of 1, the sum in Equation 2.1 will be taken over all categories.

There are a number of alternative ways to characterise spend profiles. We could calculate profiles based on the distribution of transaction values rather than counts. We could also calculate profiles based on inter-temporal rather than intra-temporal distributions, focusing on consistency of purchasing behaviour over time rather than on predictability at any given time (Krumme et al. 2013). Further, we could focus on time-based rather than category-based measures, focusing, for instance, on whether purchases of the same type tend to occur on the same day of the week (Guidotti et al. 2015). Finally, one could also create composite measures based on principal component analysis, an approach used in Eagle et al. (2010). We leave these extensions for future research.

One slight limitation introduced by the imperfect transaction labelling in the MDB data is that entropy scores for high-entropy individuals will be biased downwards. This happens because unlabelled transactions tend to be transactions that are rare (i.e. not grocery or Amazon purchases), and it is high-entropy individuals that are more likely to engage in rare transactions. Because our analysis mainly relies on relative entropy levels, this is not of major consequence and we do not pursue this further.

[10] The precise mapping from MDB transaction tags into these 9 categories is available on Github.

[11] The precise mapping from MDB transaction tags into these 48 categories is available on Github.

[12] This is automatically handled by the entropy implementation of Python's SciPy package, which is what we use to calculate entropy scores.

### 2.2.5 Summary statistics

Table A.2 provides summary statistics.

Table 2.2: Summary statistics

| Statistic | Mean | St. Dev. | Min | Pctl(25) | Median | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| Txn count | 111.9 | 59.7 | 10 | 70 | 101 | 142 | 327 |
| Month income | 2,853.5 | 2,495.2 | 0.0 | 1,407.2 | 2,217.4 | 3,595.9 | 15,027.5 |
| Savings account inflows | 747.2 | 2,448.3 | 0.0 | 0.0 | 0.0 | 400.0 | 18,809.5 |
| Savings account outflows | 762.4 | 2,422.9 | 0.0 | 0.0 | 0.0 | 400.0 | 18,099.4 |
| Savings account netflows | −7.4 | 2,883.7 | −20,000.0 | 0.0 | 0.0 | 50.0 | 21,675.4 |
| Month spend | 2,760.4 | 2,609.9 | 200.0 | 1,225.4 | 2,016.7 | 3,318.9 | 17,092.2 |
| Age | 37.5 | 10.0 | 18 | 30 | 36 | 44 | 65 |
| Female dummy | 0.4 | 0.5 | 0 | 0 | 0 | 1 | 1 |
| Urban dummy | 0.8 | 0.4 | 0 | 1 | 1 | 1 | 1 |
| Discretionary spend | 3.1 | 1.7 | 1 | 2 | 3 | 4 | 10 |
| Active accounts | 860.6 | 736.1 | 0.0 | 369.3 | 663.0 | 1,118.3 | 4,181.7 |

Notes: Income and spend variables in ’000s of Pounds, number of unique categories for spend transaction classification based on 9 categories, 48 categories, and merchant names.

Figure 2.1

Figure 2.1: Demographic characteristics of Money Dashboard users

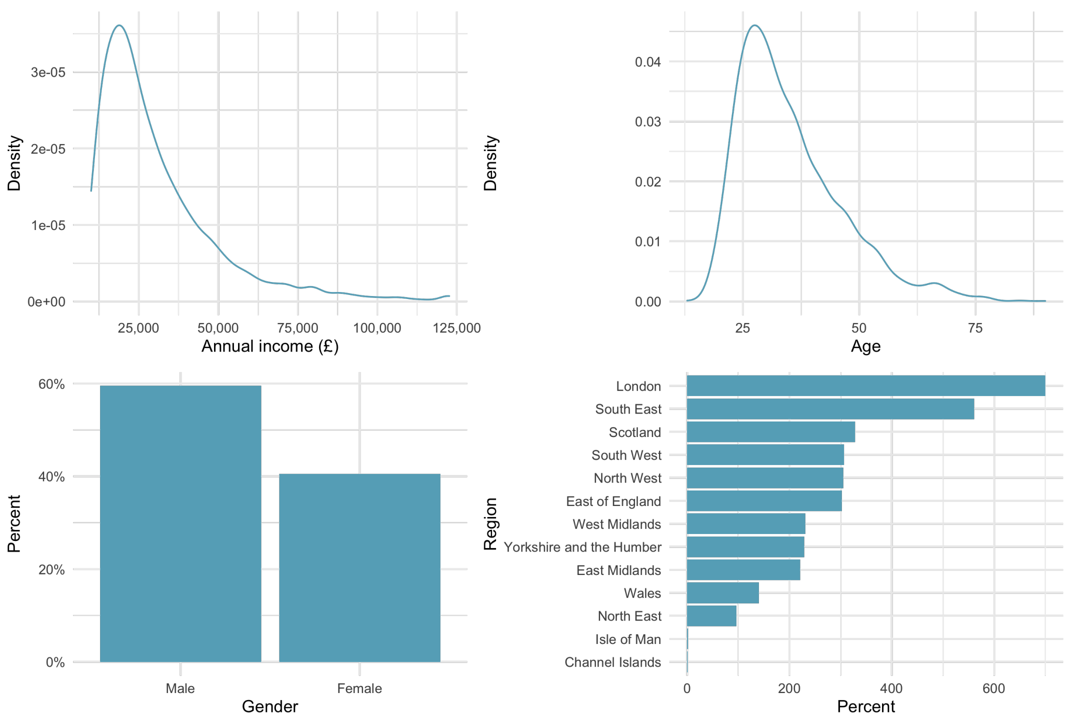
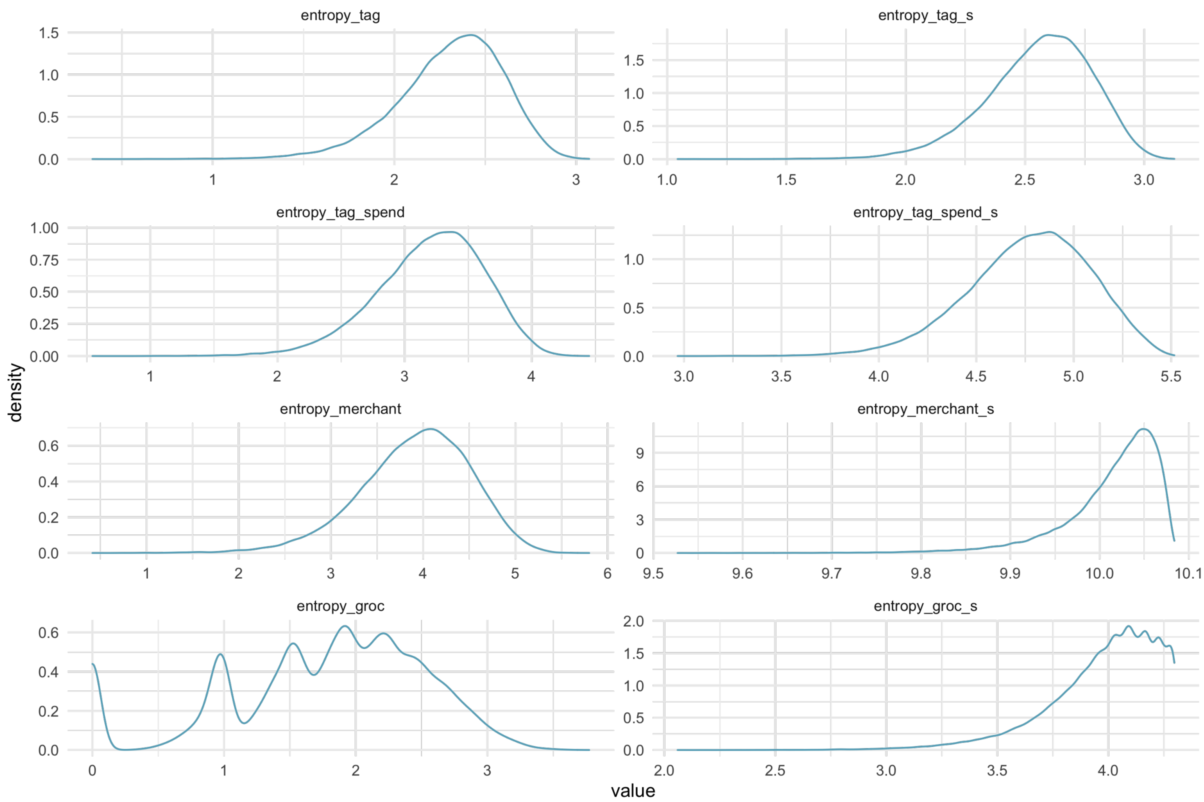



Figure 2.2

Figure 2.2: Entropy distributions




Notes:

### 2.2.6 Estimation

We estimate models of the form:

$$y_{i,t} = \alpha_i + \lambda_t + \beta H_{i,t} + x_{i,t}'\delta + \epsilon_{i,t}, \tag{2.4}$$

where $y_{i,t}$ is an indicator variable equal to one if individual $i$ made one or more transfers to any of their savings account in year-month period $t$ and zero otherwise, $H_{it}$ is $i$'s spending entropy in year-month period $t$, $x_{i,t}$ a vector of control variables, $\alpha_i$ an individual fixed effect, $\lambda_t$ a year-month fixed effect, and $\epsilon_{i,t}$ the error term.

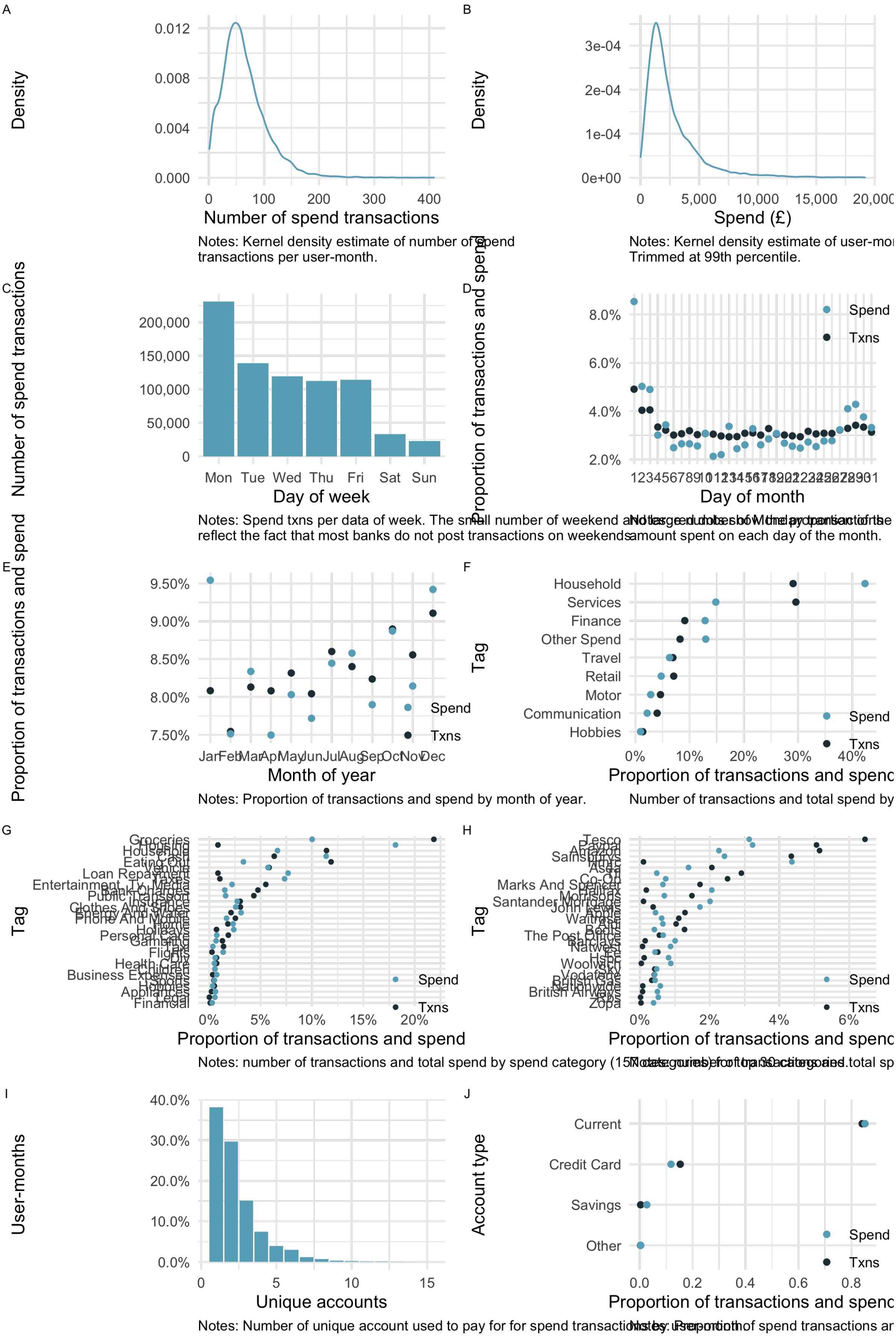
The vector of controls includes month spend, month income, an indicator for whether a user had positive income in a given month, and income variability, calculated as the standard deviation of month income over the previous 12 months.

Note that while we might in principle be worried about reverse causality, since making payments into savings accounts might lead to a non-zero count in an additional spend category and thus change entropy, this is not a concern here. As discussed in Section 2.2.3 and Section 2.2.4, we define savings as inflows into savings accounts and define entropy based on the classification of spend transactions on current accounts. If a user pays money from their current into one of their savings account, this will usually be labelled in their current account as a transfer and not enter the calculation of their entropy score. In Appendix .2.1, we provide robustness checks using lagged entropy scores, which produces very similar results.

## 2.3 Spending profiles

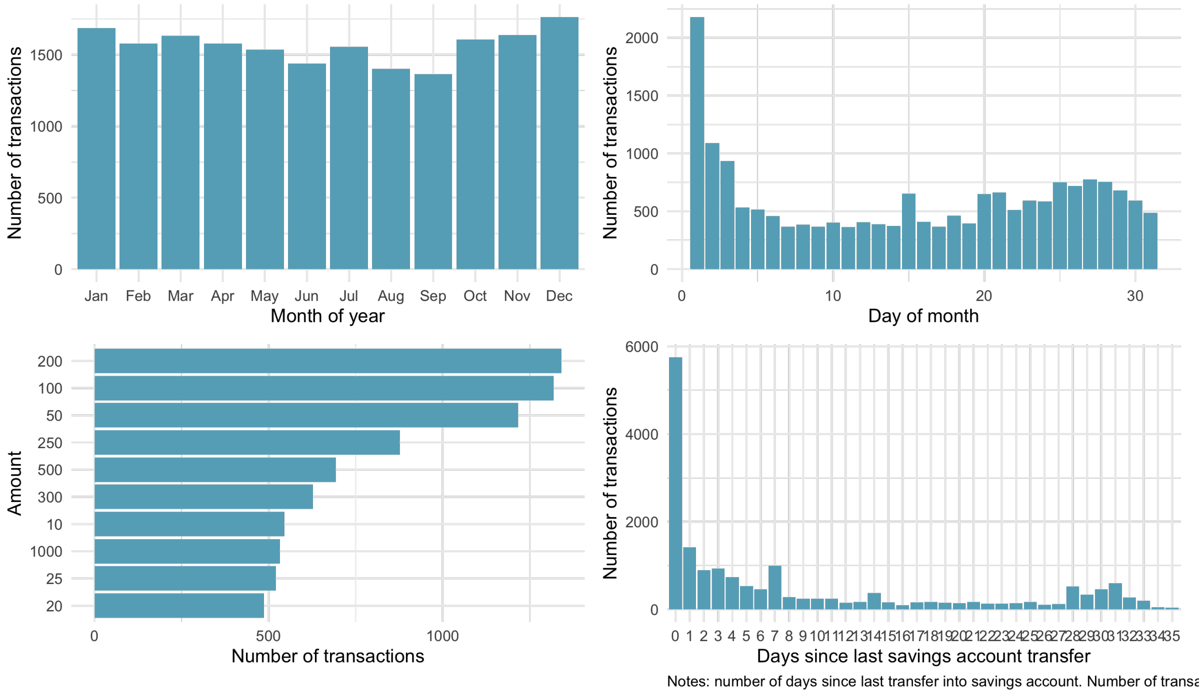
Figure 2.3

Figure 2.3: Spending behaviour



## 2.4 Savings patterns

Figure 2.4: Savings behaviour




## 2.5 Spending profiles predict emergency savings

Table 2.3 shows the effect of entropy on the probability of having emergency savings. Columns (1)-(3) show results for unsmoothed entropy based on 9 categories, 48 categories, and merchant names, respectively. Columns (4)-(6) results for smoothed entropy based on the same variables. All models include user and year-month fixed effects, and standard errors are clustered at the user-level. Confidence intervals are shown in brakets.

