Essays in Consumer Financial Behaviour

Fabian Gunzinger

A thesis submitted in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy

Warwick Business School
University of Warwick

September 2023

Contents

List of Tables

List of Figures

Acknowledgements

Warwich Business School has made this project possible by funding it through one of its generous scholarships.

Redzo Mujcic, Zvi Safra, and Ivo Vlaev have provided encourating and critical comments along the way.

Neil Steward has been a tremendous supervisor. His insights and ideas have helped me move forward when I felt stuck, and his unfailing good humour and enthusiasm have cheered me up in moments of frustration.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. All work presented was carried out in collaboration with co-authors as follows:

Chapter 2 is co-authored with Neil Stewart (Warwick Business School, University of Warwick). Data was provided by Money Dashboard to Neil Stewart. Neil developed the concept of the paper and provided comments on the manuscript. I carried out the analysis and wrote the manuscripts.

Chapter 3 is also co-authored with Neil Stewart (Warwick Business School, University of Warwick). Data was provided by Money Dashboard to Neil Stewart. I developed the concept of the paper, carried out the analysis, and wrote the manuscript. Neil provided comments on the manuscript.

Chapter 4 is a single-author paper. Data was provided by Money Dashboard to Neil Stewart. I developed the concept of the paper, carried out the analsysi, and wrote the manuscript.

Abstract

Insights from behavioural science, if combined with FinTech apps and the data these apps generate, offers the promise to make it vastly simpler for individuals to manage their finances and, through that, to improve their financial wellbeing. This thesis consists of three stand-alone chapters, all of which aim to contribute to fulfilling that promise. They all use the same dataset of financial transaction data from a large number of users of a financial aggregator app. The first chapter introduces and tests a new approach that allows researchers to elicit behaviour traits of the kind used in lab and survey studies for individuals in large datasets. The approach allows researchers to augment the information contained in large datasets and thus to fully leverage such datasets for the study of human behaviour. The second chapter tests whether the way individuals spend their money is related to how frequently they make transfers into their emergency savings funds. The overall aim of the chapter is to test whether simple summary statistics based on spending spending profiles – the way in which individuals allocate their spending transactions across different product and merchant categories – capture information about individuals' life circumstances that are predictive of their financial behaviour. If so, such statistics could be used in financial management apps, to, for instance, provide extra and customised assistance in time of need. The third chapter tests whether using Money Dashboard – the financial aggregator app that provides the data for all chapters of this thesis – is associated with a reduction in discretionary spend and an increase in emergency savings. The aim is to contribute to a better understanding of the extent to which the potentially large benefits of FinTech apps already help users achieve their financial goals and thus raise financial wellbeing.

Chapter 1

Introduction

The overall aim of this thesis is to use transaction-level data from a financial aggregator app, insights from behavioural science, and methods from econometrics and machine learning to contribute to a better understanding of the determinants of how individuals spend their money and save for the short-term. This introductory chapter first discusses the context within which this research takes place, and then introduces the main chapters of the thesis.

1.1 Context

1.1.1 Financial wellbeing

Financial wellbeing is a state in which a person can make ends meet in the present, can feel comfortable about their financial future, can feel comfortable about money, and has the financial freedom to make choices that allow for the enjoyment of life. This includes having control over one's day-to-day and month-to-month finances, not having to borrow to meet ongoing obligations, being free of debt or being in control of it, having the capacity to meet unexpected expenses, and – for a working-age individual – being on track to build enough savings for retirement.[1]

Financial wellbeing is important because a lack of it can have severe consequences both in the short-term and in the long-term. In the short-term, low

[1] The US Consumer Financial Protection Bureau defines financial wellbeing as "a state of being wherein a person can fully meet current and ongoing financial obligations, can feel secure in their financial future, and is able to make choices that allow enjoyment of life" (CFPB 2015). The UK's Money and Pension service defines current financial wellbeing as "being able to pay the bills and feel comfortable about money" and longer-term financial security as "having the savings to deal with the expected, the unexpected and the longer-term; or having some form of loss protection such as home contents or life insurance" (MPS 2018). Both organisations also differentiate some aspects of financial wellbeing depending on whether an individual is in their working or retirement age. Throughout, this discussion, as well as in the analysis in the following chapters, I focus on working age individuals.

financial wellbeing means a constant struggle to make ends meet, being overwhelmed by debt, having circumstances dictate one's of and approach to money, a higher likelihood of experiencing material hardship – which (CFPB 2017) defines as running out or worrying about running out of food, not being able to afford medical treatment or a place to live, or have utilities turned off – and the need to cut back on essentials for one's children.[2] All of these experiences can cause a decline in physical and mental health, a loss of productivity, and an overall deterioration in the quality of life.

In the long-term, the danger is that the situation is not only self-sustaining but leads to a vicious circle that becomes increasingly hard to escape. This can happen purely because having to borrow for regular expenses like food and bills – as, the Money and Pension service estimates, about 9 million adults in the UK do (MPS 2018) – can lead to a situation where an increasing amount has to be spent on servicing debt and thus increases monthly expenses and the borrowed amounts required to cover them. In addition, however, this could happen because of increasingly impaired decision making. A literature on scarcity documents that our minds tend to focus on what is scarce and neglect what is not, concentrating our mental resources where they are most needed but reducing cognitive bandwidth in other domains, which can lead to poorer decision making (Shah et al. 2012, Mullainathan and Shafir 2013, Haushofer and Fehr 2014). For instance, Mani et al. (2013) find that low-income shoppers in New Jersey perform worse on cognitive tasks when first promoted to think about their financial situation while the same prompts had no effect for wealthier shoppers, and sugar cane farmers in India perform worse on similar cognitive tasks shortly before the annual harvest (when money is scarce) than shortly thereafter (when money is plentiful). There is also evidence that scarcity might lower productivity in the present: Kaur et al. (2021) randomise the timing of wage payments to low-income piece-rate manufacturing workers and find that workers that receive their wages early and are thus no longer liquidity constrained make fewer mistakes and increase their output by 7 percent. Hence, if the anxiety caused by having little money to cover current expenses lowers individuals' productivity in the present and makes it harder to engage in rational planning for the future – how to reduce spending, how to build skills that would help find a higher-paying job, how to transition to a healthier lifestyle that would help them feel and perform better physically and mentally – then low financial wellbeing in the present might beget low financial wellbeing in the future.

The determinants of financial wellbeing are a combination of circumstantial

[2] For more detailed descriptions about the consequences of financial hardship, see CFPB (2017), MPS (2018), and Step Change (2017).

and external factors as well as the capabilities, beliefs, and behaviours of the individual. MPS (2018) provides a useful categorisation that sees financial wellbeing as a function of four broad factors: external factors that include economic conditions, demographics, and a person's social environment; enablers that include financial confidence and numeracy, a sense of control, one's spending and savings mindset, and one's engagement with money, advice, and technology; day-to-day behaviours like managing the use of credit, avoiding to borrow for everyday spends, active saving, keeping track of and making adjustments to one's spending, and shopping around; and, finally, planning ahead behaviours like building financial resilience through saving and planning for retirement.[3]

While no one factor is deterministic for financial wellbeing, research from the US (CFPB 2017) and the UK (MPS 2018) agrees that, as expected, external economic factors such as access to education and higher paying jobs are important. In the UK, for instance, individuals with an annual income of £20,000 or below account for 41 percent of the working age population but for 69 percent of those with less than £100 of savings. Similarly, 50 percent of people who borrow to cover everyday expenses earn less than £17,000. So, clearly, a search for and support of effective economic and social policy measures should be an important part of any effort to improve societal financial wellbeing. But the same research also shows that higher incomes are not sufficient. In the UK, 18 percent of individuals with less than £100 in savings have a household income of £30,000 or higher, and 20 percent of those who borrow to cover everyday expenses have an income of £50,000 or higher. In the US, too, there is large variation in the characteristics of individuals at each level of financial wellbeing; the financial wellbeing of the top quarter of people with a high-school degree, for instance, is higher than that of the bottom half of those with graduate degrees.

In both countries, research shows that the level of savings is a key contributor to financial wellbeing. In the US, it is the one factor that discriminates between different levels of financial wellbeing better than any other examined factor (CFPB 2017). In the UK, it is, together with behaviour towards credit, the strongest predictor of financial wellbeing (MPS 2018). In fact, research by a UK charity suggests that having £1,000 in liquid savings could reduce the probability of being in debt by almost half. In particular, having a habit of saving regularly – even more so than the amounts saved – has been found to be a key determinant. Other factors that are positively associated with higher financial wellbeing are confidence in one's ability to achieve one's financial goals, not using debt to cover everyday expenses, paying one's bills on time, staying within one's

[3]The US Consumer Financial Protection Bureau uses a similar classification in its definition of financial wellbeing (CFPB 2015).

budget and spending plan, paying credit card balances in full, and checking bank statements for errors.[4]

But many people struggle with these behaviours. For instance, in the UK, 21 percent of the working-age population (10.7 million adults) report to rarely or never save, and 22 percent of the population have less than £100 in savings, with those holding that amount in a formal savings account being even lower (MPS 2018). Unsurprisingly, then, one in four adults could not pay an unexpected bill of £300 from their own money (Phillips et al. 2021). In the US, the situation is similar: 30 percent of adults report that they would be unable to cover a bill of $400 in cash or its equivalents (Governors of the Federal Reserve System 2022), and, according to a 2013 survey, liquid net worth for the median household with a head aged 41-51 is only $813, whereas at the 25th percentile it is -$1,885 (Beshears, Choi, Laibson, and Madrian 2018).[5]

This is problematic because many households do face unexpected financial shocks over the course of a year. Also, many households experience high income volatility: in the US between 2013 and 2018, the median month-to-month change in household income was 36 percent, with low-income households experiencing more frequent and larger income dips (Chase 2019). The same research finds that while families need roughly six weeks of take-home income in liquid assets to weather a simultaneous income dip and expenditure shock, 65 percent of households lack such a buffer. In line with these findings, Roll and Despard (2020) find that during the coronavirus pandemic, households with liquid asses of above $2,000 had significantly lower risk to experience financial distress (indicators like skipping essential bills, being behind on credit card debt, being in overdraft) than households with lower savings.

Managing debt is similarly challenging for many. 9 million people in the UK also borrow to cover expenses for food and bills (MPS 2018), and many struggle to stay on top of their credit card debt: 2 million cards were in arrears or default, another 2 million carried persistent debt, and for another 1.6 million cards, owners were persistently making minimum payments only. Altogether, the study found that 5 million accounts (9 percent of the total) that were active in January 2015 would, under their current repayment pattern and without further borrowing, take 10 years or more to repay their balance (FCA 2016). In addition to holding high and persistent balances, a large body of research also indicates that individuals make other mistakes in debt management, especially in dealing with credit cards: they choose suboptimal credit card contracts (Agarwal, Chomsisengphet,

[4]For details on contributors to financial wellbeing see (CFPB 2017) for the US and MPS (2018) for the UK.

[5]Liquid net worth includes all assets (except pension wealth, retirement savings, homes, and durable assets) net of all debt (except student loans and collateralised debt).

Liu, et al. 2015), sometimes because they are overly susceptible to temporarily low teaser rates (Shui and Ausubel 2004, Ausubel 1991); they borrow on payday loans despite being far from having exhausted their credit card limits (Agarwal, Skiba, et al. 2009), their repayment amount is overly influenced by stated minimum payments (Sakaguchi et al. 2022), they pay down debt across different cards proportionally to outstanding balances instead of prioritising high-interest cards (Gathergood, Mahoney, et al. 2019), and they sometimes hold credit card debt and liquid assets at the same time (Gross and Souleles 2002, Gathergood and Olafsson 2020).[6]

1.1.2 Behavioural science

Behavioural science can help address these challenges by understanding the factors that determine financial behaviour and by designing and testing possible solutions. The main factors that have been found to influence financial decision making are: cognitive limitations and financial literacy (Agarwal, Driscoll, et al. 2009, Agarwal and Mazumder 2013, Korniotis and Kumar 2011, Agarwal, Amromin, et al. 2010, Fernandes et al. 2014, Jørring 2020); time-preferences and self-control (Frederick et al. 2002, Read, McDonald, et al. 2018, Ericson and Laibson 2019, Cohen et al. 2020); attitude towards money and spending (Rick, Cryder, et al. 2008, Rick, Small, et al. 2011); one's perceived locus of control (Perry and Morris 2005), degree of optimism (Puri and D. T. Robinson 2007), ability to frame decisions broadly rather than narrowly (Kumar and S. S. Lim 2008), and propensity to gamble (Kumar 2009); one's social network (Bailey, Cao, et al. 2018, Kuchler and Stroebel 2021); the degree of one's financial planning (Ameriks, Caplin, and Leahy 2003); and habits (J. Blumenstock, Callen, et al. 2018, Schaner 2018, De Mel et al. 2013).[7]

Researchers have also developed and tested a large number of approaches designed to help people make better decisions. One main area of research here aims to address limited self-control. I briefly discuss four such approaches that have the potential to help people make the financial choices they themselves would like to make. A complete review of that literature is provided by Duckworth et al. (2018).

One extensively studied approach is the commitment device, whereby an individual restricts their future choice set in order to avoid choosing a self-defeating action. While not everybody makes use of such devices when offered the opportu-

[6] See Agarwal, Chomsisengphet, and C. Lim (2017) for a more complete review of a large body of research documenting consumer choice inefficiencies and suboptimal financial behaviour.

[7] For two thorough reviews, see Agarwal, Chomsisengphet, and C. Lim (2017) and Greenberg and Hershfield (2019).

nity (Bryan et al. 2010), and while they do not work in all contexts (Laibson 2015, C. D. Robinson et al. 2018), they have been found to help individuals increase their savings rates (Ashraf et al. 2006), quit smoking (Giné et al. 2010), make healthier food choices (Schwartz et al. 2014), and exercise more regularly (Royer et al. 2015). A popular implementation of a commitment device is the platform https://www.stickk.com, which helps users define their goals and leverages the power of loss aversion and accountability to help people follow through with their desired behaviour change: it allows users, for instance, to commit to donating money to a cause they do not support if they fail to exercise as often as they wanted to, or to nominate a friend or family member who will verify whether they managed to stick to their spending goal.

Another approach are implementation intentions, a particular type of planning for the achievement of one's goals that involves "if-then" intentions, such as "if I get paid, then I transfer 10 percent of it into my savings account" (P. M. Gollwitzer and Sheeran 2006, Rogers, Milkman, et al. 2015). Such intentions have been found to support perseverance in pursuing one's goals (Oettingen and P. Gollwitzer 2010) and to increase overall goal attainment across different age groups, life domains, and types of obstacles (P. M. Gollwitzer and Sheeran 2006).

A third intervention that has the potential to alter financial behaviour is social norms messaging, whereby people are informed about how their own behaviour compares with that of a relevant peer group. Such information can be especially useful for domains where such information is usually not available, as is the case with spending and saving, and has been successful inducing high energy use households to lower their energy use without inducing low-use households to increase theirs (Schultz et al. 2007, Allcott 2011, Allcott and Rogers 2014, Brandon et al. 2017). However, the information can also backfire and has been found to lower participation in pension savings plans (Beshears, Choi, Laibson, Madrian, and Milkman 2015) and completion rates of an online course (Rogers and Feller 2016), making it important to test messages before deploying them on a large scale to avoid unintended consequences.

Finally, changing default options has been found to be a powerful tool for behaviour change. Defaults are consequential because people often stick with the status quo (W. Samuelson and Zeckhauser 1988), and because they tend to interpret defaults as a recommendation (McKenzie, Liersch, and Finkelstein 2006), implicitly view it as a reference point moving away from which would feel costly (E. J. Johnson and Goldstein 2003, Kahneman and Tversky 1979), and because even if they resolve make a different choice, they often procrastinate (Carroll et al. 2009, Ericson 2017). Default options have been applied across range of areas and have, for instance, been found to increase retirement savings contributions (Madrian and Shea 2001, Beshears, Choi, Laibson, and Madrian 2009) and organ donations (E. J. Johnson and Goldstein 2003, Gimbel et al. 2003, Abadie and Gay 2006).

These approaches, together with many others, have the potential to make it easier for individuals to manage their money and to spend and save in line with their own goals. To realise that potential, however, they have to be tested on a large scale and at low cost to gather reliable results and to have the ability to tweak and refine them. Furthermore, solutions that are found to work have to be deployed on a large scale to have a meaningful impact. The advent of online and mobile banking and the large amount of data they create enable researchers and financial institutions to do just that.

1.1.3 FinTech

The availability of large amounts of granular data about human behaviour, together with ongoing progress in the development of machine-learning algorithms elicit information from them, and the increase in computer power that allows for the storage and processing of the data and the deployment of these algorithms, offers enormous potential for better understanding human behaviour (Jaffe 2014, Buyalskaya et al. 2021) and for building applications that help people live in line with their goals.

Platforms like Facebook provide the opportunity to recruit study participants at a scale and with an ease that was not possible before (Kosinski et al. 2015), and data from its platform has been used, for instance, to predict participant personalities and life outcomes (Youyou et al. 2015), and to study the effect of social interactions on house purchasing decisions (Bailey, Cao, et al. 2018, Bailey, Dávila, et al. 2019), the spatial structure of urban networks (Bailey, Farrell, et al. 2020), the role of peer effects in product adoption (Bailey, Johnston, Kuchler, et al. 2019), the role of social networks in determining social distancing during the coronavirus pandemic (Bailey, Johnston, Koenen, et al. 2020), and the determinants of social mobility (Chetty et al. 2022a, Chetty et al. 2022b). Similarly, log and sensor data from mobile phones, another increasingly popular data source in the study of human behaviour, has been used to successfully predict personality traits (Montjoye et al. 2013, Stachl et al. 2020), predict levels of wealth (J. Blumenstock, Cadamuro, et al. 2015), improve human aid targeting (Aiken et al. 2022), and study the role of social networks for migration patterns (J. E. Blumenstock et al. 2019). Some studies also combine data from a range of sources; in one ambitious project, Chi et al. (2022) use data from satellites, mobile phone networks, topographical maps, and aggregated and de-identified connectivity data from Facebook to map levels of wealth and poverty at a 2.4km resolution for all 135 low and middle-income countries (Chi et al. 2022).

In the area of household finance, the advent of online and mobile banking is creating an every growing amount of transaction-level data and a range of FinTech apps aimed to help people manage their money. There are three ways in which such data and apps act as critical enablers for an improved understanding of financial behaviour and for the development, testing, and deployment of possible solutions that help increase financial wellbeing.

First, large-scale and granular transaction data are an invaluable resource to help researchers refine their understanding of the behavioural determinants of financial decision making because they tends to have one or many of the following attributes: it is collected automatically and thus more reliably, it provides a more complete view of an individual's financial life, it spans a longer time horizon, it covers a larger set of individuals, and the cost of collection (though not necessarily cleaning) is zero. Accordingly, such data has already been used to study a range of issues. For instance, Kuchler and Pagel (2020) show that a failure to stick to self-set debt paydown schedules is best explained by individual's present biasedness, Gelman et al. (2014) and Olafsson and Pagel (2018) show that consumer spending varies across the pay cycle, S. R. Baker (2018) and Baugh et al. (2014) study consumer spending responses in response to exogenous shocks, Carlin et al. (2019) document generational differences in financial platform use, 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.[8]

Second, the FinTech apps from which such data are often gathered allow for testing and the gradual refinement of possible solutions at high speed and relatively low cost. Furthermore, because these apps are often built to help users manage their money, they have a clear incentive to do so, something that is not true to the same extent for more traditional banks. Some work that tests the effectiveness of FinTech apps aimed to facilitate money management has already been conducted: Gargano and Rossi (2021) find that an app that allows users to set savings goals does indeed increase savings, and Levi and Benartzi (2020) and Carlin et al. (2022) find, respectively, that improved access to financial information reduces discretionary spending and non-sufficient fund fees.