Results for unsmoothed entropy suggest that a one unit increase in entropy is associated with an increase in the probability of a user making at least one transfer into their savings accounts of between 1.5 and 2.7 percentage points – an effect up to two times larger than that of a £1000 increase in monthly income.

Conversely, the effect for unsmooth entropy is smaller in magnitude but runs in the reverse direction: a one-unit increase in the smoothed entropy score is associated with a reduction in the probability of transferring money into savings account of between 0.4 and 1.6 percentage points – an effect that, in absolute magnitude, is about equal to that of a £1000 increase in monthly income.

Overall, then, the effect of entropy in spending profiles is statistically and economically significant, and robust across different definitions. In other words, the scores seem to pick up a feature of the spending distribution that is predictive of savings behaviour.

Table 2.3: Effect of entropy on P(has savings)

| Model: | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| *Variables* | | | | | | |
| Entropy (9 cats) | 0.016*** | | | | | |
| | [0.013; 0.019] | | | | | |
| Entropy (48 cats) | | 0.029*** | | | | |
| | | [0.025; 0.033] | | | | |
| Entropy (merchant) | | | 0.032*** | | | |
| | | | [0.029; 0.036] | | | |
| Entropy (9 cats, smooth) | | | | -0.008*** | | |
| | | | | [-0.010; -0.006] | | |
| Entropy (48 cats, smooth) | | | | | -0.023*** | |
| | | | | | [-0.025; -0.020] | |
| Entropy (merchant, smooth) | | | | | | -0.019*** |
| | | | | | | [-0.021; -0.016] |
| Month spend | 0.009*** | 0.009*** | 0.008*** | 0.009*** | 0.008*** | 0.007*** |
| | [0.009; 0.010] | [0.008; 0.009] | [0.008; 0.009] | [0.009; 0.010] | [0.007; 0.009] | [0.007; 0.008] |
| Month income | 0.012*** | 0.012*** | 0.012*** | 0.012*** | 0.011*** | 0.011*** |
| | [0.011; 0.013] | [0.011; 0.013] | [0.011; 0.013] | [0.011; 0.013] | [0.011; 0.012] | [0.010; 0.012] |
| Has income in month | 0.086*** | 0.084*** | 0.083*** | 0.087*** | 0.085*** | 0.086*** |
| | [0.077; 0.094] | [0.075; 0.092] | [0.074; 0.091] | [0.079; 0.096] | [0.076; 0.093] | [0.078; 0.095] |
| Income variability | 0.001* | 0.001* | 0.001* | 0.001* | 0.001* | 0.000 |
| | [-0.000; 0.001] | [-0.000; 0.001] | [-0.000; 0.001] | [-0.000; 0.001] | [-0.000; 0.001] | [-0.000; 0.001] |
| *Fixed-effects* | | | | | | |
| User | Yes | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | | |
| Observations | 1,043,727 | 1,043,727 | 1,043,416 | 1,043,727 | 1,043,727 | 1,043,416 |
| $R^2$ | 0.45368 | 0.45395 | 0.45410 | 0.45363 | 0.45415 | 0.45410 |
| Within $R^2$ | 0.00719 | 0.00768 | 0.00807 | 0.00709 | 0.00805 | 0.00808 |

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

But how can we account for the opposite sign for smoothed and unsmoothed entropy scores? We don't have the answer to this yet...

### 2.5.1 Effect of financial resilience

Table 2.4: Effect of entropy on P(has savings) by income quintile

| Model: | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Income quintile: | 1 | 2 | 3 | 4 | 5 |
| *Variables* | | | | | |
| Entropy (48 cats) | 0.033*** | 0.019*** | 0.017*** | 0.016*** | 0.018*** |
| | [0.028; 0.039] | [0.013; 0.025] | [0.011; 0.023] | [0.010; 0.022] | [0.012; 0.025] |
| Month spend | 0.010*** | 0.007*** | 0.008*** | 0.007*** | 0.006*** |
| | [0.009; 0.011] | [0.006; 0.008] | [0.006; 0.009] | [0.006; 0.008] | [0.005; 0.007] |
| Month income | 0.015*** | 0.010*** | 0.012*** | 0.013*** | 0.008*** |
| | [0.012; 0.019] | [0.004; 0.017] | [0.005; 0.019] | [0.008; 0.019] | [0.007; 0.010] |
| Income variability | -0.000 | 0.001 | 0.002*** | 0.001** | -0.000 |
| | [-0.001; 0.001] | [-0.000; 0.002] | [0.001; 0.004] | [0.000; 0.002] | [-0.001; 0.001] |
| *Fixed-effects* | | | | | |
| User | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | |
| Observations | 223,462 | 215,151 | 208,417 | 201,319 | 195,378 |
| $R^2$ | 0.59138 | 0.58949 | 0.58354 | 0.58212 | 0.57162 |
| Within $R^2$ | 0.00535 | 0.00180 | 0.00211 | 0.00191 | 0.00271 |

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

Table 2.5: Effect of entropy on P(has savings) by income quintile

| Model: | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Income quintile: | 1 | 2 | 3 | 4 | 5 |
| *Variables* | | | | | |
| Entropy (48 cats, smooth) | -0.026*** | -0.020*** | -0.017*** | -0.015*** | -0.017*** |
| | [-0.030; -0.023] | [-0.024; -0.016] | [-0.021; -0.013] | [-0.019; -0.011] | [-0.021; -0.012] |
| Month spend | 0.009*** | 0.006*** | 0.007*** | 0.007*** | 0.006*** |
| | [0.008; 0.011] | [0.005; 0.007] | [0.006; 0.008] | [0.005; 0.008] | [0.005; 0.007] |
| Month income | 0.015*** | 0.010*** | 0.012*** | 0.013*** | 0.008*** |
| | [0.012; 0.019] | [0.004; 0.017] | [0.004; 0.019] | [0.007; 0.019] | [0.007; 0.010] |
| Income variability | -0.000 | 0.001 | 0.002*** | 0.001** | -0.000 |
| | [-0.001; 0.001] | [-0.000; 0.002] | [0.001; 0.004] | [0.000; 0.002] | [-0.001; 0.001] |
| *Fixed-effects* | | | | | |
| User | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | |
| Observations | 223,462 | 215,151 | 208,417 | 201,319 | 195,378 |
| $R^2$ | 0.59152 | 0.58970 | 0.58369 | 0.58223 | 0.57175 |
| Within $R^2$ | 0.00570 | 0.00230 | 0.00246 | 0.00216 | 0.00302 |

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

Table 2.6: Effect of entropy on P(has savings) by income variability quintile

| Model: | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Income variability quintile: | 1 | 2 | 3 | 4 | 5 |
| *Variables* | | | | | |
| Entropy (48 cats) | 0.020*** | 0.014*** | 0.022*** | 0.028*** | 0.021*** |
| | [0.014; 0.026] | [0.008; 0.020] | [0.016; 0.028] | [0.022; 0.034] | [0.015; 0.027] |
| Month spend | 0.008*** | 0.008*** | 0.007*** | 0.007*** | 0.007*** |
| | [0.007; 0.009] | [0.007; 0.009] | [0.006; 0.008] | [0.006; 0.008] | [0.006; 0.008] |
| Month income | 0.014*** | 0.010*** | 0.011*** | 0.010*** | 0.010*** |
| | [0.011; 0.017] | [0.008; 0.012] | [0.010; 0.013] | [0.009; 0.012] | [0.009; 0.011] |
| Has income in month | 0.092*** | 0.076*** | 0.064*** | 0.055*** | 0.067*** |
| | [0.073; 0.111] | [0.056; 0.095] | [0.047; 0.081] | [0.040; 0.070] | [0.053; 0.081] |
| Income variability | -0.001 | 0.006** | -0.000 | 0.000 | -0.004** |
| | [-0.003; 0.001] | [0.001; 0.011] | [-0.004; 0.003] | [-0.001; 0.002] | [-0.007; -0.001] |
| *Fixed-effects* | | | | | |
| User | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | |
| Observations | 223,462 | 215,151 | 208,417 | 201,319 | 195,378 |
| $R^2$ | 0.61166 | 0.59951 | 0.59062 | 0.59379 | 0.63519 |
| Within $R^2$ | 0.00558 | 0.00398 | 0.00484 | 0.00579 | 0.00781 |

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

Table 2.7: Effect of entropy on P(has savings) by income variability quintile

| Model: | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Income variability quintile: | 1 | 2 | 3 | 4 | 5 |
| *Variables* | | | | | |
| Entropy (48 cats, smooth) | -0.019*** | -0.020*** | -0.017*** | -0.022*** | -0.024*** |
| | [-0.024; -0.015] | [-0.024; -0.016] | [-0.021; -0.013] | [-0.026; -0.018] | [-0.028; -0.019] |
| Month spend | 0.007*** | 0.007*** | 0.006*** | 0.006*** | 0.006*** |
| | [0.006; 0.008] | [0.006; 0.008] | [0.005; 0.008] | [0.005; 0.007] | [0.005; 0.007] |
| Month income | 0.014*** | 0.010*** | 0.011*** | 0.010*** | 0.010*** |
| | [0.011; 0.017] | [0.008; 0.012] | [0.009; 0.013] | [0.009; 0.012] | [0.009; 0.011] |
| Has income in month | 0.093*** | 0.075*** | 0.064*** | 0.056*** | 0.066*** |
| | [0.074; 0.112] | [0.055; 0.094] | [0.047; 0.081] | [0.040; 0.071] | [0.052; 0.081] |
| Income variability | -0.001 | 0.005** | -0.000 | 0.000 | -0.004** |
| | [-0.003; 0.001] | [0.001; 0.010] | [-0.004; 0.003] | [-0.001; 0.002] | [-0.006; -0.001] |
| *Fixed-effects* | | | | | |
| User | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | |
| Observations | 223,462 | 215,151 | 208,417 | 201,319 | 195,378 |
| $R^2$ | 0.61181 | 0.59979 | 0.59070 | 0.59394 | 0.63550 |
| Within $R^2$ | 0.00597 | 0.00466 | 0.00502 | 0.00614 | 0.00864 |

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

## 2.6 Discussion

# .1 Interpreting entropy

To see how we can interpret entropy as the predictability of a user's spending behaviour, it is useful to have a more complete understanding of Equation 2.1. The building blocks of entropy is the information content of a single event. The key intuition Shannon (1948) aimed to capture was that learning of the occurrence of a low-probability event is more informative than learning of the occurrence of a high-probability event. The information of an event $I(E)$ is thus inversely proportional to is probability $p(E)$. One way to capture this would be to define the information of event E as $I(E) = \frac{1}{p(E)}$. Yet this implied that an event that is certain to occur had information 1, when it would make sense to have information 0. To remedy this (and also satisfy additional desireable characteristics of an information function), we can use the log of the expression. Hence, the information of event E, often called *Shannon information*, *self-information*, or just *information*, is defined as:

$$I(E) = log\left(\frac{1}{p(E)}\right) = -log(p(E)). \tag{5}$$

Entropy, often called *Information entropy*, *Shannon entropy*, or just *entropy*, is the information of a random variable, $X$, and captures the expected amount of information of an event drawn at random from the probability distribution of the random variable. It is calcualted as:

$$H(X) = -\sum_x p(x) \times log(p(x)) = \sum_x p(x)I(x) = \mathbb{E}I(x). \tag{6}$$

For a single event, the key intution was that the less likely an event, the more information is conveyed when it occurs. The related idea for distributions is similar: the less skewed a distribution of a random variable, the less certain the realised value of a single draw from the distribution, the higher is entropy - the maximum entropy distribution is the uniform distribution.

# .2 Additional results

## .2.1 Endogeneity

As discussed in Section 2.2.6, one concern one might have about our results in Section A.4 is reverse causality: transferring money into savings accounts might change the distribution of spend categories and thus change entropy. As noted previously, this is not a major concern because of the way we calculate savings and spend profiles: savings are calculated as the sum of inflows into savings accounts, while spend profiles are based on the classification of spend transactions in current accounts, and transfers from current to savings accounts are labelled as such and not treated as spend transactions.

However, because transaction labelling is imperfect, it is possible that some transfers are misclassified as spends and included in the calculation of entropy scores. One way to deal with this is to lag entropy scores by one period. Table 8 presents results similar to the main results in the main text, but using entropy lagged by one year-month period as the independent variable of interest. We can see that the results are very similar to thos presented above.

Table 8: Effect of entropy on P(has savings)

| Model: | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| *Variables* | | | | | | |
| Entropy lag (9 cats) | 0.015*** | | | | | |
| | [0.012; 0.019] | | | | | |
| Entropy lag (48 cats) | | 0.024*** | | | | |
| | | [0.021; 0.028] | | | | |
| Entropy lag (merchant) | | | 0.027*** | | | |
| | | | [0.023; 0.030] | | | |
| Entropy lag (9 cats, smooth) | | | | -0.004*** | | |
| | | | | [-0.006; -0.002] | | |
| Entropy lag (48 cats, smooth) | | | | | -0.016*** | |
| | | | | | [-0.018; -0.013] | |
| Entropy lag (merchant, smooth) | | | | | | -0.013*** |
| | | | | | | [-0.016; -0.011] |
| Month spend | 0.009*** | 0.009*** | 0.009*** | 0.009*** | 0.009*** | 0.008*** |
| | [0.009; 0.010] | [0.008; 0.010] | [0.008; 0.010] | [0.009; 0.010] | [0.008; 0.009] | [0.008; 0.009] |
| Month income | 0.012*** | 0.012*** | 0.012*** | 0.012*** | 0.012*** | 0.011*** |
| | [0.011; 0.013] | [0.011; 0.013] | [0.011; 0.013] | [0.011; 0.013] | [0.011; 0.013] | [0.010; 0.012] |
| Has income in month | 0.083*** | 0.082*** | 0.082*** | 0.084*** | 0.083*** | 0.084*** |
| | [0.074; 0.092] | [0.073; 0.091] | [0.073; 0.091] | [0.075; 0.093] | [0.074; 0.092] | [0.075; 0.092] |
| Income variability | 0.001* | 0.001* | 0.001* | 0.001* | 0.001* | 0.001* |
| | [-0.000; 0.001] | [-0.000; 0.001] | [-0.000; 0.001] | [-0.000; 0.001] | [-0.000; 0.001] | [-0.000; 0.001] |
| *Fixed-effects* | | | | | | |
| User | Yes | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | | |
| Observations | 1,009,843 | 1,009,843 | 1,009,554 | 1,009,843 | 1,009,843 | 1,009,554 |
| $R^2$ | 0.45785 | 0.45802 | 0.45809 | 0.45775 | 0.45803 | 0.45800 |
| Within $R^2$ | 0.00691 | 0.00722 | 0.00748 | 0.00671 | 0.00724 | 0.00731 |

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

## .2.2 Further results by financial resilience

The Tables below show results presented in Section 2.5.1 for alternative entropy variables.