But there is large potential for further work in this area. For instance, while

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

the work of Levi and Benartzi (2020) suggests that lowering the cost of access to financial information can improve financial outcomes, little is known about what design features of mobile apps are particularly helpful. There are also features that are not currently in widespread use that could address known biases. One example is our tendency to underestimate the power of exponential growth. McKenzie and Liersch (2011) find that highlighting to participants the effects of exponential growth motivates them to save more. It would be easy for mobile apps to project accumulated future savings for different monthly savings amounts, say, and thus make the future loss from spending more money in the present more salient. Another example is mental accounting. Soman and Cheema (2011) find that earmarking part of peoples' salaries as savings increases savings rates. FinTech apps could, for instance, suggest to automatically transfer fixed amount into labelled savings pot, thus making that money less likely to be spent later. Finally, increasingly sophisticated "robo-advisers" – financial advice provided by machine-learning algorithms – has the potential to democratise personalised financial advice across an increasing number of domains, from portfolio allocation to consumption decisions to debt management (Philippon 2019, D'Acunto and Rossi 2021).

In testing and refining possible solutions, the large size of the data is critical because it allows for the identification of heterogeneous effects and allows for the appropriate customisation of solutions. Having a view of users' entire financial lives – as in the case of financial aggregator apps that allow users to add accounts from all their different financial providers – is also critical, since it allows for the identification of unintended consequences; for example, inducing credit card users to make higher automatic card repayments may increase automatic repayments but leave the total repayment amount unchanged (Guttman-Kenney et al. 2021), and auto-enrolling employees into pension schemes might increase their debt levels (Beshears, Choi, Laibson, Madrian, and Skimmyhorn 2022).

Finally, FinTech apps will make it relatively simple and cheap to make effective solutions and tools widely available to a large audience, which is critical to the goal of increasing financial wellbeing on a large scale.

All these opportunities from ever-growing data and increasingly sophisticated algorithms and powerful computers bring – like all innovations – a set of challenges. They can, for instance, compromise individual privacy and propagate social biases. Inevitably, technological progress has largely outpaced regulatory responses, and it will take time to understand all the implications of these new technologies and enshrine proper safeguards into law. In the meantime, leading researchers in the field are charting the course in thinking about how to work with these data and tools responsibly (De Montjoye et al. 2015, Kosinski et al.

2015, J. Blumenstock 2018), and private companies as well as research institutions have put in place safeguarding procedures to ensure that such best practices and current regulations are followed at the level of individual research projects.

1.2 Thesis outline

The thesis consists of three stand-alone chapters, each of which addresses a different aspect of the overall aim of the thesis – to combine large-scale transaction data, insights from from behavioural science, and methods from econometrics and machine learning to study how individuals spend and save.

Chapter 2 introduces and tests a new method that allows researchers to elicit behaviour traits of the type used in lab and survey studies for individuals in large datasets. This work contributes to the overall thesis purpose indirectly in that it aims to augment the information contained in large datasets with additional information that opens up additional avenues of research and helps researchers fully leverage the data in their study of human behaviour.

Chapter 3 tests whether the way individuals spend their money is related to how frequently they make transfers into their emergency savings funds. The overall aim of the chapter is to test whether simple summary statistics based on spending spending profiles – the way in which individuals allocate their spending transactions across different product and merchant categories – capture information about individuals' life circumstances that are predictive of their financial behaviour. If so, such statistics could be used in financial management apps to, for instance, provide extra and customised assistance in time of need.

Chapter 4 tests whether using Money Dashboard – the financial aggregator app that provides the data for all chapters of this thesis – is associated with a reduction in discretionary spend and an increase in emergency savings. The aim is to contribute to a better understanding of the extent to which the potentially large benefits of FinTech apps already help users achieve their financial goals and thus raise financial wellbeing.

Chapter 2

Machine-learned behaviour traits

2.1 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 instalments, and swapping car and home insurer frequently or infrequently. Each of these choices 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 different 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 of salary receipt and spend on immediately consumed goods 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 are unavailable,

[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 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), 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 2.2 provides a brief introduction to the economic models of intertemporal choice that provide the conceptual framework for our approach, Section 4.2 introduces the dataset used for our analysis and discusses data pre-processing and sample selection; Section 4.3 introduces in detail our empirical approach and estimates time-preferences; Section 2.5 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.

2.2 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 far-sighted 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 P. A. 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 τ periods from now relative to immediately experienced utility. Such a model can be written as:

$$U_t = \sum_{\tau=0}^{T-t} D(\tau)u(c_\tau). \tag{2.1}$$

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 ρ is the *discount rate*. We can thus rewrite the equation above as:

$$U_t = \sum_{\tau=0}^{T-t} \delta^\tau u(c_\tau). \tag{2.2}$$

This is the exponential discounting model. Because δ 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 83 percent, implying that they value consumption one year from today at 83 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 δ, 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

[2] In general, we have indifference if $u(c_t) = \delta u(c_{t+1})$, where $\delta = 1/(1+\rho)$. Here, we have indifference if $u(c_{t+1}) = 1.2u(c_t)$. Hence: $u(c_t) = \delta 1.2u(c_t) = \frac{1.2}{1+\rho}u(c_t)$, which implies that $1.2 = 1+\rho$, so that we have $\rho = 0.2$ and $\delta = 0.83$.

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

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 τ 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 (2.2), the model rapidly became economists' standard framework to study intertemporal choice.

Yet over 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, 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 Ericson and Laibson (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 behavioural 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

$$D(\tau) = \begin{cases} 1 & \text{if } t = 0 \\ \beta\delta^t & \text{if } t > 0, \end{cases} \tag{2.3}$$

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

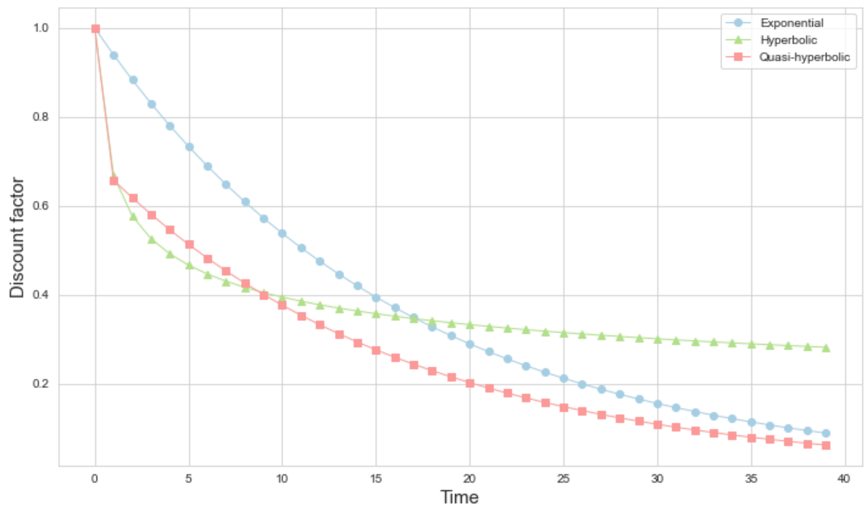
so that the full model becomes:

$$U_t = u(c_t) + \beta \sum_{\tau=1}^{T-t} \delta^{\tau} u(c_{\tau}), \tag{2.4}$$

where $0 < \beta, \delta \leq 1$. The simple addition to the DU model is the parameter β, 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 trade-offs and impatient when making short-run trade-offs.

Figure 2.1: Comparison of different discount functions

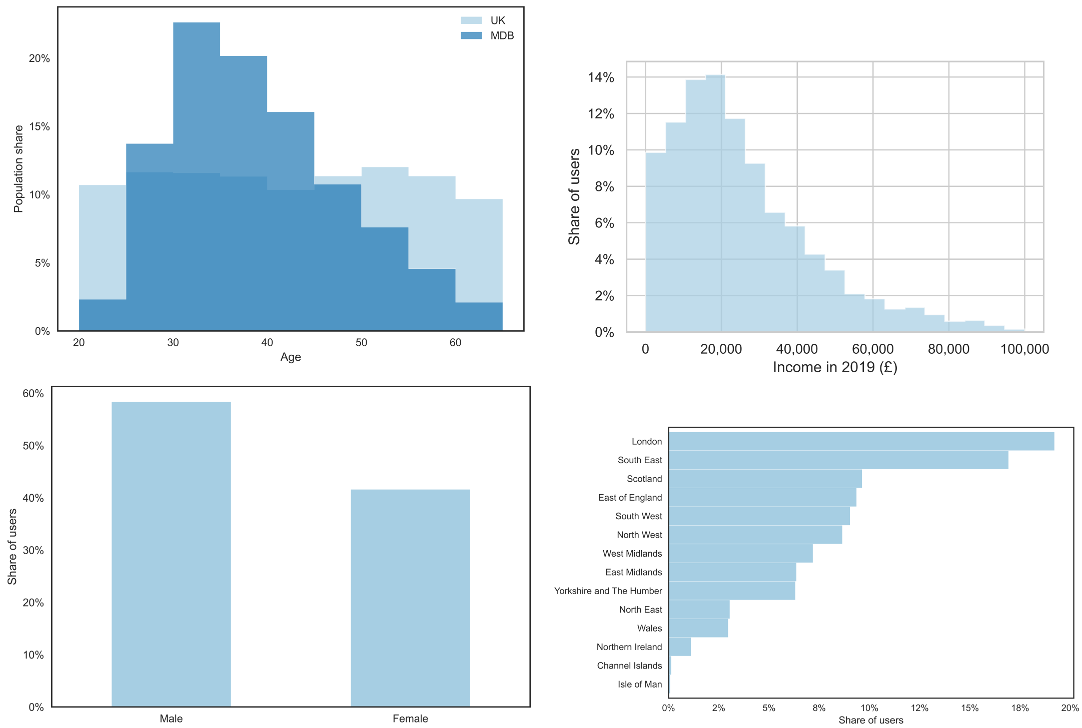


Notes: Comparison of exponential, hypterbolic, and quasi-hyperbolic discount functions. The functions are parametrised such as to provide an illustrative comparison of the different ways to model discounting behaviours. The precise parameter values are: exponential with $\delta = 0.94$, hyperbolic with $\alpha = 4$ and $\gamma = 1$, and quasi-hyperbolic with $\delta = 0.94$ and $\beta = 0.7$.

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 2.1 visualises the differences and similarities between stylised versions of the three different discount functions.

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

Figure 2.2: 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 in the MDB data.

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.

2.3 Data

2.3.1 Dataset description

Data are 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).

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 2.2 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 are not dependent on having a representative sample of people.

2.3.2 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.

2.3.3 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 4.1 shows how each criteria affects sample size.

Table 2.1: Sample selection

	Users	Accounts	Transactions	Value (£M)
Raw sample	27,111	130,325	55,125,624	1,187.8
At least 6 months of data	23,478	122,312	54,541,339	1,176.2
At least one current account	22,184	118,794	52,854,156	1,158.0
At least 5 monthly debits totalling £200	14,750	77,796	38,758,413	866.4
Income payments in 2/3 of all observed months	8,817	53,228	25,420,978	561.8
Yearly incomes between 5k and 100k	5,202	28,786	13,475,729	216.7
No more than 10 active accounts in any year	4,936	23,129	11,922,181	179.1
Debits of no more than 100k in any month	4,665	21,129	10,940,880	101.3
Working-age	3,724	16,608	9,388,334	83.8
Final sample	3,724	16,608	9,388,334	83.8

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 removing all users that do not meet the selection criteria at 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.

2.4 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.

2.4.1 Identifying choice scenarios

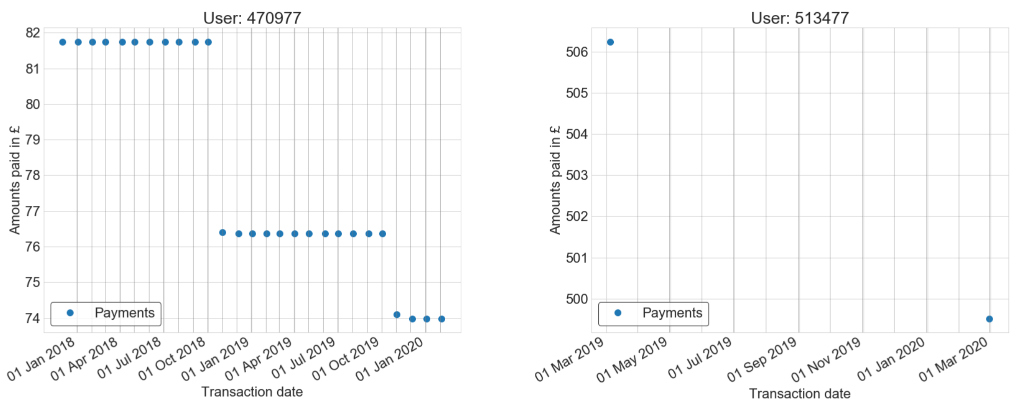
For each of the three choice scenarios, this section discusses the rational, the approach used to classify users into different types, and the limitations of that approach.

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 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 2.3 provides an example of a typical monthly payer (left panel) and a typical yearly payer (right panel).

Figure 2.3: Monthly and yearly car insurance payment patterns



Notes: Illustrative case studies for a users paying for their car insurance monthly (left) and annually (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

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

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 handset 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 rely on a contractual feature of O2, one of the largest UK mobile providers, which takes handset 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 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 £20,000 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.

2.4.2 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 features 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

$$H = -\sum p_i log_2(p_i), \tag{2.5}$$

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

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}, \tag{2.6}$$

and

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}. \tag{2.7}$$

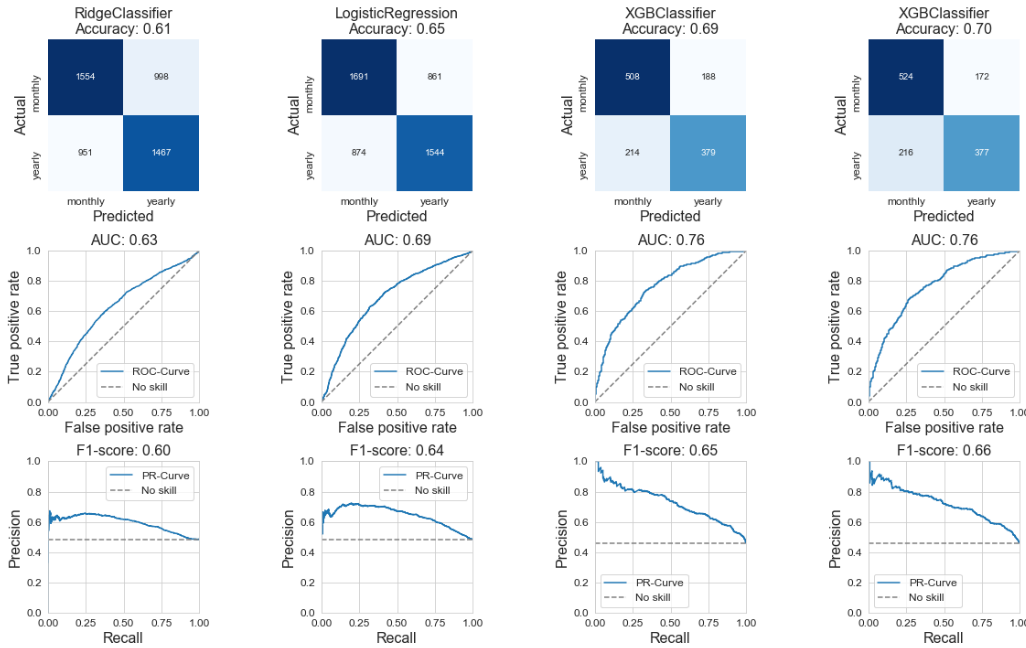
The two concepts originate in information retrieval, and capture two intuitive ideas: precision is the proportion of all true positives out of all returned positives, while recall is the proportion of all true positives that have been returned as such out of all existing true positives. A model might 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:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{2.8}$$

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 2.4 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 2.4: 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

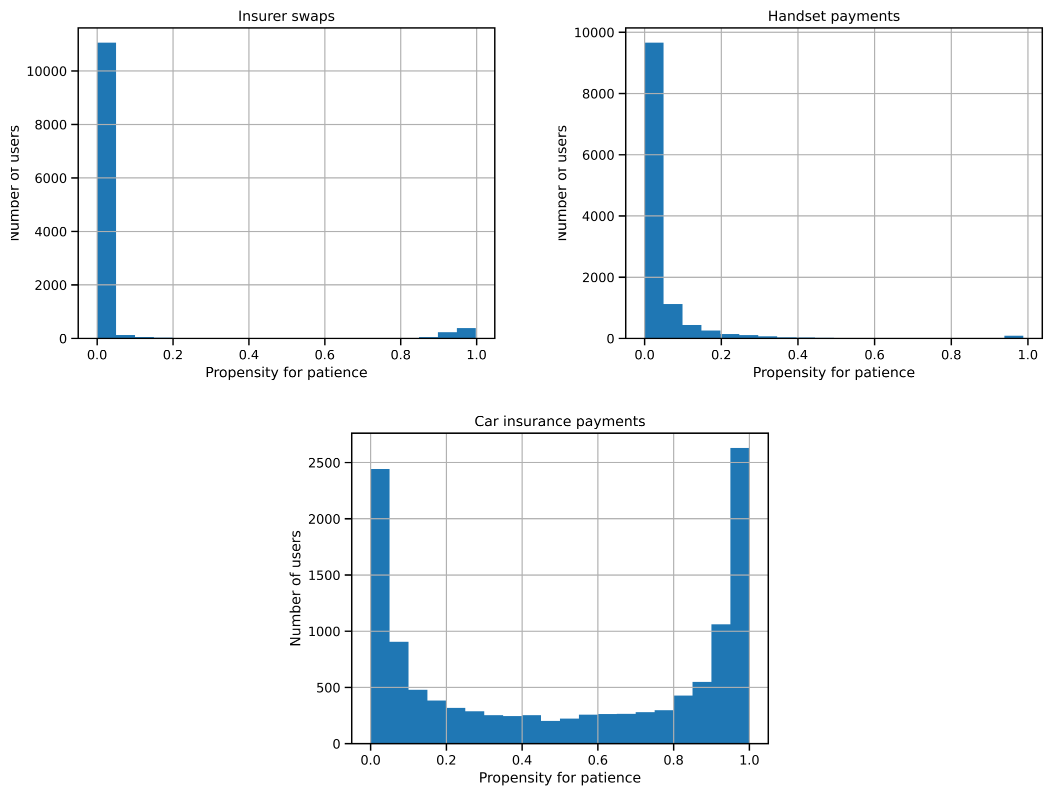
The XGBoost model outperforms all other models for all three scenarios. We optimise its performance by tuning some of its hyperparameters using grid search and 10-fold cross validation. Parameters we optimise for are the maximal tree depth (2, 3, 4, 5, 6), the learning rate (0.01, 0.05, 0.1, 0.3, 0.5),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

Column four in Figure 2.4 shows the XGBoost model performance after fine-tuning and shows that we managed to perform performance only slightly.