Table 9: Effect of entropy on P(has savings) by income quintile

| Model: | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Income quintile: | 1 | 2 | 3 | 4 | 5 |
| *Variables* | | | | | |
| Entropy (9 cats) | 0.022*** | 0.006** | 0.009*** | 0.007** | 0.006** |
| | [0.017; 0.026] | [0.001; 0.011] | [0.004; 0.015] | [0.002; 0.013] | [0.000; 0.012] |
| Month spend | 0.011*** | 0.007*** | 0.008*** | 0.007*** | 0.007*** |
| | [0.010; 0.012] | [0.006; 0.009] | [0.007; 0.009] | [0.006; 0.008] | [0.005; 0.008] |
| Month income | 0.016*** | 0.011*** | 0.012*** | 0.013*** | 0.008*** |
| | [0.013; 0.019] | [0.004; 0.017] | [0.005; 0.020] | [0.008; 0.019] | [0.007; 0.010] |
| Income variability | 0.000 | 0.001 | 0.002*** | 0.001** | -0.000 |
| | [-0.001; 0.001] | [-0.000; 0.002] | [0.001; 0.004] | [0.000; 0.002] | [-0.001; 0.001] |
| *Fixed-effects* | | | | | |
| User | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | |
| Observations | 223,462 | 215,151 | 208,417 | 201,319 | 195,378 |
| $R^2$ | 0.59111 | 0.58937 | 0.58347 | 0.58205 | 0.57151 |
| Within $R^2$ | 0.00469 | 0.00151 | 0.00193 | 0.00175 | 0.00246 |

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

Table 10: Effect of entropy on P(has savings) by income quintile

| Model: | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Income quintile: | 1 | 2 | 3 | 4 | 5 |
| *Variables* | | | | | |
| Entropy (9 cats, smooth) | -0.008*** | -0.009*** | -0.006*** | -0.006*** | -0.008*** |
| | [-0.012; -0.005] | [-0.012; -0.006] | [-0.009; -0.002] | [-0.009; -0.002] | [-0.012; -0.005] |
| Month spend | 0.011*** | 0.007*** | 0.008*** | 0.007*** | 0.006*** |
| | [0.010; 0.012] | [0.006; 0.008] | [0.007; 0.009] | [0.006; 0.008] | [0.005; 0.007] |
| Month income | 0.017*** | 0.011*** | 0.012*** | 0.013*** | 0.008*** |
| | [0.013; 0.020] | [0.004; 0.017] | [0.005; 0.020] | [0.008; 0.019] | [0.007; 0.010] |
| Income variability | 0.000 | 0.001 | 0.002*** | 0.001** | -0.000 |
| | [-0.001; 0.001] | [-0.000; 0.002] | [0.001; 0.004] | [0.000; 0.002] | [-0.001; 0.001] |
| *Fixed-effects* | | | | | |
| User | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | |
| Observations | 223,462 | 215,151 | 208,417 | 201,319 | 195,378 |
| $R^2$ | 0.59095 | 0.58944 | 0.58347 | 0.58207 | 0.57158 |
| Within $R^2$ | 0.00431 | 0.00168 | 0.00193 | 0.00179 | 0.00260 |

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

Table 11: Effect of entropy on P(has savings) by income variability quintile

| Model: | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Income variability quintile: | 1 | 2 | 3 | 4 | 5 |
| *Variables* | | | | | |
| Entropy (9 cats) | 0.011*** | 0.005* | 0.012*** | 0.016*** | 0.011*** |
| | [0.006; 0.016] | [-0.001; 0.010] | [0.006; 0.018] | [0.010; 0.022] | [0.006; 0.017] |
| Month spend | 0.008*** | 0.008*** | 0.007*** | 0.007*** | 0.007*** |
| | [0.007; 0.009] | [0.007; 0.009] | [0.006; 0.008] | [0.006; 0.009] | [0.006; 0.008] |
| Month income | 0.015*** | 0.010*** | 0.011*** | 0.010*** | 0.010*** |
| | [0.012; 0.017] | [0.008; 0.012] | [0.010; 0.013] | [0.009; 0.012] | [0.009; 0.011] |
| Has income in month | 0.094*** | 0.077*** | 0.065*** | 0.057*** | 0.068*** |
| | [0.075; 0.113] | [0.057; 0.096] | [0.048; 0.082] | [0.041; 0.072] | [0.054; 0.083] |
| Income variability | -0.001 | 0.006** | -0.000 | 0.001 | -0.004*** |
| | [-0.003; 0.002] | [0.001; 0.011] | [-0.004; 0.003] | [-0.001; 0.002] | [-0.007; -0.001] |
| *Fixed-effects* | | | | | |
| User | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | |
| Observations | 223,462 | 215,151 | 208,417 | 201,319 | 195,378 |
| $R^2$ | 0.61156 | 0.59945 | 0.59051 | 0.59361 | 0.63509 |
| Within $R^2$ | 0.00532 | 0.00383 | 0.00456 | 0.00534 | 0.00753 |

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

Table 12: Effect of entropy on P(has savings) by income variability quintile

| Model: | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Income variability quintile: | 1 | 2 | 3 | 4 | 5 |
| *Variables* | | | | | |
| Entropy (9 cats, smooth) | -0.007*** | -0.009*** | -0.005*** | -0.007*** | -0.010*** |
| | [-0.011; -0.004] | [-0.013; -0.006] | [-0.009; -0.002] | [-0.010; -0.003] | [-0.014; -0.007] |
| Month spend | 0.008*** | 0.008*** | 0.007*** | 0.007*** | 0.007*** |
| | [0.007; 0.009] | [0.007; 0.009] | [0.006; 0.008] | [0.006; 0.009] | [0.006; 0.008] |
| Month income | 0.015*** | 0.010*** | 0.012*** | 0.011*** | 0.010*** |
| | [0.012; 0.018] | [0.008; 0.012] | [0.010; 0.013] | [0.009; 0.012] | [0.009; 0.011] |
| Has income in month | 0.095*** | 0.077*** | 0.066*** | 0.058*** | 0.069*** |
| | [0.076; 0.115] | [0.057; 0.096] | [0.049; 0.083] | [0.042; 0.073] | [0.055; 0.083] |
| Income variability | -0.001 | 0.006** | -0.000 | 0.000 | -0.004** |
| | [-0.003; 0.001] | [0.001; 0.011] | [-0.004; 0.003] | [-0.001; 0.002] | [-0.007; -0.001] |
| *Fixed-effects* | | | | | |
| User | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | |
| Observations | 223,462 | 215,151 | 208,417 | 201,319 | 195,378 |
| $R^2$ | 0.61156 | 0.59954 | 0.59048 | 0.59355 | 0.63514 |
| Within $R^2$ | 0.00533 | 0.00404 | 0.00448 | 0.00519 | 0.00767 |

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

Table 13: Effect of entropy on P(has savings) by income quintile

| Model: | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Income quintile: | 1 | 2 | 3 | 4 | 5 |
| *Variables* | | | | | |
| Entropy (merchant) | 0.032*** | 0.027*** | 0.024*** | 0.017*** | 0.021*** |
| | [0.027; 0.037] | [0.022; 0.033] | [0.019; 0.030] | [0.011; 0.023] | [0.015; 0.027] |
| Month spend | 0.010*** | 0.006*** | 0.007*** | 0.007*** | 0.006*** |
| | [0.009; 0.011] | [0.005; 0.008] | [0.006; 0.008] | [0.006; 0.008] | [0.005; 0.007] |
| Month income | 0.015*** | 0.010*** | 0.012*** | 0.013*** | 0.008*** |
| | [0.012; 0.019] | [0.004; 0.017] | [0.005; 0.019] | [0.008; 0.019] | [0.007; 0.010] |
| Income variability | -0.000 | 0.001 | 0.002*** | 0.001** | -0.000 |
| | [-0.001; 0.001] | [-0.000; 0.002] | [0.001; 0.004] | [0.000; 0.002] | [-0.001; 0.001] |
| *Fixed-effects* | | | | | |
| User | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | |
| Observations | 223,357 | 215,087 | 208,364 | 201,280 | 195,328 |
| $R^2$ | 0.59137 | 0.58963 | 0.58366 | 0.58211 | 0.57163 |
| Within $R^2$ | 0.00544 | 0.00229 | 0.00245 | 0.00198 | 0.00288 |

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

Table 14: Effect of entropy on P(has savings) by income quintile

| Model: | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Income quintile: | 1 | 2 | 3 | 4 | 5 |
| *Variables* | | | | | |
| Entropy (merchant, smooth) | -0.025*** | -0.017*** | -0.014*** | -0.013*** | -0.015*** |
| | [-0.029; -0.022] | [-0.021; -0.013] | [-0.018; -0.011] | [-0.018; -0.009] | [-0.019; -0.012] |
| Month spend | 0.009*** | 0.006*** | 0.007*** | 0.006*** | 0.005*** |
| | [0.008; 0.010] | [0.005; 0.007] | [0.005; 0.008] | [0.005; 0.007] | [0.004; 0.006] |
| Month income | 0.015*** | 0.010*** | 0.011*** | 0.013*** | 0.008*** |
| | [0.012; 0.019] | [0.003; 0.017] | [0.004; 0.018] | [0.007; 0.019] | [0.007; 0.010] |
| Income variability | -0.000 | 0.001 | 0.002*** | 0.001** | -0.000 |
| | [-0.001; 0.001] | [-0.000; 0.002] | [0.001; 0.004] | [0.000; 0.002] | [-0.001; 0.001] |
| *Fixed-effects* | | | | | |
| User | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | |
| Observations | 223,357 | 215,087 | 208,364 | 201,280 | 195,328 |
| $R^2$ | 0.59153 | 0.58963 | 0.58368 | 0.58225 | 0.57180 |
| Within $R^2$ | 0.00581 | 0.00230 | 0.00249 | 0.00232 | 0.00327 |

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

Table 15: Effect of entropy on P(has savings) by income variability quintile

| Model: | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Income variability quintile: | 1 | 2 | 3 | 4 | 5 |
| *Variables* | | | | | |
| Entropy (merchant) | 0.025*** | 0.020*** | 0.025*** | 0.036*** | 0.026*** |
| | [0.020; 0.031] | [0.014; 0.025] | [0.019; 0.031] | [0.030; 0.042] | [0.020; 0.031] |
| Month spend | 0.007*** | 0.008*** | 0.007*** | 0.006*** | 0.007*** |
| | [0.006; 0.009] | [0.006; 0.009] | [0.005; 0.008] | [0.005; 0.007] | [0.005; 0.008] |
| Month income | 0.014*** | 0.010*** | 0.011*** | 0.010*** | 0.010*** |
| | [0.011; 0.017] | [0.008; 0.012] | [0.010; 0.013] | [0.009; 0.012] | [0.009; 0.011] |
| Has income in month | 0.091*** | 0.075*** | 0.063*** | 0.054*** | 0.066*** |
| | [0.072; 0.111] | [0.055; 0.094] | [0.046; 0.080] | [0.039; 0.069] | [0.052; 0.080] |
| Income variability | -0.001 | 0.006** | -0.000 | 0.000 | -0.004** |
| | [-0.003; 0.001] | [0.001; 0.010] | [-0.004; 0.003] | [-0.001; 0.002] | [-0.007; -0.001] |
| *Fixed-effects* | | | | | |
| User | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | |
| Observations | 223,369 | 215,092 | 208,349 | 201,276 | 195,330 |
| $R^2$ | 0.61172 | 0.59956 | 0.59064 | 0.59404 | 0.63529 |
| Within $R^2$ | 0.00592 | 0.00422 | 0.00506 | 0.00644 | 0.00814 |

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

Table 16: Effect of entropy on P(has savings) by income variability quintile

| Model: | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Income variability quintile: | 1 | 2 | 3 | 4 | 5 |
| *Variables* | | | | | |
| Entropy (merchant, smooth) | -0.018*** | -0.018*** | -0.016*** | -0.017*** | -0.016*** |
| | [-0.021; -0.014] | [-0.022; -0.014] | [-0.019; -0.012] | [-0.020; -0.013] | [-0.021; -0.012] |
| Month spend | 0.007*** | 0.006*** | 0.006*** | 0.006*** | 0.006*** |
| | [0.006; 0.008] | [0.005; 0.008] | [0.005; 0.007] | [0.005; 0.007] | [0.005; 0.007] |
| Month income | 0.014*** | 0.010*** | 0.011*** | 0.010*** | 0.010*** |
| | [0.011; 0.017] | [0.007; 0.012] | [0.009; 0.013] | [0.009; 0.011] | [0.009; 0.011] |
| Has income in month | 0.094*** | 0.075*** | 0.065*** | 0.057*** | 0.068*** |
| | [0.075; 0.113] | [0.056; 0.095] | [0.048; 0.082] | [0.042; 0.072] | [0.054; 0.082] |
| Income variability | -0.001 | 0.005** | -0.000 | 0.000 | -0.004** |
| | [-0.003; 0.001] | [0.000; 0.010] | [-0.004; 0.003] | [-0.001; 0.002] | [-0.006; -0.001] |
| *Fixed-effects* | | | | | |
| User | Yes | Yes | Yes | Yes | Yes |
| Year-month | Yes | Yes | Yes | Yes | Yes |
| *Fit statistics* | | | | | |
| Observations | 223,369 | 215,092 | 208,349 | 201,276 | 195,330 |
| $R^2$ | 0.61177 | 0.59980 | 0.59070 | 0.59387 | 0.63537 |
| Within $R^2$ | 0.00607 | 0.00480 | 0.00520 | 0.00602 | 0.00836 |

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

# Appendix A

# Do Money Dashboard users spend more and save less?

## A.1 Introduction

This paper evaluates whether Money Dashboard, a UK-based financial aggregator app, helps its users reduce their discretionary spend and increase their "rainy-day savings".

The question is important because a large number of adults in the UK and the US do not have enough savings to cover unexpected expenses like car or medical bills: in the UK, 25 percent of adults would be unable to cover an unexpected bill of £300 (Philipps et al. 2021), while in the US, about 30 percent would be unable to cover a $400 bill (Governors of the Federal Reserve System 2022). But while there is a large body of research that studies reasons for why savings are low, little is known about what could help people save more.[1]

FinTech aggregator apps like Money Dashboard are promising because they provide easy access to financial information that make it easier to monitor ones spending and saving, and often also offer tools such as budgeting and the setting of spending goals. Financial information that is more easily accessible and is aggregated in ways that help people keep track of their goals might be beneficial because rational inattention theory predicts that it makes people more likely to access that information, which, in turn, might lead to better consumption decisions. Similarly, tools that help with budgeting and with

[1]Well-documented behavioural biases that help explain undersaving are, among others, present bias (Laibson 1997, Laibson and Marzilli-Ericson 2019), inertia (Madrian and Shea 2001), over-extrapolation (Choi et al. 2009), and limited self-control and willpower (R. H. Thaler and Shefrin 1981, Benhabib and Bisin 2005, Fudenberg and Levine 2006, Loewenstein and O'Donoghue 2004, Gul and Pesendorfer 2001). One danger of viewing low savings mainly as a result of behavioural biases is that while these biases likely do play some role and designing environments and tools to help correct them are thus part of the solution, it is at least conceivable that this is an area where the focus on behaviour-level solutions distracts from an effort to find more effective society-level solutions, a danger inherent in behavioural science research convincingly highlighted in Chater and Loewenstein (2022): if the main problem is that many people are unable to earn enough to save, then the effectiveness of helping them manage their low incomes more effectively pales in comparison with efforts to help them earn more.

setting spending goals have the potential to help users make consumption decisions more in line with their intentions because such tools can act as commitment devices that – if users experience disutility from falling short of their goals – introduce a cognitive cost to overspending or undersaving.[2]

The nascent literature that studies the effect of FinTech apps on financial outcomes suggests that these apps can indeed lead to improved financial outcomes: they have been found to reduce spending by providing users with information about their spending relative to peers (D'Acunto et al. 2020) and offering budgeting options (Lukas and Howard 2022), to increase savings by offering budgeting options (Gargano and Rossi 2021), and to reduce non-sufficient fund fees by facilitating access to information (Carlin et al. 2022).

In this paper, I specifically test whether using Money Dashboard is associated with a reduction in discretionary spending and an increase in "rainy-day savings". I use a new estimator proposed by Callaway and Sant'Anna (2021) that corrects for recently identified problems in two-way fixed effects estimates.

I find that users reduce their discretionary spend by between £100 and £150 (11-17% of average discretionary spend) once they start using the app and sustain that reduction throughout the six-month post-signup period I consider. Looking at disaggregated measures of discretionary spend further shows that the reduction is the result of maintained month-to-month changes in behaviour rather than one-off cancellations of direct-debit transactions, that it results from reducing spending on a number of different categories of purchases rather than a single one, and that it is a result of changes along the extensive rather than the intensive margin – users reduce the number of transactions they make rather than the value of the average transaction. Interestingly, users do not seem to use these additional funds to build up "rainy-day savings": net-inflows into savings accounts do not change after signup. I can also neither find significant increases in flows into investment and pension accounts or additional savings accounts that are not linked to the app, not additional loan repayments.