2.4.3 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 2.5: Propensity for patience estimates



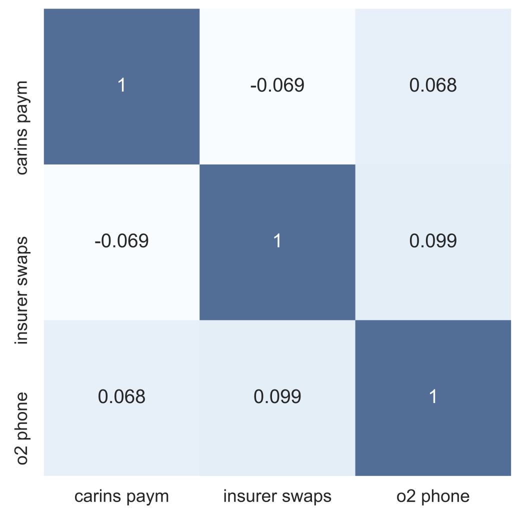
Notes: Estimated propensity for patience scores for based on Insurer swaps (top left), mobilephone handset payments (top right), and car insurance payments (bottom).

Figure 2.5 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.

2.4.4 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 2.6 shows, however, that the correlation among the three propensities is virtually zero.

Figure 2.6: 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. Numbers represent Pearson correlation coefficients.

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

Noisy data are a problem because identifying choice scenarios and classifying users's choice is not straightforward. Figure 2.3 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 are 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 handset even though they really paid monthly or maybe did not even buy a new handset). 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 and is so large that risk-preferences do not generalise evena cross very similar contexts. 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 handset 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.

2.4.5 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).

2.4.6 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 (E. 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, Caplin, Leahy, and Tyler 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.

2.5 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 4.3, 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 4.3, 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 2.2, 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

[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.

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 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 3

Spending entropy predicts savings behaviour

3.1 Introduction

This paper documents variation in spending profiles – the way spending transactions are allocated across product of merchant categories – for a large set of users of a financial management app and shows that spending profiles predict "emergency savings" – short-term savings intended as a buffer against unexpected financial shocks.

Emergency 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 wellbeing (MPS 2018, CFPB 2017), yet while there is a large literature on pension savings – particularly on the effect on auto-enrollment on savings plan participation – little is known about how individuals save for the short-term.[1] Apart from Beshears, Choi, Iwry, et al. (2020), who study design options for automated contributions to emergency saving pots, and Phillips et al. (2021), who test such a design in the UK, there is no research on how we can help individuals build such emergency savings. Furthermore, around a quarter of adults in the UK and the US are unable to cover irregular expenses like car and medical bills: in the UK, 25 percent of adults would be unable to cover an unexpected bill of £300 (Phillips 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). Research from the UK suggests, however, that even having £1000 in savings reduces by more than half a household's chances of falling into debt that leads to financial

[1] For main contributions to the to literature on auto-enrollment into savings plans, see Madrian and Shea (2001), Choi et al. (2002), Choi et al. (2004), and Beshears, Choi, Laibson, and Madrian (2009).

problems (Phillips et al. 2021).

There is a large literature on the determinants of financial behaviour, and some of the main factors that have been identified to influence have individuals and households make financial decisions are: cognitive limitations and financial literacy (Agarwal, Driscoll, et al. 2009, Agarwal and Mazumder 2013, Korniotis and Kumar 2011, Agarwal, Amromin, et al. 2010, Fernandes et al. 2014, Jørring 2020); time-preferences and self-control (Frederick et al. 2002, Read, McDonald, et al. 2018, Ericson and Laibson 2019, Cohen et al. 2020); attitude towards money and spending (Rick, Cryder, et al. 2008, Rick, Small, et al. 2011); ones perceived locus of control (Perry and Morris 2005), degree of optimism (Puri and D. T. Robinson 2007), ability to frame decisions broadly rather than narrowly (Kumar and S. S. Lim 2008), and propensity to gamble (Kumar 2009); ones social network (Bailey, Cao, et al. 2018, Kuchler and Stroebel 2021); the degree of ones financial planning (Ameriks, Caplin, and Leahy 2003); and habits (J. Blumenstock, Callen, et al. 2018, Schaner 2018, De Mel et al. 2013).[2]

In this paper, we use simple summary statistics of individuals' spending profiles and test whether differences in spending profiles are predictive of the frequency with which individuals make transactions into their savings accounts. To summarise spending profiles, we calculate the entropy of an individual's spending profile in a given period, which captures the predictability of spending transactions in said period. We focus on spending profiles for three reasons: first, our understanding of how individuals spend their money is largely based on survey data from a relatively small number of individuals. Large-scale transaction data of the type used in this paper has become available to researchers only in recent years and has not, to the best of our knowledge, been used to explore patterns in consumer spending behaviour.

Second, research suggests that spending profiles might reflect circumstances in an individual's life that are also related to their saving behaviour. For instance, Muggleton et al. (2020) find that higher spending entropy predicts financial distress, and hypothesise that this is because disorder in ones personal life might simultaneously be reflected in ones spending behaviour and make it harder to deal with financial obligations, based on a literature in psychology that suggests that disorder is 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 addition to financial outcomes, high entropy behaviour has also been found to be associated with higher calorie intake (Skatova et al. 2019) and with a higher

[2]For two thorough reviews, see Agarwal, Chomsisengphet, and C. Lim (2017) and Greenberg and Hershfield (2019).

number of visits to and higher spend in supermarkets (Guidotti et al. 2015).

Finally, we think of our focus on spending profile as a proof of concept for the broader agenda of testing whether we can extract simple summary statistics from large-scale transaction data that capture relevant information about individuals' life circumstances. If successful, such information could be used in the design of financial management apps to provide timely and customised assistance to users when they need it most.

Our results show that spending entropy is predictive of the frequency of savings transactions, with an effect size similar to that of an increase in monthly income income of between £1,000 and £2,000. This holds true if we control for simple component parts of entropy, suggesting that the non-linear way in which entropy combines these components captures something about spending profiles that is of relevance. The results are also largely consistent across entropy measures calculated based on different product and merchant categories. However, they direction of the effect changes for different types of entropy that differ in the way in which we treat product or merchant categories in which an individual makes no transactions in a given period. Exploratory research suggests that the number of such zero count categories, together with the variation of counts for categories with a positive number of transactions, goes some way in explaining the outcome. But more work is necessary to more fully understand this outcome, which we leave for future research.

Apart from the small literature on emergency savings mentioned above, our work contributes to a broader literature that uses financial-transaction data from banks or financial aggregator apps to understand consumer financial behaviour. For instance, 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]

The remainder of this paper is organised as follows: Section 3.2 discusses data pre-processing and our methodological approach, Section 4.4 presents the main

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

results and further exploratory analysis, and Section 4.5 concludes.

3.2 Methods

3.2.1 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.[4] The full 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 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 can be cheaply collected for a very large number of users, that collection is automatic 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.[5] Another major limitation is imperfect transaction labelling. Money Dasbhoard automatically classifies transaction into different categories based on transaction descriptions and merchant information. While this classification is quite mature and reliable for transactions with major retailers, it is imprecise and at times incorrect for less common transactions. We discuss this further in the next section.

3.2.2 Preprocessing and sample selection

We 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

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

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

repository and not run as part of this project.[6] Here, we briefly describe the main cleaning steps and their rationale. We drop all transactions with a missing description string because these cannot be categorised, and all transactions that are not automatically categorised by the app. Dropping these transactions makes it likely that we will underestimate amounts spent and saved, but minimises the risk of incorrectly classified transactions. We 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.[7] Finally, we 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.[8]

To minimise the influence of outliers, we 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] We 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.

The overall aim of the sample selection is to restrict our sample to users for whom we can observe a regular income, can be reasonably sure that they have added all their bank accounts to MDB, and for whom we observe at least six months of data. Table 4.1 summarises the sample selection steps we applied, associated data losses, and the size of our final sample.

We start by dropping the first and last month of data for each user because we are unlikely to observe users' complete data in these months, which might affect selection for the criteria below. Next, we keep only users whom we observe for at least 6 months to ensure that we have a suitable amount of data for each user. We also require users to have at least one current and one savings accounts to be able to observe their spending and savings behaviour. The next few steps are all aimed at selecting users about whom we can be reasonably sure that they have added all their financial accounts. We thus select on criteria we would

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

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

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

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

Table 3.1: 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.

expect to see for all such users: an annual income of at least £5,000, at least 10 monthly spending transaction, at least 4 monthly grocery transactions, and a total monthly spend of at least £200. Finally, we keep only users for whom we observe all relevant demographic characteristics, and who are between 18 and 65 years old. The reason for the last step is that younger users and retires will plausibly have different financial objectives than people in their working age, which is the population we are interested in.[10]

3.2.3 Dependent variables

Our outcome variable is a binary indicator for whether of not a user made a payment into any of their savings account in a given month. We classify as payments into savings accounts all savings account credits of £5 or more.

Such transactions will include those that are intended for purposes other than building to deal with unexpected financial shocks. For instance, users might transfer money to their savings accounts to save for a particular goal, such as new car, a holiday, or a wedding. While Money Dashboard allows users to create custom tags for transactions that might, in principle, be helpful to distinguish between these different types of savings transactions, only a very small number of transactions actually have such tags, so that we cannot use them to identify emergency savings more precisely. As a result, we refer to all observed transactions into savings accounts as emergency savings.

The reason we focus on a binary indicator is twofold: first, because we hypothesis that chaotic or difficult life circumstances that are also reflected in spending entropy might make it harder to remember to save, for which the accurate test is

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

to see whether spending entropy is related to the likelihood of making a savings transaction. Second, research by the UK's Money and Pension Service suggests that to build up sufficient emergency funds over time forming a habit of saving regularly is more important than the specific amounts saved month to month (MPS 2018).

3.2.4 Spending profiles

We define a user's spending profile as the distribution of the number of spending transactions 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 occurring with the logarithm of that probability.[11] The base of the logarithm is often chosen to be 2, though other choices are possible. 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 in a particular period as (we omit individual and time subscripts to keep the notation simpler):

$$H = -\sum_{c \in \mathcal{C}} p_c log(p_c), \tag{3.1}$$

where $\mathcal{C}$ is the set of all spending categories, p_c the probability that an individual makes a purchase in spending category c, and log the base 2 logarithm, which expresses entropy in bits.

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.[12] For simpler interpretation of our regression coefficients below, we standardise entropy scores to have a mean of 0 and a standard deviation of 1.

One 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

[11] 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 analogous.

[12] For further discussion on how to interpret Equation 3.1, see Appendix 3.A.

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.

We calculate entropy based on three sets of spend categories. The first measure is based on 9 spending categories used by Muggleton et al. (2020). The second measure is based on our own, more fine-grained, categorisation into 48 different categories.[13] The third measure is based on merchant names, as labelled by Money Dashboard.

We also calculate spending category probabilities in two different ways. To calculate what we call "unsmoothed" entropy scores, we calculate the p_cs in Equation 3.1 as simple frequentist probabilities

$$p_c = \frac{f_c}{F}, \tag{3.2}$$

where f_c is the number of transactions in spend category c (the frequency with which c occurs) and $F = \sum_{c \in \mathcal{C}} f_c$ the total number of spending transactions. To avoid taking the log of zero for categories with zero transactions, the sum in Equation 3.1 is taken over categories with positive transaction counts only.[14] To calculate "smoothed" entropy scores, we apply additive smoothing to calculate probabilities as

$$p_c^s = \frac{f_c + 1}{F + |\mathcal{C}|}, \tag{3.3}$$

where the size of set $\mathcal{C}$, $|\mathcal{C}|$, is the number of unique spending categories. Hence, we simply increment all frequency counts by one and then calculate probabilities as in Equation 3.2. Because categories with a zero transaction count will have a numerator of 1, the sum in Equation 3.1 will be taken over all categories.

3.2.5 Summary statistics

Table 4.2 provides summary statistics and Figure 3.1 distributions for the key variables used in our analysis as well as for some sample demographics. Income and spending distributions are being compared to those of the Living Cost and Food Survey, which is the ONS's main survey on household spending. We can see that the sample distributions broadly match the UK-wide ones but that we

[13]The precise mapping from MDB transaction tags into 9 and 48 categories is available on Github here and here, respectively. While we group all transaction tags available in the Money Dashboard dataset into 48 categories, there are no "mortgage release" transactions in our final analysis sample, so that, effectively, all calculations are based on 47 categories.

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

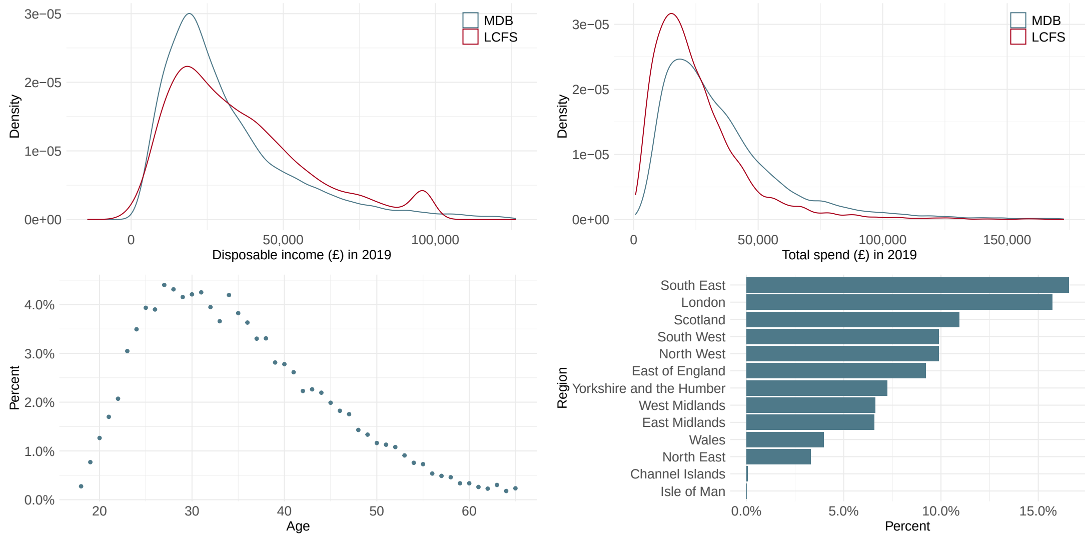
tend to underestimate incomes and overestimate spend. As discussed above, our sample is biased towards high-earning individuals, so that we would expect both income and spend to be higher than across the whole population. The main reason this is not the case is likely due to the imperfect identification of income transactions in the Money Dasbhoard data.

Table 3.2: Summary statistics

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Txn count	111.87	59.67	10	70	101	142	327
Month income (£)	2,853.49	2,495.24	0.00	1,407.21	2,217.37	3,595.92	15,027.54
Savings account inflows (£)	747.25	2,448.33	0.00	0.00	0.00	400.00	18,809.50
Savings account outflows (£)	762.40	2,422.94	0.00	0.00	0.00	400.00	18,099.38
Savings account netflows (£)	−7.39	2,883.70	−20,000.00	0.00	0.00	50.00	21,675.41
Month spend (£)	2,760.39	2,609.92	200.02	1,225.41	2,016.70	3,318.95	17,092.19
Age	37.46	9.97	18	30	36	44	65
Female dummy	0.44	0.50	0	0	0	1	1
Urban dummy	0.84	0.37	0	1	1	1	1
Active accounts	3.11	1.74	1	2	3	4	10
Discretionary spend (£)	860.63	736.08	0.00	369.34	662.98	1,118.28	4,181.71

Notes: Summary statistics for main variables used in analysis and user characteristics. Unique categories refer to the number of distinct spending categories that individuals made purchases in within a given month for spending transaction categorisation based on 9 categories, 48 categories, and merchant names.

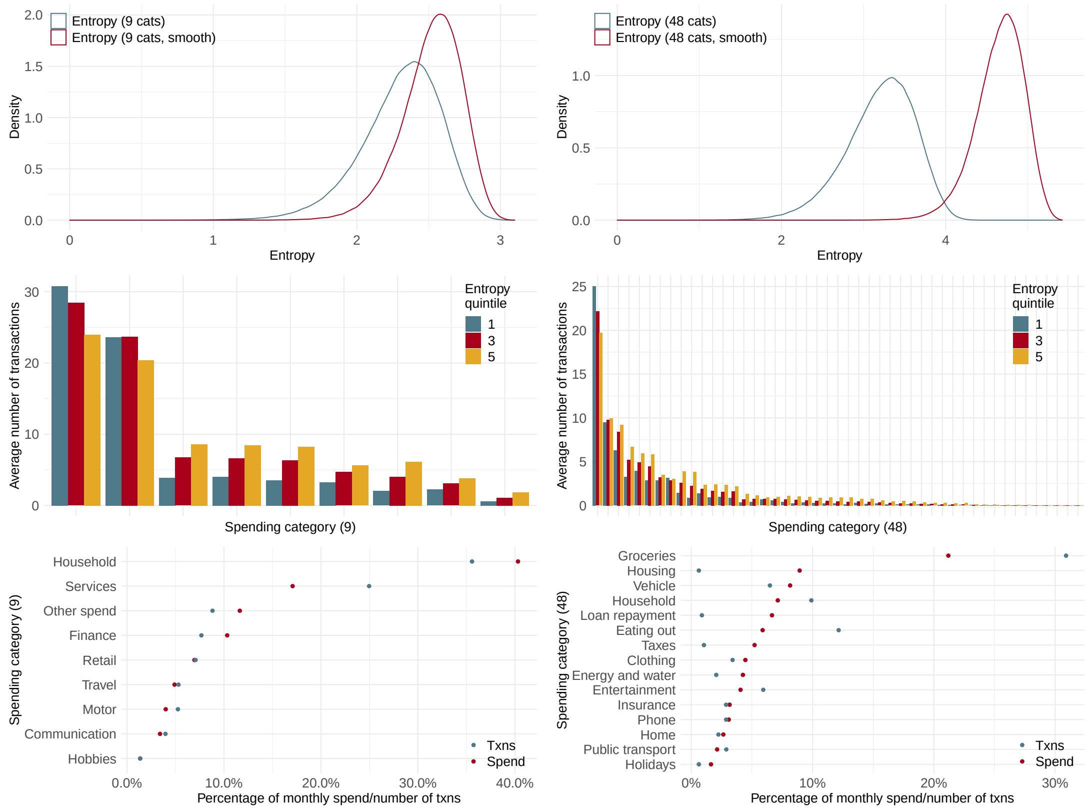
Figure 3.1: Demographic characteristics of Money Dashboard users



Notes: The top left and top right panels 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 bottom left panel shows the distribution of age, the bottom right panel that of the regions.

Individuals in our data make at least one savings transaction in half of all observed periods, and have at least some income in 98 percent of those periods. As discussed above, our sample is skewed towards men in their 20s, 30s, and 40s, who live in urban areas, mainly the South East of England. Mean monthly spend is £2,000 and the median is £2,200, and the average individual spends money across 27 different merchants in 17 different categories (of the 48 category classification).

Figure 3.2: Entropy distributions



Notes: The top row shows distributions of unsmoothed and smoothed versions of entropy. The middle row shows associated spending profiles for individuals in the first, third, and fifth quintile of the entropy distribution. Spending category labels are deliberately omitted to not overcrowd the plots and to focus on the distribution of transaction counts. The bottom row shows the distributions of spend and number of transactions that goes to the top categories. The plot on the left shows all 9 categories, that on the right the top 15 out of 48 categories. The left column shows data for the 9 category classification, the right column data for the 48 categories classification.