My work mainly contributes to three strands of the literature. The first, is the aforementioned recent literature that studies the effect of FinTech apps on financial outcomes. The second, is the very recent literature on studying interventions to help increase "rainy-day savings" an area of household finances that has until recently had no attention. In addition to studies testing the effect of FinTech apps on savings, there is a strand of research that studies the use of auto-enrolment into employer-sponsored savings accounts – similar to the ones used to increase pension savings (R. H. Thaler and Benartzi 2004, Choi et al. 2004, Choukhmane 2019) – and finds an increase in both participation (relative to opt-in accounts) and account balances (Beshears, Choi, et al. 2020, Berk et al. 2022). Finally, my

[2] On rational inattention theory, see, for instance, Brunnermeier and Nagel (2008), DellaVigna (2009), and Sims (2003). On commitment devices see, among others, R. H. Thaler and Shefrin (1981), Laibson (1997), and O'Donoghue and Rabin (1999a) for theoretical foundations, and Beshears, Milkman, et al. (2016) and Hsiaw (2013) for a discussion of soft commitment devices.

work also contributes to a rapidly growing literature of using financial-transaction data from banks or financial aggregator apps to understand consumer financial behaviour. As already mentioned, Kuchler and Pagel (2020) use data from a financial aggregator app to estimate time preferences. Similar data has been used to show that consumer spending varies across the pay cycle (Gelman et al. 2014, Olafsson and Pagel 2018), to test the consumer spending response to exogenous shocks (S. R. Baker 2018, Baugh et al. 2014), and to better understand the generational differences in financial platform usage patterns (Carlin et al. 2019). Some researchers use transaction-data directly provided by banks. Ganong and Noel (2019) show that consumer spending drops sharply after the predictable income drop from exhausting unemployment insurance benefits, Meyer and Pagel (2018) analyse how individuals reinvest realised capital gains and losses, and Muggleton et al. (2020) show that chaotic spending behaviour is a harbinger of financial distress.[3]

There are two main limitations to my approach. First, the data is not generated by a randomised experiment. The gold-standard to evaluate whether use of Money Dashboard improves financial outcomes would be a randomised controlled-trial, where out of a sample of potential users (ideally random and representative of the UK population), we would randomly grant access to the app to some users and then compare outcomes of those treated users with the control group of users who did not have access. Instead, the data I have access to only contains data for individuals who self-selected into using the app. Individuals will choose to do so for a number of different reasons, all of which are unobserved in the data, and at least some of which would probably have changed their financial outcomes even if they had not signed up to the app. Any changes in financial outcomes we observe are thus "aggregate" or "net" effects of these unobservables and the "pure" causal effect of app use.

To see this, think of the net effect as $\textit{net effect} = \textit{causal effect} + \textit{``need''}(\downarrow) + \textit{``motivation''}(\uparrow)$, where the arrows indicate the direction of the bias, and consider three cases that illustrate three stylised but plausible scenarios for signup. First, consider a user who signs up in the hope that the app will help them reign in discretionary spending that has has gotten out of hand. If it takes the user some time to fully adjust their spending, then even if the app does help them make these adjustments, the estimated positive effect of app use will be biased downward. Next, consider a user who decides to start bringing their own lunch to work instead of eating out in an effort to save for a new car and signs up to MDB in the hope that the app will help them keep track of their spending. Such a user would probably have reduced their discretionary spend even if they had not signed up to the app, thus creating an upward bias on our estimated net effect. Finally, consider a user who signs up to MDB purely because they happened to see an advert for the app on the Bus and got curious. In this case, we can think of signup being close to random –

[3]For a comprehensive review of the literature using financial transaction data, see S. R. Baker and Kueng (2022).

almost as if the user had been allocated to the treatment group in our ideal experiment – and the estimated net effect will closely resemble the causal effect of app use. Hence, under the weak assumption that at least some users sign up for reasons that are not as good as random, our estimated effects will be biased upwards or downwards depending on the relative proportion of users whose unobservable reasons for signup create an upward and downward bias.

The second limitation is that even if I were able to isolate the effect of the app, I am not able to differentiate between the contributions of different features of the app such as improved access to information and budgeting.

Despite these limitations, which mean that we cannot interpret results as causal effects of using Money Dashboard, the finding that app use is associated with an economically significant and sustained drop in discretionary spend, and the finding that this reduction comes about by users making fewer transactions in a number of different spending categories month-by-month provides interesting insight of how users reduce discretionary spend, either because of or simply while using Money Dashboard. Furthermore, the results make it very plausible that the Money Dashboard, and possibly apps like it, does have a positive causal effect, and thus suggests that further research, focused on identifying the causal effect of the app overall as well as that of its component parts, would be interesting and worthwhile.

The remainder of this paper is organised as follows: Section A.2 introduces the dataset used, discusses preprocessing and presents summary statistics; Section A.3 introduces the empirical approach used in the analysis; Section A.4 presents the results; and Section A.5 concludes. To make it easier for interested readers to clarify questions about details and subtleties of data preprocessing and analysis steps, I provide links to the scripts that implement the steps discussed in the text in the relevant places throughout the text.[4]

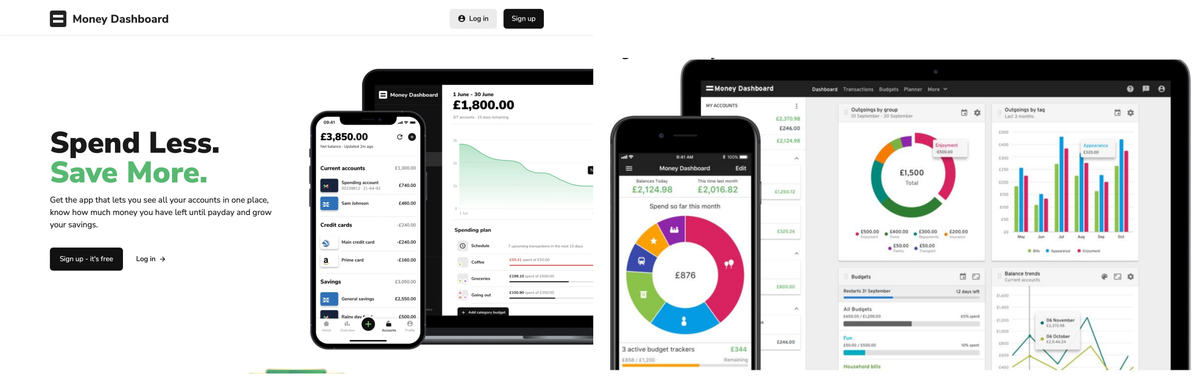
## A.2 Data

### A.2.1 Dataset description

I use data from Money Dashboard (MDB), a UK-based financial aggregation app that allows users to link accounts from different banks to obtain an integrated view of their finances in order to help them to reduce their spending and increase their savings. Figure **??** shows a screenshot from MDB's website (left) and an example of its user interface (right). The complete dataset contains more than 500 million transactions made between 2012 and June 2020 by about 270,000 users, and provides information such as date, amount, and description of the transaction as well as account and user-level information.

[4] The projects GitHub repo that contains all files used to produce the results can be found at https://github.com/fabiangunzinger/mdb_eval.

Figure A.1: Money Dashboard website and user interface




Notes: Screenshot from the top of the Money Dashboard website (left) and example of its user interface (right), illustrating some of the transaction grouping and budgeting functions.

Crucially, for this paper, MDB can access up to three years of historic data for each linked account. Figure A.2 shows how we can leverage pre-signup data provided in the MDB data to construct valid control groups. Each row of cells shows data for one of 200 randomly selected users, with red cells indicating data from pre-signup months and blue cells indicating data from post-signup months. Signup thus occured sometime during the first blue coloured month, and we can use users who have not yet signed up (whose cells are still red) as a comparison group. In my analysis below, I will use this data structure to compare the outcomes of users who start using the Money Dashboard app at a particular point in time with users who have not yet started using the app at that time.

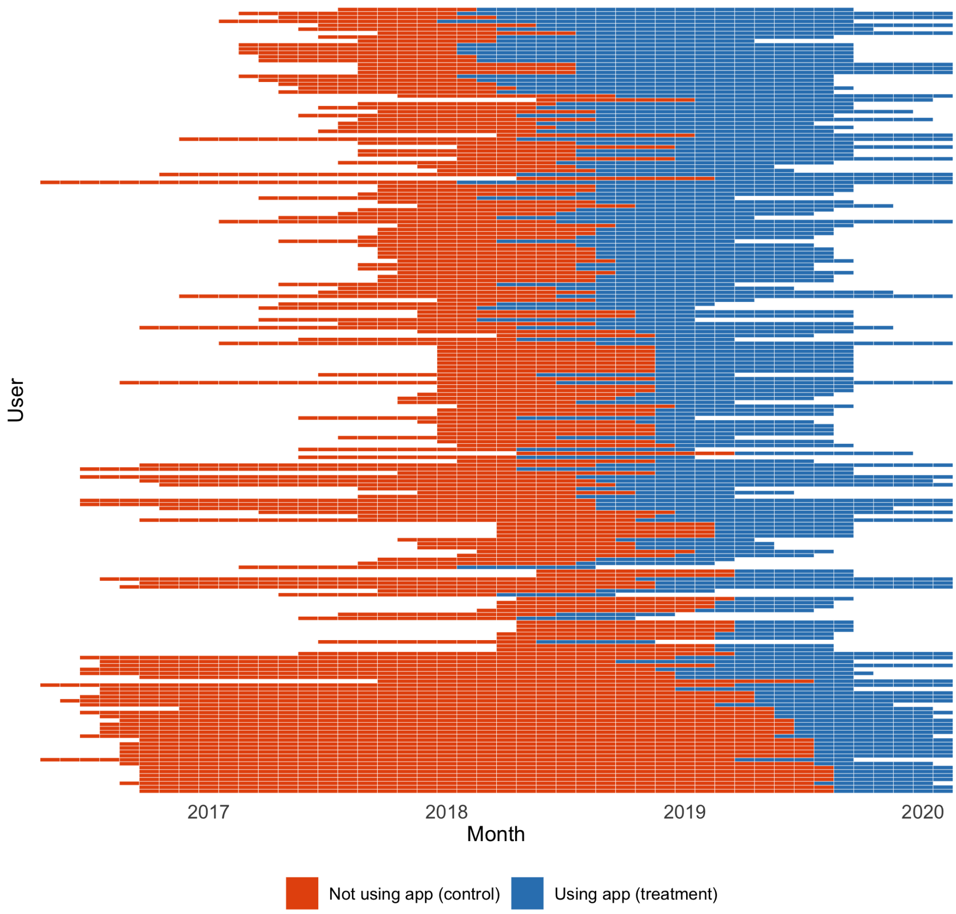
The data's main advantage for the study of consumer financial behaviour is that it allows us to observe all savings and spending transactions for users who linked all their financial accounts. This is important since it means that we can be sure that any changes we observe in some of a user's account are not offset by changes in unliked accounts.[5] Furthermore, data is collected automatically and in real-time rather than through surveys that collect data with a time-lag and often rely on consumers's ability and willingness to provide accurate information.

The data's main limitation is that because users's self select into using MDB, the sample is not representative of the wider UK population: it is well documented that FinTech app users are more likely to be male, younder, and higher-income earners than the average person Carlin et al. (2019). Also, as pointed out in Gelman et al. (2014), a willingness to share financial information with a third party might not only select on demographic characteristics, but also for an increased need for financial management or a higher degree of financial sophistication. However, because, as discussed in the introduction, the aim of this paper is to assess the effect of MDB on people who choose to use it, the lack of representativeness is not an issue.[6] A second limitation is that while

[5]For an example where debt-reduction in one set of accounts is offset by an increase in debt by another set of account, see Medina (2021).

[6]For an example of how re-weighing can be used to mitigate the non-representative issue, see Bourquin

Figure A.2: Pre and post-signup data availability




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.

we can observe user's complete financial behaviour if they add all their financial accounts to the app, it is not trivial to distinguish between users who do and do not do that. I address this challenge in the sample selection process documented below. A third issue is that while the app is able to classify many transactions into types, it misclassifies some transactions and cannot classify others altogether. I address this as part of the cleaning process documented below.

### A.2.2 Preprocessing

**Data cleaning:** I use the dataset described above for a number of projects, and perform a number of steps to create a minimally cleaned version of the dataset that is the basis for all such projects. These steps are performed in a dedicated data repository and

et al. (2020).

not run as part of this project.[7] Here, I briefly describe the main cleaning steps and their rationale. I drop all transactions with a missing description string because these cannot be categoriesed, and all transactions that are not automatically categoriesed by the app. Dropping these transactions makes it likely that I will underestimate amounts spent and saved, but minimises the risk of incorrectly classified transactions. I also group transactions into transfer, spend, and income subgroups, following Muggleton et al. (2020) to define spend subgroups and Hacıoğlu-Hoke et al. (2021) to define income subgroups.[8] Finally, I classify as duplicates and drop transactions with identical user ID, account ID, date, amount, and transaction description. This will drop some genuine transactions, such as when a user buys 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.

To minimise the influence of outliers, I winsorise all variables at the 1 percent level or – if we winsorise on both ends of the distribution – at the 0.5 percent level.[9] I rely on winsorisation (replacing top values with percentile values) instead of trimming (replacing top values with missing values) because data inspection suggests that in most cases, very large (absolute) values are not the result of data errors, which would call for trimming, but reflect genuine outcomes, which makes winsorising appropriate because it leaves these observations in the data while lowering their leverage to influence results.

**Sample selection:** The three main goals of sample selection are to select a sample of users for whom I can be reasonably certain to observe all relevant financial accounts,[10] account histories of at least 12 months, and who are not using MDB for business purposes. Table A.1 lists the precise conditions I apply to implement these criteria and their effect on sample size.[11]

In my main analysis, I show effects of app use for the 12-months period from 6 months before and 5 months after MDB signup, treating the month of signup as period 0. To ensure that results are not affected by the number of accounts we observe for an individual, it is thus critical that we can observe at least 12 months of history for all accounts a user adds to the platform. This ensures that in the extreme case where a user adds an account they had used for some time in the fifth month after they signed up, I observe all transactions on that account throughout the 12-month period of interest.[12] Data

[7]The module with all cleaning functions is available on Github.

[8]The precise list used to classify transactions is available on Github.

[9]The code that performs the winsorisation are available on Github.

[10]Relevant accounts include all current, savings, and credit-card accounts, but exclude long-term savings and investment accounts.

[11]The code that implements the selection criteria is available on Github.

[12]For example: if a user has made monthly payments of £100 into a savings account for two years by the time they sign up to MDB but links that account five months after joining, we can only observe the historical payments if MDB can access at least 12 months of history. If this is not the case, and MDB can only, say, access 6 months of history, we would erroneously conclude that the user started saving £100 more starting in the month they signed up.

Table A.1: Sample selection

| | Users | User-months | Txns | Txns (m£) |
|---|---|---|---|---|
| Raw sample | 271,856 | 7,948,520 | 662,112,975 | 124,573 |
| Drop test users | 270,782 | 7,878,398 | 656,047,534 | 122,887 |
| App signup after March 2017 | 88,368 | 2,320,421 | 202,580,838 | 38,816 |
| At least one savings account | 50,226 | 1,334,328 | 125,841,337 | 26,645 |
| At least one current account | 48,794 | 1,303,164 | 123,468,715 | 26,263 |
| At least £5,000 of annual income | 20,647 | 541,746 | 55,857,451 | 11,760 |
| At least 10 txns each month | 14,229 | 369,944 | 40,662,904 | 8,529 |
| At least £200 of monthly spend | 10,438 | 272,228 | 31,529,498 | 6,837 |
| No more than 10 active accounts | 9,788 | 248,975 | 27,589,696 | 5,426 |
| Complete demographic information | 7,720 | 202,633 | 22,671,753 | 4,378 |
| Working age | 7,568 | 197,951 | 22,279,279 | 4,213 |
| Final sample | 7,568 | 197,951 | 22,279,279 | 4,213 |

Notes: Number of users, user-months, transactions, and transaction volume in millions of British Pounds left in our sample after each sample selection step.

exploration suggests that all major banks start providing 12 months of historical data from April 2017 onwards, which is why include only users who sign up after that data.[13]

To ensure that I can be reasonably certain to observe users have added all their financial accounts to the app, I 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, I drop users with more than 10 active accounts in any given month.