Figure 3.2 shows the distributions of the unsmoothed and smoothed version of entropy based on 9 and 48 spending categories, in the top row, the associated average spending profiles for individuals in the first, third and fifth quintiles in the middle row, and an overview of the main categories individuals spend on in the bottom row. The entropy distributions are bounded on the right by the maximum entropy value, given by the value for a discrete uniform distribution with 9 or 48 distinct values,[15] and on the left by zero, the value for cases where an individual makes all spending transactions in a single category. The different effect of smoothing on the two variables is a result of the different number of categories with counts of zero.

The spending profiles in the middle row show that higher entropy individuals make both fewer transactions in high-count categories and more transactions in low-count categories. The spend profile based on the 48 category classification also shows that spend counts follow approximately a power law distributions, with very few of the 48 categories accounting for the large majority of transactions.

[15] For instance, the upper bound for the 9 category variable is $9 \times \left(\frac{1}{9}\right) log_2 \left(\frac{1}{9}\right) \approx 3.17$.

The relative ordering of the top spending categories based on the 48 category classification is about as expected, as are the relationship between the number of transactions and spend: there are, for instance, a large number of relatively low value transactions and meals eaten out (which includes snacks and coffee) and a very small number of high value transactions for housing. The proportion of spend that goes towards housing, less than 10%, is clearly an underestimate, however, and reflects the challenge and imperfect nature of automatic transaction labelling in our data that we discussed above.

3.2.6 Estimation

To estimate the relationship between spending entropy and savings behaviour we estimate fixed-effect models of the form:

$$y_{i,t} = \alpha_i + \lambda_t + \beta H_{i,t} + x'_{i,t}\delta + \epsilon_{i,t}, \tag{3.4}$$

where $y_{i,t}$ is an indicator variable equal to one if individual i made at least one 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, α_i an individual fixed effect, λ_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 3.2.3 and Section 3.2.4, 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, such a transaction will usually be labelled in their current account as a transfer rather than a spending transaction, and thus not enter the calculation of their entropy score.

3.3 Spending profiles predict emergency savings

Table 3.3 shows the effect of entropy on the probability of building emergency savings in a given month. Columns (1)-(3) show results for unsmoothed entropy based on 9 categories, 48 categories, and merchant names, respectively. Columns

(4)-(6) show 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. 95% confidence intervals are shown in brackets.[16]

Results for unsmoothed entropy suggest that a one standard-deviation increase in spending entropy is associated with an increase in the probability of a user making at least one transfer into their savings accounts of between 1.1 and 2.3 percentage points – an effect equal to an increase in monthly income of between £1000 and £2000. Combined, the results in the three columns present very strong evidence that entropy captures something about users's spending distribution that is related to their likelihood for making payments into their savings accounts. Furthermore, entropy variables defined based on a larger number of unique categories, that allow for a more precise segmentation of spending behaviour, capture features of spending behaviour that are as strongly or more strongly related to savings behaviour as monthly income.

The results for smoothed entropy are similar but a tend to be smaller in magnitude, and – together – also provide strong evidence that smoothed entropy captures features of spending behaviour that is related to savings behaviour in a statistically significant and economically relevant way. The main difference to the results for unsmoothed entropy is the reversal of the sign of the coefficients: across all three measures, an increase in entropy is estimated to be associated with a decrease in the likelihood of any savings transactions. The effect of the measure based on 9 categories is not meaningfully different from zero, but the estimates for the measures based on 48 categories and merchants are almost identical and suggest that the magnitude of the effect is between 1.2 and 2.1 percentage points – equalling the effect of a reduction of monthly income of between £1000 and £2000.

[16] In line with an emerging consensus for how to avoid over-reliance on statistical significance, we deliberately do not show significance stars in our regression tables and instead emphasise the implications of the bounds of the confidence intervals when discussing our results. For two recent discussions, see Imbens (2021) and Romer (2020).

Table 3.3: Effect of entropy on P(savings transactions)

	(1)	(2)	(3)	(4)	(5)	(6)
Entropy (9 cats)	0.011					
	[0.005; 0.017]					
Entropy (48 cats)		0.023				
		[0.017; 0.029]				
Entropy (merchant)			0.017			
			[0.011; 0.023]			
Entropy (9 cats, smooth)				-0.004		
				[-0.007; -0.000]		
Entropy (48 cats, smooth)					-0.017	
					[-0.021; -0.012]	
Entropy (merchant, smooth)						-0.017
						[-0.021; -0.013]
Month spend (£'000)	0.010	0.009	0.010	0.010	0.009	0.008
	[0.009; 0.011]	[0.008; 0.011]	[0.008; 0.011]	[0.009; 0.011]	[0.008; 0.010]	[0.006; 0.009]
Month income (£'000)	0.012	0.012	0.012	0.012	0.012	0.012
	[0.011; 0.014]	[0.011; 0.014]	[0.011; 0.014]	[0.011; 0.014]	[0.010; 0.014]	[0.010; 0.013]
Has income in month	0.086	0.085	0.086	0.087	0.086	0.086
	[0.070; 0.102]	[0.069; 0.101]	[0.069; 0.102]	[0.071; 0.103]	[0.070; 0.102]	[0.070; 0.102]
Income variability	0.001	0.001	0.001	0.001	0.001	0.001
	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]
Observations	389,790	389,790	389,790	389,790	389,790	389,790
R^2	0.46824	0.46838	0.46830	0.46820	0.46844	0.46870
Within R^2	0.00621	0.00648	0.00633	0.00615	0.00659	0.00709
User fixed effects	✓	✓	✓	✓	✓	✓
Year-month fixed effects	✓	✓	✓	✓	✓	✓

Notes: Results from estimating Equation 3.4. The dependent variable in all columns is a dummy variable indicating whether a user made at least one transaction into any of their savings accounts in a given period. Terms in brackets denote the upper and lower ends of a 95% confidence interval. Standard errors are clustered at the user-level.

As discussed in Section 3.2.6, these results are not due to reverse causality. While we might think that making a savings transactions might change some or all of the components of entropy discussed in Section 3.2.4 – the number of unique spending categories with positive frequency count, the standard deviation of these counts, and the total number of spend transactions – and thus change entropy, this is not the case because of the way we define entropy and savings, and the way spending transactions are categorised. We define entropy based on all current account debits that are identified as spends, while we define savings transactions as the sum of all savings accounts credits. If a user transfers money from their current account to their savings account, this will be identified as a savings transaction, but be identified as a transfer on their current account and thus not considered when calculating their entropy score.

The estimates of our control variables are largely as expected, with the exception of monthly spend, which one might have expected to be negatively correlated with savings. Also, it is evident that the strongest predictor among the included controls for whether a user makes any savings transfer is whether they receive any income in that month. Income variability, in contrast, is not correlated with savings behaviour in any economically significant way, suggesting that people with variable incomes do not build savings cushions for periods where they have no income.

Overall, 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. The obvious question raised by the results is why smoothing entropy scores flips the direction of the effect of entropy. We address this next.

3.3.1 Why does smoothing flip the direction of the effect

One way to think about the sign change in Table 3.3 is to realise that it implies that at least for some individuals, the relative rank of smoothed and unsmoothed entropy must differ considerably – there must be some individuals that have low unsmoothed entropy but high smoothed entropy or some that have high unsmoothed entropy and low smoothed entropy or both. Understanding who those individuals are might thus help us understand the sign flip.

To understand rank differences between unsmoothed and smoothed entropy scores it is useful to rewrite Equation 3.1 in a way that makes it easy to see its component parts. Remember from Section 3.2.4 that f_c is the number of transactions made by a user in a given period in spending category c, $\mathcal{C}$ the set of all spending categories, and F the total number of transactions made by a given user in a given period. Additionally, let $\mathcal{C}^+ = \{c : f_c > 0\}$ be the set of all spending categories with positive frequency counts (i.e. with at least one transaction) and $\mathcal{C}^0 = \{c : f_c = 0\}$ the set of all spending categories with a zero frequency count, so that $\mathcal{C} = \mathcal{C}^0 \cup \mathcal{C}^+$. Then, using our definitions of unsmoothed and smoothed probabilities, we can write unsmoothed entropy as

$$H = -\sum_{c \in \mathcal{C}^+} \left(\frac{f_c}{F}\right) log \left(\frac{f_c}{F}\right), \tag{3.5}$$

and smoothed entropy as:

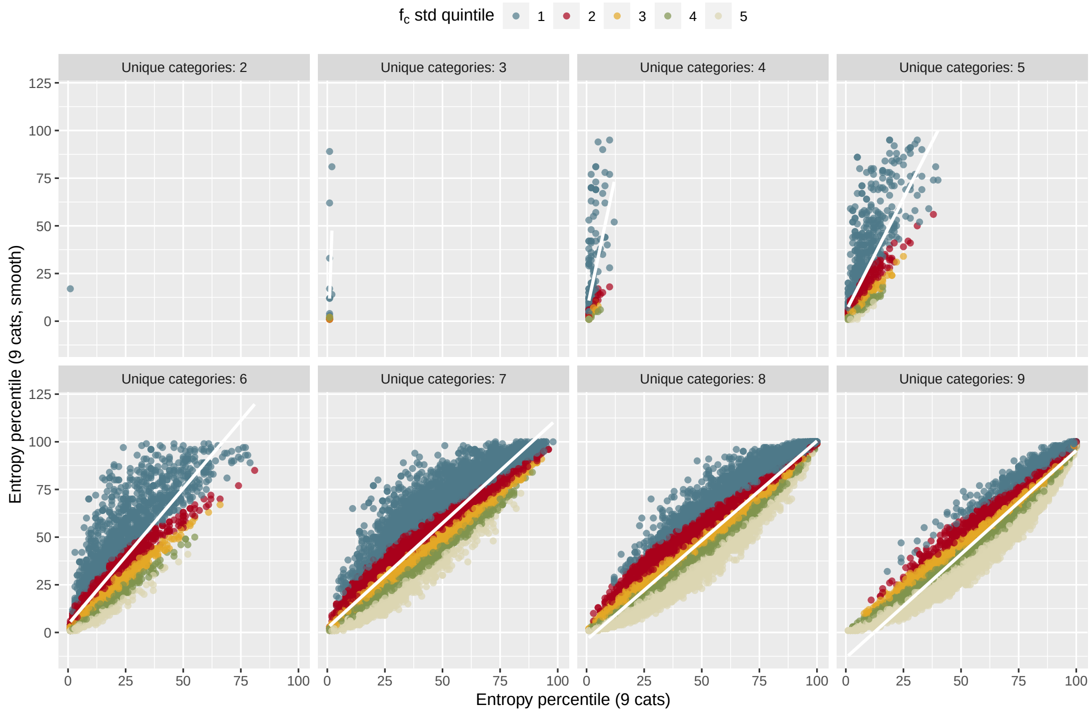
$$H^s = -\sum_{c \in \mathcal{C}^+} \left(\frac{f_c + 1}{F + |\mathcal{C}|}\right) log \left(\frac{f_c + 1}{F + |\mathcal{C}|}\right) - |\mathcal{C}^0| \left(\frac{1}{F + |\mathcal{C}|}\right) log \left(\frac{1}{F + |\mathcal{C}|}\right), \tag{3.6}$$