Finally, I drop test users and users that are not of working age (younger than 18 or older than 65) because these groups might have objectives other than to reduce spending and increase savings.[14] And I retain only users for whom we can observe the complete set of demographic covariates (gender, age, region) to ensure all users can be used throughout the analysis. These steps do impact the sample size significantly.

### A.2.3 Summary statistics

Figure A.3 provides a view of some salient sample characteristics. Panels A and B show that while distributions of disposable income and total spend in 2019 broadly mirror that of the ONS Living Cost and Food Survey (LCFS) data, the MDB data tends to slightly underestimate incomes and overestimate spending.[15] Given that our sample is likely biased towards high-earners, as discussed above, we would expect both income and spend to

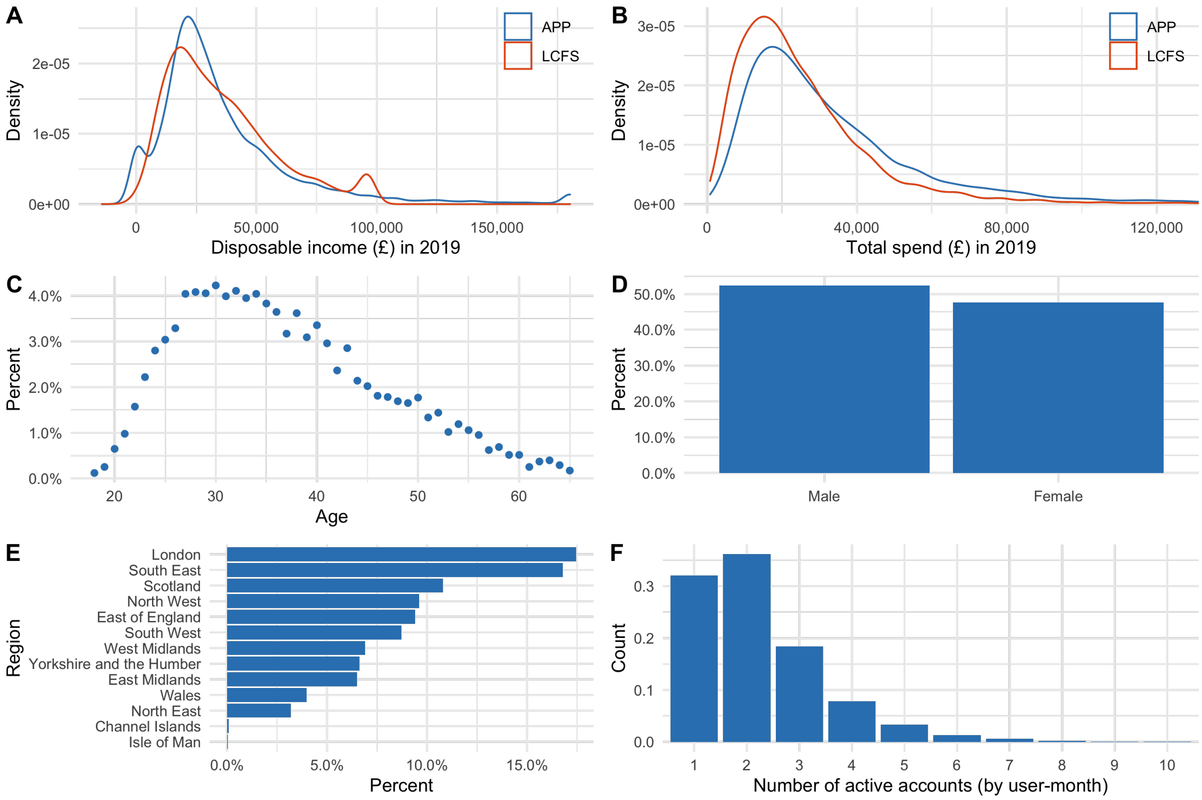
[13]The Jupyter Notebook containing the data exporation is available on Github.

[14]We cannot identify test users precisely, but drop users who signed up prior to or during 2011, the first year the app was in operation.

[15]I accessed the LCFS data via the UK Data Service at the following url: https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8686.

be higher than in the LCFS. Two likely caveats to the data probably create discrepancies relative to the LCFS data. First, as discussed in Section A.2.2, I drop unclassified transactions from the data, which biases both income and spending downwards (i.e. spending would probably be even higher compared to the LCFS). Second, it is likely that it is more challenging to automatically classify income transactions than spend transactions, which might create an additional downward bias on the income distribution.[16] Panels C, D, and E show that, as discussed in Section A.2.1, the sample is skewed towards users that are younger, male, and live in the South East. Finally, Panel D shows that most users use transact with one or two accounts per month.

Figure A.3: 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 A.2 provides additional summary statistics. It shows, for instance, that the average number of monthly transactions is slightly above 100 (with a mean of 112 and a median of 101), that half of all user-month incomes lie between £1,407 and £3,596, and that inflows into savings accounts closely mirror outflows. It also suggests that highly discretionary spend accounts for about 30 percent of total monthly spend

[16] Bourquin et al. (2020) present an alternative algorithm to identify income transactions and find that this leads to higher estimated incomes. I do not use their algorithm because I only use disposable income as a covariate to capture relative income differences between users.

Table A.2: Summary statistics

| Statistic | Mean | St. Dev. | Min | Pctl(25) | Median | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| Txn count | 111.9 | 59.7 | 10 | 70 | 101 | 142 | 327 |
| Month income | 2,853.5 | 2,495.2 | 0.0 | 1,407.2 | 2,217.4 | 3,595.9 | 15,027.5 |
| Savings account inflows | 747.2 | 2,448.3 | 0.0 | 0.0 | 0.0 | 400.0 | 18,809.5 |
| Savings account outflows | 762.4 | 2,422.9 | 0.0 | 0.0 | 0.0 | 400.0 | 18,099.4 |
| Savings account netflows | −7.4 | 2,883.7 | −20,000.0 | 0.0 | 0.0 | 50.0 | 21,675.4 |
| Month spend | 2,760.4 | 2,609.9 | 200.0 | 1,225.4 | 2,016.7 | 3,318.9 | 17,092.2 |
| Age | 37.5 | 10.0 | 18 | 30 | 36 | 44 | 65 |
| Female dummy | 0.4 | 0.5 | 0 | 0 | 0 | 1 | 1 |
| Urban dummy | 0.8 | 0.4 | 0 | 1 | 1 | 1 | 1 |
| Discretionary spend | 3.1 | 1.7 | 1 | 2 | 3 | 4 | 10 |
| Active accounts | 860.6 | 736.1 | 0.0 | 369.3 | 663.0 | 1,118.3 | 4,181.7 |

## A.3 Estimation

We want to estimate the effect of app use over time. Given our data, a natural way to do this would be to use a dynamic two-way fixed effects model that includes user and year-month fixed effects and dummies indicating time since app signup. The estimated coefficients on these dummies are then conventionally interpreted as dynamic treatment effects. However, a series of papers in econometric research demonstrate that while this approach is extremely common in applied research, two-way fixed effects models do not produce valid estimates of dynamic treatment effects in most settings (Roth et al. 2022). In particular, in settings with staggered treatment assignment, where units are first exposed to treatment at different points in time (as is the case in our setting), dynamic two-way fixed effects are valid only if there is homogeneity in treatment effects across treatment adoption cohorts. In most settings, including my own, this is a very strong assumption.[17]

Because of this, I use a new estimator proposed by Callaway and Sant'Anna (2021), which allows for arbitrary treatment effect heterogeneity across treatment adoption cohorts and time, and allows for the incorporation of a parallel trends assumption conditional on covariates.[18] In describing the estimator, I follow the approach of Callaway and

[17]The problem is exacerbated in static two-way fixed effects models, which require homogeneity in treatment effects across treatment adoption cohorts and in time since treatment. Goodman-Bacon (2021) shows that the static two-way fixed effect DiD estimator is a weighted average of all possible two-units and two-time-periods difference-in-differences in which one unit changes its treatment status and the other does not. Because this includes "forbidden comparisons", where already treated units are used as control units, this can lead to a scenario where the treatment effect of some units has a negative weight, unless one makes the additional assumption of time-invariant treatment effects (in which case, the "forbidden" comparisons have an effect size of zero). These negative weights, in turn, can lead to situations where the treatment effect estimate is negative even though all unit-level treatment effects are positive. The dynamic specification suffers from very similar problems, as shown in Sun and Abraham (2021). For two excellent reviews of this new literature, see Roth et al. (2022) and A. C. Baker et al. (2022).

[18]I implement the estimator using the R package 'did'.

Sant'Anna (2021) of first defining the causal parameter of interest, and then discussing identification, estimation, and inference, while using the simplified notation used in Roth et al. (2022).

**Causal parameter of interest:** The basic building block of the framework, and the causal effect of interest, is the group-time average treatment effect: the average treatment effect at time $t$ for the group of individuals first treated at time $g$, defined as:

$$ATT(g,t) = \mathbb{E}[Y_{i,t}(g) - Y_{i,t}(0)|G_i = g], \tag{A.1}$$

where $Y_{i,t}(g)$ is the potential outcome in time period $t$ of an individual $i$ in group $g$, and $Y_{i,t}(0)$ is the (counterfactual) potential outcome of that same individual if they had remained untreated.

**Identification:** These effects are identified if two main assumptions hold: if there is limited and known anticipation of treatment, and if the assumption of parallel trends between treatment and comparison groups holds either unconditionally or conditionally on a set of covariates.[19] Because the purpose of this section is to convey the core idea of the estimation approach, I keep things as simple as possible and discuss only the case with no anticipation effects and where the parallel trend assumption holds unconditionally. Callaway and Sant'Anna (2021) show that the same overall approach also works when allowing for known anticipation and conditional parallel trends.[20]

Given these two assumptions, Callaway and Sant'Anna (2021) show that $ATT(g,t)$ is identified by comparing the expected change of group $g$ between periods $t$ and $g-1$ with that of a comparison group that is not yet treated at time $t$:

[19] Additional assumptions are (i) that the treatment is absorbing in the sense that once an individual is treated they will remain treated forever, (ii) that individuals in the data are randomly and independently drawn from a larger population, and (iii) an overlap condition that ensures that there is a positive number of users that is first exposed to the treatment at any period and that – under the conditional parallel trends assumption – propensity scores for initial treatment times based on covariates are bounded away from zero. The first assumption could be violated only if a user closes their account on the app and then signed up again later on, all within the roughly within the two-year data periods I use. This cannot be more than a tiny minority of users. The second assumption holds less trivially. One way to think of a super population from which the users in the dataset are drawn is to think of knowledge about the app as partially random, and about the super population of all individuals who would have signed up had they learned about the apps existence.

[20] Intuitively, known anticipation merely shifts the reference period from the period immediately before treatment to the period before anticipation of treatment begins. When relying on the conditional parallel trends assumption, the overall the group-time average treatment effect $ATT(g,t)$ is the average of unconditional group-time average treatment effects for each value of the covariate vector $X_i$. As discussed in Roth et al. (2022), estimation is challenging when $X_i$ is continuous or can take on a large number of values, since then we will typically lack data to estimate unconditional group-time treatment effects for each value of $X_i$. There are different semi- and non-parametric approaches that can be used in such cases. I use the first step regression estimator throughout for all results shown in this paper. See Callaway and Sant'Anna (2021) and Roth et al. (2022) for more details.

$$ATT(g,t) = \mathbb{E}[Y_{i,t} - Y_{i,g-1}|G_i = g] - \mathbb{E}[Y_{i,t} - Y_{i,g-1}|G_i = g'], \text{ for any } g' > \text{t}. \quad \text{(A.2)}$$

As this holds for all $g'$ that are not yet treated at time $t$, it also holds for an average over all such groups, collected in set $\mathcal{G}$, so that

$$ATT(g,t) = \mathbb{E}[Y_{i,t} - Y_{i,g-1}|G_i = g] - \mathbb{E}[Y_{i,t} - Y_{i,g-1}|G_i \in \mathcal{G}]. \quad \text{(A.3)}$$

This equation encapsulates two main results: that the period just before treatment, $g-1$, is a valid reference period, and that the group of all individuals that have not yet been treated at time $t$ are a valid comparison group for estimating treatment effects in time $t$.[21] As a result of this, units who are treated in the very first period in the dataset are dropped from the sample, since there exists no possible control group based on which to identify their treatment effect, and since they are not useful as a control group themselves. Similarly, unit treated in the very last period in the data are also dropped, since there exists no “not-yet-treated“ group that could serve as a comparison group for them.

**Estimation:** $ATT(g,t)$ can be estimated by replacing expectations with their sample analogues,

$$\widehat{ATT(g,t)} = \frac{1}{N_g} \sum_{i:G_i=g} [Y_{i,t} - Y_{i,g-1}] - \frac{1}{N_{\mathcal{G}}} \sum_{i:G_i \in \mathcal{G}} [Y_{i,t} - Y_{i,g-1}]. \quad \text{(A.4)}$$

Once these building blocks are estimated, aggregating them to event-study type treatment effects that provide the (weighted) average treatment effect $l$ periods away from treatment adoption across different adoption groups can be achieved by calculating

$$ATT_l = \sum_g w_g ATT(g, g+l), \quad \text{(A.5)}$$

where $w_g = P(G = g|g \in \mathcal{G})$ are the groups relative frequencies in the treated population. When calculating these event study parameters, I use a panel balanced in event times, with all units being observed for at least 5 treatment periods. This avoids the $ATT_l$ being influenced by different group compositions at different periods $l$.[22]

**Inference:** Inference is based on a bootstrap proceedure that can account for clustering and produces simultaneous (or uniform) confidence bands that account for multiple

[21] If a group of never-treated individuals is available, then these could also serve as a comparison group. But because my dataset does not contain such individuals, I do not discuss this case.

[22] See Section 3.1.1 in Callaway and Sant'Anna (2021) for a more detailed discussion.

hypothesis testing.[23]

## A.4 Results

The main outcome variables in my analysis are "discretionary spend", defined as spend transactions over which users likely have a lot of control, and "net-inflows into savings-accounts", defined as the difference between inflows into and outflows from all of a user's savings accounts.[24]

### A.4.1 Main results

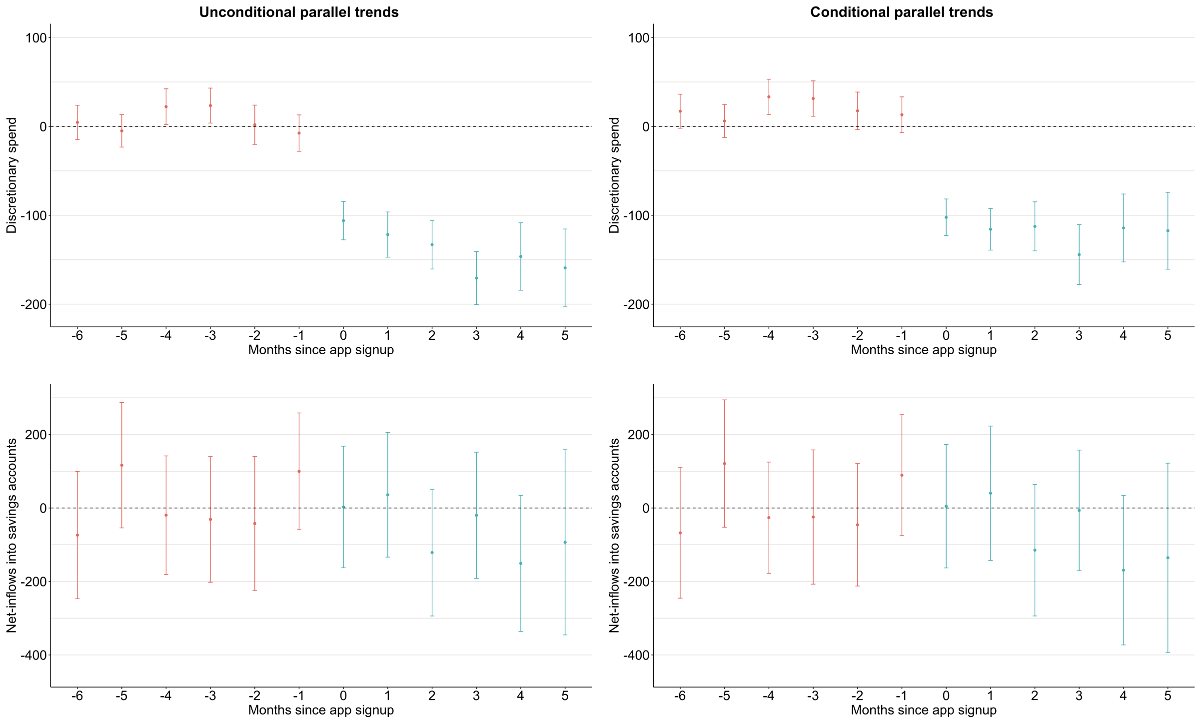
Figure A.4 shows the effect of app use on monthly discretionary spend (top row) and monthly net-inflows into savings accounts (bottom row) under the unconditional (left column) and conditional (right column) parallel trends assumptions. As discussed in Section A.3, the conditional parallel trends assumption states that the outcome variable of treatment and control units with the same set of covariates would have evolved in parallel fashion. Throughout my analysis, the set of covariates I use are month income, month spend, the number of active accounts, and age. Estimates are group-time average treatment effects aggregated by time since treatment exposure, as defined in Equation A.5. All results are presented with a uniform 95% confidence band, based on bootstrapped standard errors that account for autocorrelation in the data and are clustered at the user level, as discussed in Section A.3.

We can see that discretionary spend falls by between £100 and £150 per month once users start using the app, depending on the parallel trends assumption used. Given that average monthly discretionary spend is about £860 (see Table A.2), this corresponds to a drop in discretionary spend of about 11-17 percent, which is substantial. These results are in line with those found in Levi and Benartzi (2020), which find a 11.6 percent reduction in discretionary spend following the use of an aggregator app. Interestingly, we can also see that at least for the first six months of app use, the effect is persistent.

[23] For more details on the inference proceedure, see Section 4.1 in Callaway and Sant'Anna (2021). The interpretation of uniform and pointwise confidence intervals differ in the following way: a uniform 95% confidence band accounts for multiple hypothesis testing in that it is constructed such that *all* shown coefficients cover their corresponding true value 95 percent of the time. In contrast, a more commonly used pointwise 95% confidence band is constructed such that the confidence interval for each parameter covers the true parameter 95 percent of the time.

[24] Discretionary spend is calculated solely based on card transactioins, while cash transactions are excluded. This is because cash transactions are recorded in the MDB data as ATM withdrawals, and it is impossible to know what these funds were used for. Furthermore, I do not include gasoline expenses in discretionary spend because many people use their car to commute to work so that it is unclear what proportion of these expenses are discretionary. A list of transaction tags used to identify discretionary spend transactions and the code used create the variable are available on Github. To capture only user-generated savings-account flows and not interest payments and similar automated transactions, I include only transaction with an absolute value of £5 or higher. The code used to calculate savings account flows is also available on Github.

Figure A.4: Main results




Notes: The effect of app use on monthly discretionary spending (top row) and monthly net-inflows into savings accounts (bottom row) under the unconditional (left column) and conditional (right column) parallel trends assumption. Point estimates represent group-time average treatment effects aggregated to periods since treatment exposure, as defined in Section A.3. Red lines represent point estimates and uniform 95% confidence bands for pre-treatment periods allowing for clustering at the user level. If the null hypothesis that parallel trends hold in all periods is correct, these should be equal to zero. Blue lines provide similar information for post-treatment periods.

That the results based on the unconditional and the conditional parallel trends assumptions are very similar but not identical is not surprising. 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 treatment and control groups. A classic example of the latter would be comparing individuals wo did and did not sign up for a job-training program, where characteristics such as age and employment history are often quite different (Heckman et al. 1997). In our context, however, where all individuals eventually sign up to Money Dashboard and our comparison group is a set of "not-yet-treated" rather than "never-treated" individuals, we would not expect covariate values between individuals in the treatment and control groups to vary as much. At the same time, we would expect the results to differ somewhat. If, as is plausible, discretionary spend is a constant fraction of income and total spend, then we would expect parallel trends only for individuals with similar incomes and total spend. Similarly, if we think that discretionary spend increases in the number of accounts we observe per user, then parallel trends in spend only holds for users for whom we observe the same number of accounts.[25]

[25] As discussed in Section A.2, users choose which accounts to link to Money Dashboard, and while I select for users that appear to have linked all their accounts and for whom we could observe the complete

We can also see that in two periods, the null-hypothesis of identical pre-signup trends is being rejected (both under conditional and unconditional trends). The estimated differences from zero are not large, but a sign that the results should be interpreted with caution. In future work, I plan to use the approach introduced by Rambachan and Roth (2022) to test how sensitive my results are to deviations from parallel trends.

Finally, we can see that, in contrast to discretionary spending, net-inflows into savings accounts do not change once users start using the app. This is somewhat surprising, since we might have expected that users transfer at least some of their saved funds into their savings accounts to either build up a "rainy-day" savings cushion or as a contribution towards specific savings goals. The absence of such transfers naturally begs the question "where does the money go?", to which I turn in the next section.

### A.4.2 Where do additional funds go?

Above, we found that while users reduce their discretionary spend after using MDB by between £100 and £150, they do not transfer additional funds into their savings accounts.

One possible explanation is that users do save more but transfer money either into longer-term savings vehicles such as investment funds or into savings accounts they have not linked to MDB. The first three panels (from the top left) in Figure A.5 show flows into investment of pension funds, current account outflows that users manually labelled as "savings", and – as an alternative but cruder measure of transfers into savings accounts – the total amount of transfers from current accounts into other accounts held by the user.[26] None of these change significantly during users' pre and post signup periods, rejecting the idea that users systematically move additional funds into either investment vehicles or other savings accounts. An exception is the spikes in user-labelled savings in the month just before and the month of signup. Rather than suggesting an additional inflow of funds, however, this is is more likely to be an artifact of users only manully labelling transactions during active app use rather than also labelling historical transctions.[27]

Another possibility is that users use the additional funds to pay down debt. The bottom right panel in Figure A.5 shows loan repayments, which do not change significantly during the 12-month period around app signup either.

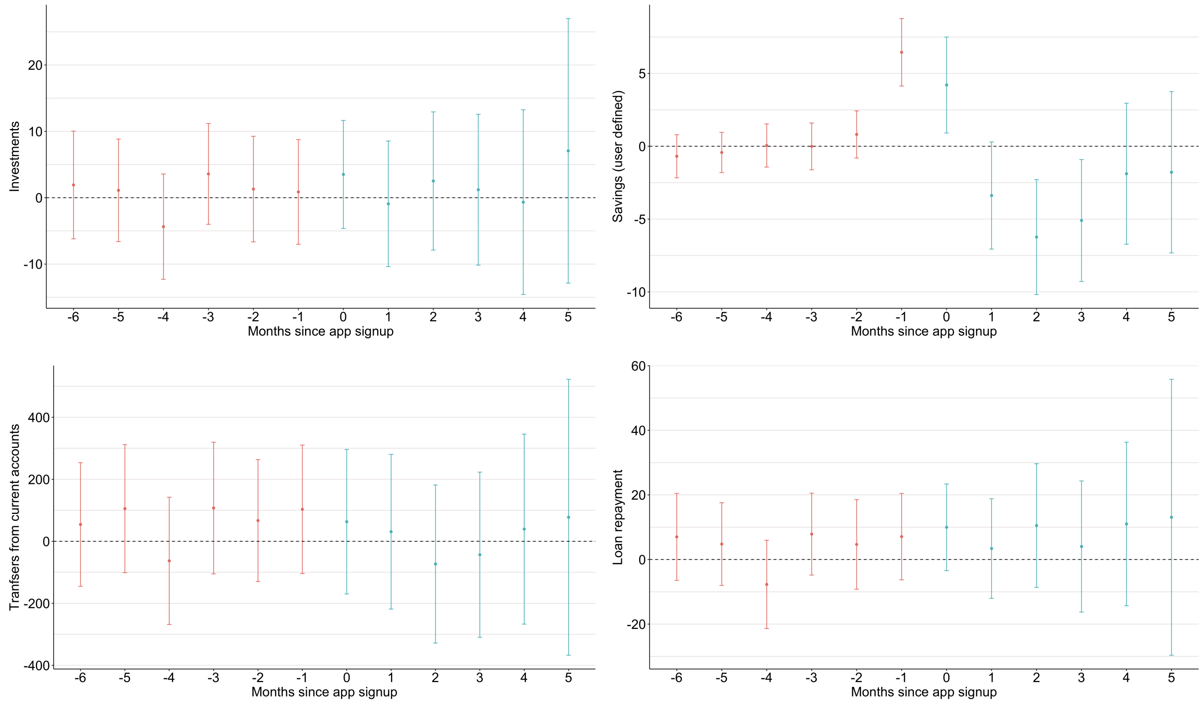
The finding that the saved funds do not flow into any of the plausible channels suggests that they might be channelled towards multiple uses either across or between individuals.

account history even if they added some of their accounts after joining, I cannot rule out that some users do add account after signup and MDB is unable to capture the complete history.

[26]The code where these variables are defined is available on Github.

[27]The pattern could be explained in two ways: either users can see the entire account historey available to MDB after linking an account but only label transactions during periods of active app use and the period immediately before signup, or MDB only shows users data from the month before signup onwards, so that users are unable to manually label transactions in earlier periods.

Figure A.5: Possible destinations of additional savings



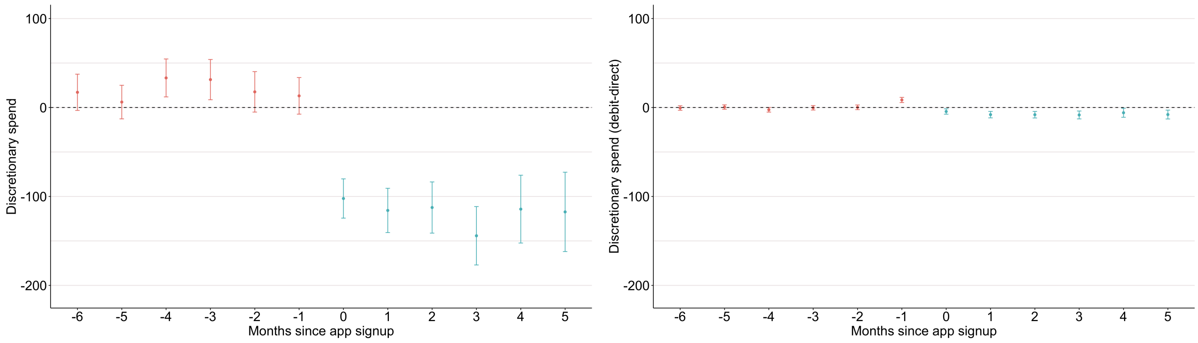

Notes: Point estimates represent group-time average treatment effects aggregated to periods since treatment exposure, as defined in Section A.3. Red lines represent point estimates and uniform 95% confidence bands for pre-treatment periods allowing for clustering at the user level. If the null hypothesis that parallel trends hold in all periods is correct, these should be equal to zero. Blue lines provide similar information for post-treatment periods.

### A.4.3 Disaggregated discretionary spend

Another question of interest based on the main findings above is along what dimensions users reduce their discretionary spend. There are three dimensions of interest.

First, users can achieve the persistent reduction in discretionary spend seen in Figure A.4 by either maintaining a change in behaviour consistently month-to-month by, for instance, making fewer purchases or spending less on certain items, or, alternatively, make a decision during the month of signup that automates the reductin in spend by cancelling debit-direct contracts. Figure A.6 shows in the left panel the reduction of discretionary spending shows in Figure A.4 above for reference, and in the right panel discretionary spend paid via debit-direct contracts. It shows that debit-direct discretionary spend falls only marginally after app use, indicating that it is a change in behaviour sustained month-to-month that drives the reduction in discretionary spend. A caveat to this analysis is that the MDB data allows for only imperfect identification of debit-direct transactions for two reasons: first, only some banks label debit-direct transactions as such in the transaction descriptions provided to MDB and it is commonly applied labels that I use to identify the transactions. Second, MDB redacts some transaction descriptions entirely or in part to ensure the anonymity of its users, and its possible that some of these transactions are

Figure A.6: Change in direct-debit discretionary spend




Notes: Effect of app use on total discretionary spend (for reference) and discretionary spend paid via debit-direct contracts. Point estimates represent group-time average treatment effects aggregated to periods since treatment exposure, as defined in Section A.3. Red lines represent point estimates and uniform 95% confidence bands for pre-treatment periods allowing for clustering at the user level. If the null hypothesis that parallel trends hold in all periods is correct, these should be equal to zero. Blue lines provide similar information for post-treatment periods.

debit-directs.[28]

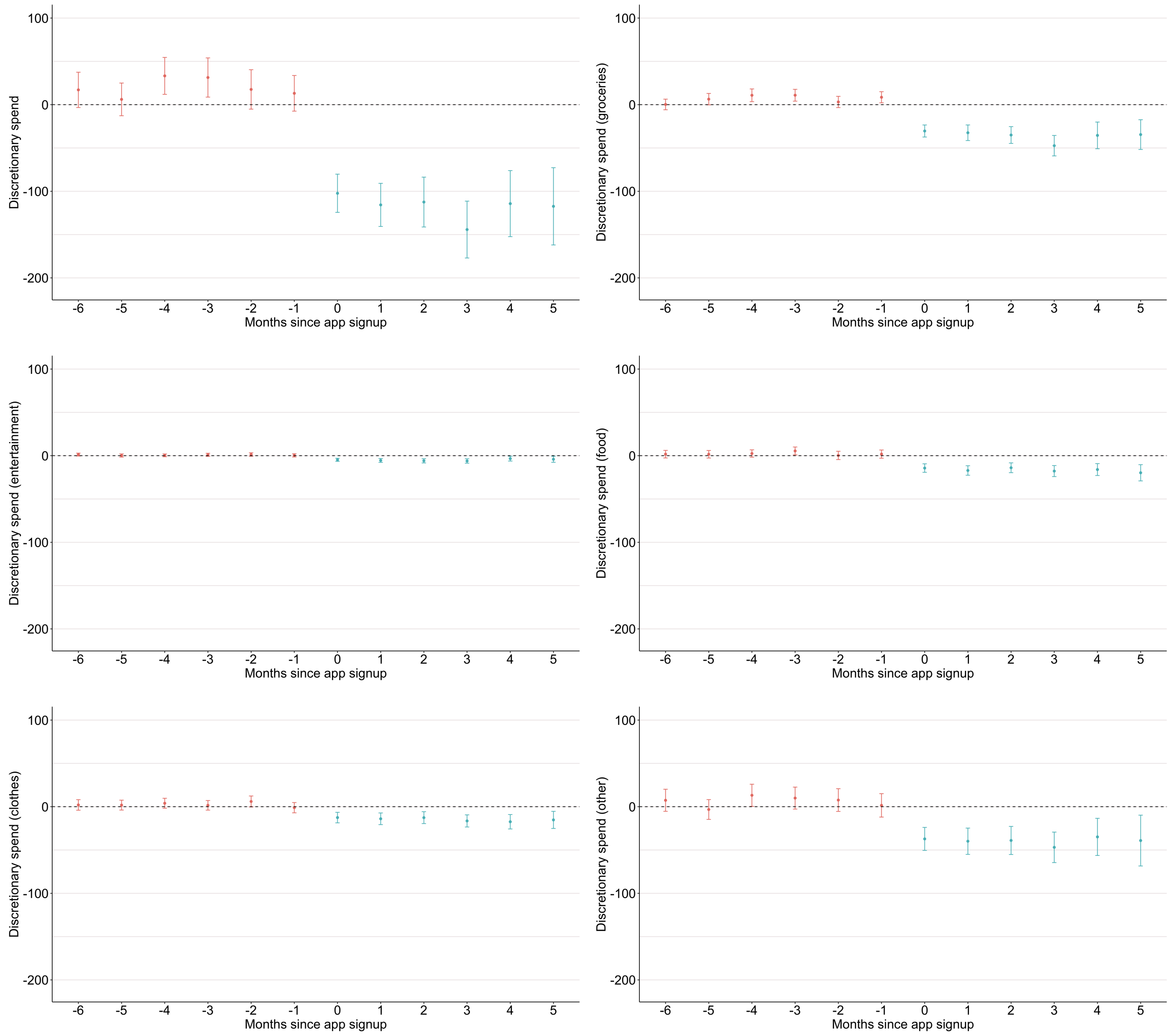
The second interesting way to disaggregate discretionary spend is into different component groups based on the type of spend; for this purpose, I group transactions into "groceries", "clothes", "entertainment", "food", which captures spending on restaurant meals and take-away, and "other".[29] Figure A.7 again reproduces the plot from the main results in the top-left panel for reference and then plots results for each of the five subgroups. While we can see that users mainly reduce spending on groceries and the residual category "other", followed by food and clothes, the overall picture that emergest is that these differences are quite small, and that the overall reduction in discretionary spend results from changes along all these margins.

Finally, it is interesting to consider whether users reduce discretionary spend along the extensive or the intensive margin – whether they make fewer transactions or reduce the average spend per transaction. Figure A.8 shows the number of transactions (the extensive margin) in the left panel and the average spend per transaction (the intensive margin) in the right panel. We can see that the reduction in spend is clearly the result of changes along the extensive margin – on average, users make about five fewer discretionary spend purchases per month once they sign up to MDB. The average transaction value of a discretionary spend purchase in the data is about £25, adding up to the total effect we find in Figure A.4.

[28]The code used to calculate debit-direct discretionary spend is available on Github.

[29]The list used to classify transactions is available on Github. I classify transactions tagged by MDB as "entertainment" as "other discretionary spend", since many such transactions are purchses with Amazon, which we cannot precisely classify.

Figure A.7: Change in discretionary spend by subgroup



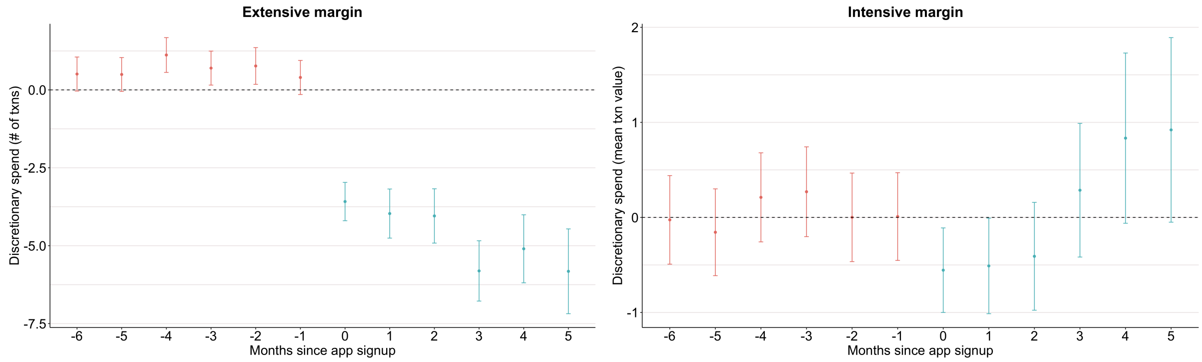

Notes: Effect of app use on total discretionary spend (for reference) and its subgroups. Point estimates represent group-time average treatment effects aggregated to periods since treatment exposure, as defined in Section A.3. Red lines represent point estimates and uniform 95% confidence bands for pre-treatment periods allowing for clustering at the user level. If the null hypothesis that parallel trends hold in all periods is correct, these should be equal to zero. Blue lines provide similar information for post-treatment periods.

## A.5 Conclusion

In this paper, I test whether using Money Dashboard is associated with a reduction in discretionary spending and an increase in "rainy-day savings". I use a new estimator proposed by Callaway and Sant'Anna (2021) that corrects for recently identified problems in two-way fixed effects estimates.

I find that users reduce their discretionary spend by between £100 and £150 (11-17% of average discretionary spend) once they start using the app and sustain that reduction throughout the six-month post-signup period I consider. Looking at disaggregated measures of discretionary spend further shows that the reduction is the result of maintained month-to-month changes in behaviour rather than one-off cancellations of direct-debit

Figure A.8: Intensive and extensive margin




Notes: The effect of app use on monthly discretionary spend disaggregated into the effect on the extensive (left column) and intensive (right column) margins. The extensive margin is the number of discretionary transactions per month, and the intensive margin is the average value of a discretionary spend transaction. Point estimates represent group-time average treatment effects aggregated to periods since treatment exposure, as defined in Section A.3. Red lines represent point estimates and uniform 95% confidence bands for pre-treatment periods allowing for clustering at the user level. If the null hypothesis that parallel trends hold in all periods is correct, these should be equal to zero. Blue lines provide similar information for post-treatment periods.

transactions, that it results from reducing spending on a number of different categories of purchases rather than a single one, and that it is a result of changes along the extensive rather than the intensive margin – users reduce the number of transactions they make rather than the value of the average transaction. Interestingly, users do not seem to use these additional funds to build up "rainy-day savings": net-inflows into savings accounts do not change after signup. I can also neither find significant increases in flows into investment and pension accounts or additional savings accounts that are not linked to the app, not additional loan repayments.

The main limitation of my approach is that I cannot isolate the effect of MDB use from possible confounding factors that let users to self-select into using the app in the first place, and that I cannot isolate the effect of individual components of the app such as information and goal-setting. However, I do find that app use is associated by a statistically and economically significantly reducuction of discretionary spend that is sustained for at least six months, and that seems to be brought about by a reduction in the number of purchases across a number of spending categories. This provides novel insights into how people that probably do have some unobserved motivation to reduce their savings actually achieve this. The additional finding that a reduction in spend is not accompanied by a commensurate increase in either short-term savings, investments, or debt-reduction suggests that people use the saved funds for a variety of different purposes.

The findings also suggest promising directions for further research. First, the substantial drop in discretionary spend associated with app use does suggest that MDB and apps like it might have a positive causal impact on financial outcomes, making it worthwhile to study their effect in more detail in ways that account for the two limitations mentioned above. Second, the finding that people may direct saved funds towards a number of different purposes can be the result of differences in funds use across or within people or a combination of the two. Within people variation in fund use could suggest the possibility that – similarly to allocating credit-card payments towards different cards proportional to outstanding debt rather than towards highest-interest cards Gathergood et al. (2019) – people direct additional cash-flows towards a number of different uses instead of towards those with the highest payoff, such as paying down credit-card debt or investing in government-subsidised high-interest investments.[30] Further research establishing to what extent there is within-person variation of fund use, and whether such funds are directed towards the most effective uses would provide further useful insights into savings behaviour, and could serve as the basis for further development of financial aggregator apps to help users direct saved funds towards the most effective users.

# .1 Additional results

## .1.1 Unbalanced aggregation

The baseline specification relies on a panel balanced in event time and thus only includes users that have been exposed to treatment for at least 5 periods. Figure 9 reproduces the baseline results in the left panel and compares it with results based on the full sample.

As discussed in section 3.1.1 of Callaway and Sant'Anna (2021), 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 $l$ periods after treatment (which is what we want) as well as compositional changes due to different groups being included for different periods $l$ 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.

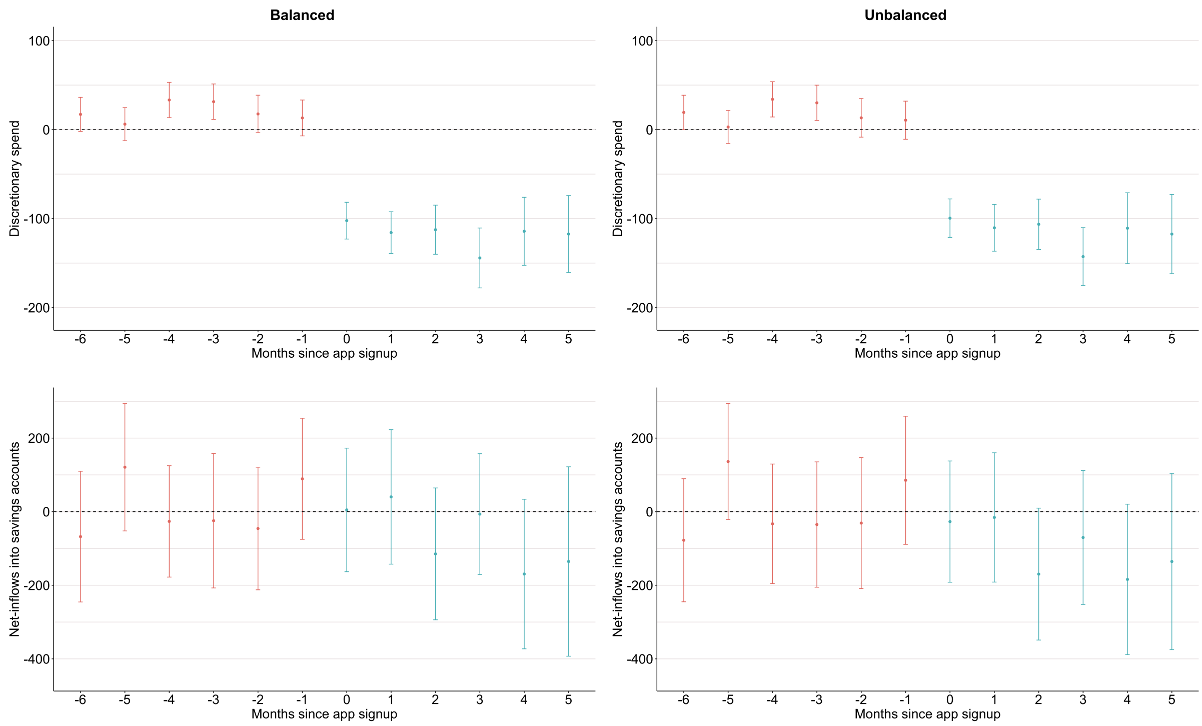
As expected, using the full data reduces the size of the confidence intervals. But the results are otherwise very similar.

## .1.2 Inflows and outflows

The leftmost panel in Figure 10 reproduces the results from Figure A.4 for net-inflows into savings accounts using the conditional parallel trends assumption for reference, and decomposes these net-inflows into inflows (middle) and outflows (right). We can see that net-inflows remain unchanged because both inflows and outflows remain unchanged and not because a change in inflows if offset by a commensurate change in outflows.

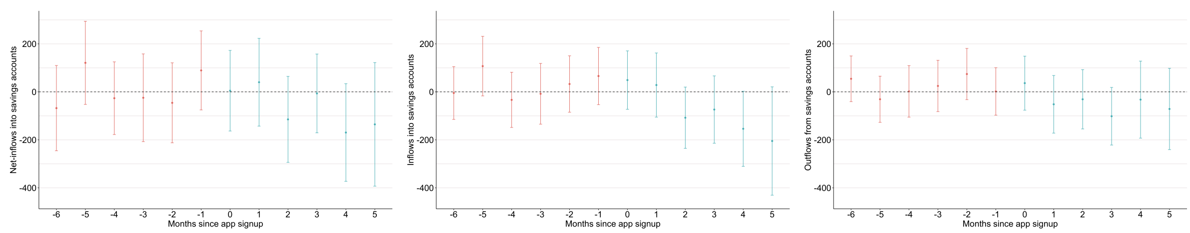
[30]In the UK, for instance, the first £4000 paid into a Lifetime ISA each year are matched by the government by payment of £1000, corresponding to a risk-free interest rate of 25 percent.

Figure 9: Balanced and unbalanced aggregation




Notes: Main results based on balanced and unbalanced panels. Point estimates represent group-time average treatment effects aggregated to periods since treatment exposure, as defined in Section A.3. Red lines represent point estimates and uniform 95% confidence bands for pre-treatment periods allowing for clustering at the user level. If the null hypothesis that parallel trends hold in all periods is correct, these should be equal to zero. Blue lines provide similar information for post-treatment periods.

Figure 10: Decomposing net-inflows into inflows and outflows




Notes: Decomposition of net-inflows into savings accounts (left) into inflows (middle) and outflows (right). Point estimates represent group-time average treatment effects aggregated to periods since treatment exposure, as defined in Section A.3. Red lines represent point estimates and uniform 95% confidence bands for pre-treatment periods allowing for clustering at the user level. If the null hypothesis that parallel trends hold in all periods is correct, these should be equal to zero. Blue lines provide similar information for post-treatment periods.

# Bibliography


Ainslie, George (1975). “Specious reward: a behavioral theory of impulsiveness and impulse control.” In: *Psychological bulletin* 82.4, p. 463.

— (2001). *Breakdown of will.* Cambridge University Press.

Allcott, Hunt and Nathan Wozny (2014). “Gasoline prices, fuel economy, and the energy paradox”. In: *Review of Economics and Statistics* 96.5, pp. 779–795.

Ameriks, John, Andrew Caplin, John Leahy, and Tom Tyler (2007). “Measuring self-control problems”. In: *American Economic Review* 97.3, pp. 966–972.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger (2015). “Working over time: Dynamic inconsistency in real effort tasks”. In: *The Quarterly Journal of Economics* 130.3, pp. 1067–1115.

Baker, Andrew C, David F Larcker, and Charles CY Wang (2022). “How much should we trust staggered difference-in-differences estimates?” In: *Journal of Financial Economics* 144.2, pp. 370–395.

Baker, Scott R (2018). “Debt and the response to household income shocks: Validation and application of linked financial account data”. In: *Journal of Political Economy* 126.4, pp. 1504–1557.

Baker, Scott R and Lorenz Kueng (2022). “Household financial transaction data”. In: *Annual Review of Economics* 14, pp. 47–67.

Baugh, Brian, Itzhak Ben-David, and Hoonsuk Park (2014). “Disentangling financial constraints, precautionary savings, and myopia: household behavior surrounding federal tax returns”. Tech. rep. National Bureau of Economic Research.

Benhabib, Jess and Alberto Bisin (2005). “Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption–saving decisions”. In: *Games and economic Behavior* 52.2, pp. 460–492.

Berk, Sarah Holmes, John Beshears, James J Choi, and David Laibson (2022). “Automating Short-Term Payroll Savings: Initial Evidence from a Large UK Experiment”. In:

Berns, Gregory S, David Laibson, and George Loewenstein (2007). “Intertemporal choice–toward an integrative framework”. In: *Trends in cognitive sciences* 11.11, pp. 482–488.

Beshears, John, James J Choi, J Mark Iwry, David C John, David Laibson, and Brigitte C Madrian (2020). “Building emergency savings through employer-sponsored rainy-day savings accounts”. In: *Tax Policy and the Economy* 34.1, pp. 43–90.

Beshears, John, Katherine L Milkman, and Joshua Schwartzstein (2016). "Beyond beta-delta: The emerging economics of personal plans". In: *American Economic Review* 106.5, pp. 430–34.

Biljanovska, Nina and Spyros Palligkinis (2018). "Control thyself: Self-control failure and household wealth". In: *Journal of Banking & Finance* 92, pp. 280–294.

Bourquin, Pascale, Isaac Delestre, Robert Joyce, Imran Rasul, and Tom Walters (2020). "The effects of coronavirus on household finances and financial distress". In: *IFS Briefing Note BN298*.

Brunnermeier, Markus K and Stefan Nagel (2008). "Do wealth fluctuations generate time-varying risk aversion? Micro-evidence on individuals". In: *American Economic Review* 98.3, pp. 713–36.

Busse, Meghan R, Christopher R Knittel, and Florian Zettelmeyer (2013). "Are consumers myopic? Evidence from new and used car purchases". In: *American Economic Review* 103.1, pp. 220–56.

Buyalskaya, Anastasia, Marcos Gallo, and Colin F Camerer (2021). "The golden age of social science". In: *Proceedings of the National Academy of Sciences* 118.5, e2002923118.

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

CAN, Commonwealth Bank of Australia (2019). "Improving the Financial Wellbeing of Australians". Tech. rep. URL: https://www.commbank.com.au/content/dam/caas/newsroom/docs/CWM0375_Financial%20Wellbeing%20Report_v4.pdf.

Carlin, Bruce, Arna Olafsson, and Michaela Pagel (2019). "Generational Differences in Managing Personal Finances". In: *AEA Papers and Proceedings*. Vol. 109, pp. 54–59.

— (2022). "Mobile Apps and Financial Decision Making". In: *Review of Finance*.

CFPB, Consumer Financial Protection Bureau (2017). "Financial Well-being in America". Tech. rep. URL: https://www.consumerfinance.gov/data-research/research-reports/financial-well-being-america/.

Chabris, Christopher F, David Laibson, Carrie L Morris, Jonathon P Schuldt, and Dmitry Taubinsky (2008). "Individual laboratory-measured discount rates predict field behavior". In: *Journal of risk and uncertainty* 37.2-3, p. 237.

Chater, Nick and George Loewenstein (2022). "The i-frame and the s-frame: How focusing on the individual-level solutions has led behavioral public policy astray". In: *Available at SSRN 4046264*.

Choi, James J, David Laibson, Brigitte C Madrian, and Andrew Metrick (2004). "For better or for worse: Default effects and 401 (k) savings behavior". In: *Perspectives on the Economics of Aging*. University of Chicago Press, pp. 81–126.

— (2009). "Reinforcement learning and savings behavior". In: *The Journal of finance* 64.6, pp. 2515–2534.

Choukhmane, Taha (2019). “Default options and retirement saving dynamics”. In: *Working Paper*.

Cohen, Jonathan, Keith Marzilli Ericson, David Laibson, and John Myles White (2020). “Measuring time preferences”. In: *Journal of Economic Literature* 58.2, pp. 299–347.

Colby, Helen and Gretchen B Chapman (2013). “Savings, subgoals, and reference points”. In:

D’Acunto, F, AG Rossi, and M Weber (2020). “Crowdsourcing peer information to change spending behavior”. In: *Chicago Booth Research Paper* 19-09.

DellaVigna, Stefano (2009). “Psychology and economics: Evidence from the field”. In: *Journal of Economic literature* 47.2, pp. 315–72.

DellaVigna, Stefano and Ulrike Malmendier (2006). “Paying not to go to the gym”. In: *American Economic Review* 96.3, pp. 694–719.

Eagle, Nathan, Michael Macy, and Rob Claxton (2010). “Network diversity and economic development”. In: *Science* 328.5981, pp. 1029–1031.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde (2018). “Global evidence on economic preferences”. In: *The Quarterly Journal of Economics* 133.4, pp. 1645–1692.

Frederick, Shane, George Loewenstein, and Ted O’Donoghue (2002). “Time discounting and time preference: A critical review”. In: *Journal of economic literature* 40.2, pp. 351–401.

Fudenberg, Drew and David K Levine (2006). “A dual-self model of impulse control”. In: *American economic review* 96.5, pp. 1449–1476.

Ganong, Peter and Pascal Noel (2019). “Consumer spending during unemployment: Positive and normative implications”. In: *American Economic Review* 109.7, pp. 2383–2424.

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

Gathergood, John (2012). “Self-control, financial literacy and consumer over-indebtedness”. In: *Journal of economic psychology* 33.3, pp. 590–602.

Gathergood, John, Neale Mahoney, Neil Stewart, and Jörg Weber (2019). “How do individuals repay their debt? The balance-matching heuristic”. In: *American Economic Review* 109.3, pp. 844–75.

Gelman, Michael, Shachar Kariv, Matthew D Shapiro, Dan Silverman, and Steven Tadelis (2014). “Harnessing naturally occurring data to measure the response of spending to income”. In: *Science* 345.6193, pp. 212–215.

Goodman-Bacon, Andrew (2021). “Difference-in-differences with variation in treatment timing”. In: *Journal of Econometrics* 225.2, pp. 254–277.

Governors of the Federal Reserve System, Board of (2022). “Economic Well-Being of U.S. Households in 2021”. Tech. rep.

Guidotti, Riccardo, Michele Coscia, Dino Pedreschi, and Diego Pennacchioli (2015). "Behavioral entropy and profitability in retail". In: *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, pp. 1–10.

Gul, Faruk and Wolfgang Pesendorfer (2001). "Temptation and self-control". In: *Econometrica* 69.6, pp. 1403–1435.

Hacioglu, Sinem, Diego Känzig, and Paolo Surico (2020). "The Distributional Impact of the Pandemic". In:

Hacıoğlu-Hoke, Sinem, Diego R Känzig, and Paolo Surico (2021). "The distributional impact of the pandemic". In: *European Economic Review* 134, p. 103680.

Harrison, Glenn W, Morten I Lau, and Melonie B Williams (2002). "Estimating individual discount rates in Denmark: A field experiment". In: *American economic review* 92.5, pp. 1606–1617.

Hausman, Jerry A (1979). "Individual discount rates and the purchase and utilization of energy-using durables". In: *The Bell Journal of Economics*, pp. 33–54.

Heckman, James J, Hidehiko Ichimura, and Petra E Todd (1997). "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme". In: *The review of economic studies* 64.4, pp. 605–654.

Hesketh, Beryl, Christine Watson-Brown, and Sonja Whiteley (1998). "Time-related discounting of value and decision-making about job options". In: *Journal of vocational behavior* 52.1, pp. 89–105.

Hirsh, Jacob B, Raymond A Mar, and Jordan B Peterson (2012). "Psychological entropy: a framework for understanding uncertainty-related anxiety." In: *Psychological review* 119.2, p. 304.

Hsiaw, Alice (2013). "Goal-setting and self-control". In: *Journal of Economic Theory* 148.2, pp. 601–626.

Hurst, Erik (2004). "Grasshoppers, ants and pre-retirement wealth: A test of permanent income consumers". In: *Michigan Retirement Research Center Research Paper No. WP* 88.

Johnson, Eric, Stephen Atlas, and John Payne (2011). "Time preferences, mortgage choice and strategic default". In: *ACR North American Advances*.

Kirby, Kris N, Nancy M Petry, and Warren K Bickel (1999). "Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls." In: *Journal of Experimental psychology: general* 128.1, p. 78.

Krumme, Coco, Alejandro Llorente, Manuel Cebrian, Alex Pentland, and Esteban Moro (2013). "The predictability of consumer visitation patterns". In: *Scientific reports* 3.1, pp. 1–5.

Kuchler, Theresa and Michaela Pagel (2020). "Sticking To Your Plan: The Role of Present Bias for Credit Card Debt Paydown". In: *Journal of Financial Economics, forthcoming.*

Laibson, David (1994). "Essays in hyperbolic discounting". In: *Ph. D. dissertation, MIT.*

Laibson, David (1997). "Golden eggs and hyperbolic discounting". In: *The Quarterly Journal of Economics* 112.2, pp. 443–478.

Laibson, David and Keith Marzilli-Ericson (2019). "Intertemporal choice". In: *Handbook of Behavioral Economics* 2.

Lawrance, Emily C (1991). "Poverty and the rate of time preference: evidence from panel data". In: *Journal of Political economy* 99.1, pp. 54–77.

Levi, Yaron and Shlomo Benartzi (2020). "Mind the app: Mobile access to financial information and consumer behavior". In:

Loewenstein, George and Ted O'Donoghue (2004). "Animal spirits: Affective and deliberative processes in economic behavior". In: *Available at SSRN 539843*.

Loewenstein, George and Drazen Prelec (1992). "Anomalies in intertemporal choice: Evidence and an interpretation". In: *The Quarterly Journal of Economics* 107.2, pp. 573–597.

Lukas, Marcel F. and Ray Charles Howard (2022). "The influence of budgets on consumer spending". In: *Journal of Consumer Research*.

Lunt, Peter K and Sonia M Livingstone (1991). "Psychological, social and economic determinants of saving: Comparing recurrent and total savings". In: *Journal of economic Psychology* 12.4, pp. 621–641.

Madrian, Brigitte C and Dennis F Shea (2001). "The power of suggestion: Inertia in 401 (k) participation and savings behavior". In: *The Quarterly journal of economics* 116.4, pp. 1149–1187.

MAS, Money Advice Service (2014). *Money Lives: the financial behaviour of the UK*. URL: https://www.moneyadviceservice.org.uk/en/corporate/money-lives (visited on 04/07/2020).

McClure, Samuel M, Keith M Ericson, David I Laibson, George Loewenstein, and Jonathan D Cohen (2007). "Time discounting for primary rewards". In: *Journal of neuroscience* 27.21, pp. 5796–5804.

Medina, Paolina C (2021). "Side effects of nudging: Evidence from a randomized intervention in the credit card market". In: *The Review of Financial Studies* 34.5, pp. 2580–2607.

Meier, Stephan and Charles Sprenger (2010). "Present-biased preferences and credit card borrowing". In: *American Economic Journal: Applied Economics* 2.1, pp. 193–210.

— (2013). "Discounting financial literacy: Time preferences and participation in financial education programs". In: *Journal of Economic Behavior & Organization* 95, pp. 159–174.

Meyer, Steffen and Michaela Pagel (2018). "Fully closed: Individual responses to realized capital gains and losses". Tech. rep. Working paper.

Milkman, Katherine L, Todd Rogers, and Max H Bazerman (2009). “Highbrow films gather dust: Time-inconsistent preferences and online DVD rentals”. In: *Management Science* 55.6, pp. 1047–1059.

— (2010). “I’ll have the ice cream soon and the vegetables later: A study of online grocery purchases and order lead time”. In: *Marketing Letters* 21.1, pp. 17–35.

Mittal, Chiraag, Vladas Griskevicius, Jeffry A Simpson, Sooyeon Sung, and Ethan S Young (2015). “Cognitive adaptations to stressful environments: When childhood adversity enhances adult executive function.” In: *Journal of personality and social psychology* 109.4, p. 604.

Money Supermarket (2020). *What’s the difference between paying car insurance monthly or annually?* URL: https://www.moneysupermarket.com/car-insurance/annual-vs-monthly/ (visited on 08/05/2020).

MPS, Money and Pension Service (2018). “Building the Financial Capability of UK Adults”. Tech. rep. URL: https://moneyandpensionsservice.org.uk/2019/02/06/adult-financial-capability-building-the-financial-capability-of-uk-adults-survey/.

Muggleton, Naomi K, Edika G Quispe-Torreblanca, David Leake, John Gathergood, and Neil Stewart (2020). “Evidence from mass-transactional data that chaotic spending behaviour precedes consumer financial distress”. Tech. rep. DOI: 10.31234/osf.io/qabgm. URL: psyarxiv.com/qabgm.

O’Donoghue, Ted and Matthew Rabin (1999a). “Doing it now or later”. In: *American economic review* 89.1, pp. 103–124.

— (1999b). “Incentives for procrastinators”. In: *The Quarterly Journal of Economics* 114.3, pp. 769–816.

O2 Mobile (2013). *O2 launches new tariff allowing customers to get the latest phone whenever they want.* URL: https://news.o2.co.uk/press-release/o2-launches-new-tariff-allowing-customers-to-get-the-latest-phone-whenever-they-want/ (visited on 01/25/2020).

— (2018). *O2 unveils its revolutionary O2 custom plans.* URL: https://news.o2.co.uk/press-release/o2-custom-plans-2/ (visited on 01/25/2020).

— (2020). *Help and support – your first bill.* URL: https://www.O2.co.uk/help/account-and-billing/first-bill (visited on 01/25/2020).

Oaten, Megan and Ken Cheng (2007). “Improvements in self-control from financial monitoring”. In: *Journal of Economic Psychology* 28.4, pp. 487–501.

Olafsson, Arna and Michaela Pagel (2018). “The liquid hand-to-mouth: Evidence from personal finance management software”. In: *The Review of Financial Studies* 31.11, pp. 4398–4446.

Payne, John W, James R Bettman, and Eric J Johnson (1992). “Behavioral decision research: A constructive processing perspective”. In: *Annual review of psychology* 43.1, pp. 87–131.

Philipps, Jo, Annick Kuipers, and Will Sandbrook (2021). “Supporting emergency savings: early learnings of the employee experience of workplace sidecar savings”. Tech. rep.

Rambachan, Ashesh and Jonathan Roth (2022). “A More Credible Approach to Parallel Trends”. Tech. rep. Working Paper.

Read, Daniel (2001). “Is time-discounting hyperbolic or subadditive?” In: *Journal of risk and uncertainty* 23.1, pp. 5–32.

Read, Daniel, George Loewenstein, and Shobana Kalyanaraman (1999). “Mixing virtue and vice: Combining the immediacy effect and the diversification heuristic”. In: *Journal of Behavioral Decision Making* 12.4, pp. 257–273.

Read, Daniel, Rebecca McDonald, and Lisheng He (2018). “Intertemporal choice: Choosing for the future”. In: *The Cambridge handbook of psychology and economic behavior*, pp. 167–197.

Read, Daniel and Barbara Van Leeuwen (1998). “Predicting hunger: The effects of appetite and delay on choice”. In: *Organizational behavior and human decision processes* 76.2, pp. 189–205.

Reimers, Stian, Elizabeth A Maylor, Neil Stewart, and Nick Chater (2009). “Associations between a one-shot delay discounting measure and age, income, education and real-world impulsive behavior”. In: *Personality and Individual Differences* 47.8, pp. 973–978.

Roth, Jonathan, Pedro HC Sant’Anna, Alyssa Bilinski, and John Poe (2022). “What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature”. In: *arXiv preprint arXiv:2201.01194*.

Sabat, Jorge and Emily Gallagher (2019). “Rules of thumb in household savings decisions: Estimation using threshold regression”. In: *Available at SSRN 3455696*.

Samuelson, Paul A (1937). “A note on measurement of utility”. In: *The review of economic studies* 4.2, pp. 155–161.

Saunders, Bob and Gerard Fogarty (2001). “Time discounting in relation to career preferences”. In: *Journal of Vocational Behavior* 58.1, pp. 118–126.

Shannon, Claude Elwood (1948). “A mathematical theory of communication”. In: *The Bell system technical journal* 27.3, pp. 379–423.

Shapiro, Jesse M (2005). “Is there a daily discount rate? Evidence from the food stamp nutrition cycle”. In: *Journal of public Economics* 89.2-3, pp. 303–325.

Sims, Christopher A (2003). “Implications of rational inattention”. In: *Journal of monetary Economics* 50.3, pp. 665–690.

Skatova, Anya, Neil Stewart, Edward Flavahan, and James Goulding (2019). “Those Whose Calorie Consumption Varies Most Eat Most”. In:

Stewart, Neil, Emina Canic, and Timothy L Mullett (2019). "On the futility of estimating utility functions: Why the parameters we measure are wrong, and why they do not generalize". In:

Strotz, Robert Henry (1955). "Myopia and inconsistency in dynamic utility maximization". In: *The review of economic studies* 23.3, pp. 165–180.

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.

Sutter, Matthias, Martin G Kocher, Daniela Glätzle-Rützler, and Stefan T Trautmann (2013). "Impatience and uncertainty: Experimental decisions predict adolescents' field behavior". In: *American Economic Review* 103.1, pp. 510–31.

Thaler, Richard (1981). "Some empirical evidence on dynamic inconsistency". In: *Economics Letters* 8.3, pp. 201–207.

Thaler, Richard H and Shlomo Benartzi (2004). "Save more tomorrow$^{\text{TM}}$: Using behavioral economics to increase employee saving". In: *Journal of political Economy* 112.S1, S164–S187.

Thaler, Richard H and Hersh M Shefrin (1981). "An economic theory of self-control". In: *Journal of political Economy* 89.2, pp. 392–406.

Urminsky, Oleg and H Bayer (2014). "A meta-analytic review of time discounting measurement". Tech. rep. working paper, University of chicago.

Urminsky, Oleg and Gal Zauberman (2015). "The psychology of intertemporal preferences". In: *The Wiley Blackwell handbook of judgment and decision making* 2, pp. 141–181.

Vernon-Feagans, Lynne, Michael Willoughby, and Patricia Garrett-Peters (2016). "Predictors of behavioral regulation in kindergarten: Household chaos, parenting, and early executive functions." In: *Developmental psychology* 52.3, p. 430.

Warner, John T and Saul Pleeter (2001). "The personal discount rate: Evidence from military downsizing programs". In: *American Economic Review* 91.1, pp. 33–53.

Zuke, Elinor (2017). *Insurance customers need to swap around every year*. URL: https://www.consumerintelligence.com/articles/insurance-customers-need-to-shop-around-every-year (visited on 05/10/2021).