where the size of set $\mathcal{C}^0$, $|\mathcal{C}^0|$, is the number of all spending categories in which a user makes no transactions in the period. These expressions make clear that, by definition, unsmoothed entropy is a function of frequency counts of categories with positive counts only while smoothed entropy has two parts: the sum over all additively smoothed frequency counts of categories with positive counts, plus the same sum for additively smoothed probabilities of categories with zero counts, which reduces to a constant term that is multiplied by the number of such categories.

The expressions also make transparent the three main components of both types of entropy that are determined by user behaviour. The first is the number of spending categories with a non-zero frequency count, $|\mathcal{C}^+|$, which determines the number of elements summed over in Equation 3.5, and partitions the categories into either contributing to the sum on the left hand side of Equation 3.6 or – given that for an exogenously fixed $|\mathcal{C}|$, $|\mathcal{C}^0| = |\mathcal{C} \backslash \mathcal{C}^+|$ – to the constant term on the right hand side. The second component is the variation of the frequency counts, f_c, and the third component is the number of total spend transactions, $F = \sum_{\mathcal{C}} f_c$. Together, these two elements determine the probabilities associated with a spend transaction occurring in a given category. The number of total spending categories, $|\mathcal{C}|$, also determines smoothed entropy and, implicitly, also unsmoothed entropy since it "scales" the number of categories with a positive frequency count, $|\mathcal{C}^+|$, as a given number of spending transactions are categorised into finer or coarser categories. But it is exogenously given and does not depend on user behaviour.

These components help us reason about rank differences between unsmoothed and smoothed entropy. From equations 3.5 and 3.6 we can see that the first part of smoothed entropy that sums over all spending categories with positive fre-

Figure 3.3: Effect of smoothing on entropy



Notes: Percentile ranks of 9-category-based unsmoothed and smoothed entropy separated by the number of categories with positive frequency counts. White reference lines indicate equal percentile ranks; colours, frequency count standard deviation quintiles. Figure is based on a 10% sample of the dataset used for analysis. There are only 8 panels since there are cases where a user has spends in only a single category.

quency counts is very similar to the entire expression of unsmoothed entropy – it is that same expression but with additively smoothed probabilities. Hence, all else equal, the higher the number of categories with positive counts, the more smoothed entropy is determined by that first part, and the more similar it will be to unsmoothed entropy. As a result, we would expect to find large (rank) differences between entropy scores among cases with few positive-counts categories. Furthermore, given that unsmoothed entropy is increasing in the number of categories with positive counts, we expect these cases to have relatively low unsmoothed entropy.

Given that we expect high rank difference cases to make spends across a small number of categories, and expect these cases to have low unsmoothed entropy, high rank differences will occur among the subset of those cases that have high smoothed entropy. Remember from Section 3.2.4 that entropy is higher the more equal the spending category probabilities are. Hence, for a given number of zero-count categories, smoothed entropy will be higher if the (additively smoothed) probabilities of all positive-count categories are close to the (additively smoothed) probabilities of the zero-count categories, which will be the case (i) if there are few overall transactions, such that frequency counts (f_c) are close to zero and (ii) if the counts are similar.

Figure 3.3 visualises this entire line of reasoning for our 9-category based entropy variable: it shows scatter plots of the percentile ranks of unsmoothed and smoothed entropy with a reference line indicating identical rank, separated by the number of categories with positive frequency counts and coloured based on the quintile of the frequency count standard deviation.[17] First, ignoring the colouring and focusing on the shape of the dots only we can see that, as expected, the relationship between the two entropy measures is tighter the higher the number of non-zero spending categories is. Cases with large entropy rank differences are thus to be found among cases with fewer positive spend categories. Among these, the colouring makes clear that, as expected, the cases with the largest rank differences – those with the furthest vertical difference to the reference line – have low variation in their frequency counts. However, it is also clear that the reverse is not true: there are cases with low count frequency variation that experience little or even a negative rank difference. Furthermore, among cases with a higher number of categories with positive counts, cases with a very high variation in frequency counts can also experience quite large rank differences, albeit in the opposite direction. Hence, while counts variation is some help in identifying cases with high entropy rank differences, it does not do so perfectly.

One reason the relationship is not perfect is that while cases with few total transactions that are evenly spread across a small number of categories will have high smoothed entropy, such will also have high unsmoothed entropy. What we are looking for, more precisely, are cases where this pattern holds broadly, but that also have a small number of frequency counts that dominate the others but are still relatively close to zero. This would lead to low unsmoothed entropy (because of the dominant counts) but high smoothed entropy (because all counts are still relatively close to zero, so that the additively smoothed distribution resembles a uniform distribution). Further research will be necessary to more fully classify such cases, and to inquire what it is about their behaviour that might be related to savings behaviour.

3.3.2 Is entropy informative beyond its component parts?

Another question that arises once we think of entropy as comprising three main component parts is whether its non-linear nature captures anything about the relationship between spending and saving behaviour that is not captured already by the three simpler components.

[17] Colouring based on the quintile of the total number of transactions leads to a very similar plot and is thus not shown.

Table 3.4: Controlling for components

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Entropy (9 cats)	0.011	0.004						
	[0.005; 0.017]	[-0.003; 0.011]						
Entropy (9 cats, smooth)			-0.004	-0.004				
			[-0.007; -0.000]	[-0.009; 0.001]				
Entropy (48 cats)					0.023	0.020		
					[0.017; 0.029]	[0.005; 0.034]		
Entropy (48 cats, smooth)							-0.017	-0.021
							[-0.021; -0.012]	[-0.031; -0.010]
Unique categories ($\lvert\mathcal{C}^+\rvert$)		0.004		0.004		0.002		0.004
		[0.002; 0.005]		[0.003; 0.005]		[0.001; 0.004]		[0.003; 0.005]
Category counts std.		-0.001		-0.003		0.004		-0.010
		[-0.005; 0.003]		[-0.008; 0.001]		[-0.002; 0.010]		[-0.016; -0.005]
Number of spend txns (F)		0.001		0.001		0.000		0.001
		[0.000; 0.001]		[0.000; 0.001]		[-0.000; 0.001]		[0.000; 0.001]
Month spend	0.010	0.006	0.010	0.006	0.009	0.006	0.009	0.006
	[0.009; 0.011]	[0.005; 0.007]	[0.009; 0.011]	[0.005; 0.007]	[0.008; 0.011]	[0.005; 0.007]	[0.008; 0.010]	[0.005; 0.007]
Month income	0.012	0.011	0.012	0.011	0.012	0.011	0.012	0.011
	[0.011; 0.014]	[0.009; 0.013]	[0.011; 0.014]	[0.009; 0.013]	[0.011; 0.014]	[0.009; 0.013]	[0.010; 0.014]	[0.009; 0.013]
Has income in month	0.086	0.082	0.087	0.082	0.085	0.081	0.086	0.081
	[0.070; 0.102]	[0.066; 0.098]	[0.071; 0.103]	[0.066; 0.098]	[0.069; 0.101]	[0.065; 0.097]	[0.070; 0.102]	[0.065; 0.097]
Income variability	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]	[0.000; 0.002]
Observations	389,790	389,790	389,790	389,790	389,790	389,790	389,790	389,790
R^2	0.46824	0.46908	0.46820	0.46908	0.46838	0.46910	0.46844	0.46914
Within R^2	0.00621	0.00778	0.00615	0.00779	0.00648	0.00783	0.00659	0.00790
User fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Year-month fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Terms in brackets denote the upper and lower ends of a 95% confidence interval. Standard errors are clustered at the user-level.

We can test this by checking whether the relationship between spending entropy and the probability of making a savings transaction remains economically and statistically significant once we control for the three components. Columns (1) and (3) in Table 3.4 replicate the results for the 48-category-based unsmoothed and smoothed entropy measures presented in Table 3.3 for reference. In columns (2) and (4) we additionally control for the three entropy components. Including these components has some effect: the coefficients change slightly – decreasing in absolute magnitude in the case of unsmoothed entropy, increasing in the case of smoothed entropy – while the width of the confidence intervals about double in both cases, reflecting the strong collinearity amount the component and entropy. However, both coefficients remain statistically significant and their confidence intervals cover values that are also economically significant. Hence, the results make clear that the results in Table 3.3 cannot be attributed simply to the effect of one or more of entropy's simple components.

3.4 Conclusion

In this paper, we calculate simple summary statistics for spending profiles for a large number of users of a UK-based financial aggregator app and show that the predictability of spending, captured using entropy, is predictive of the frequency with which users make transactions into their savings accounts. The effect of spending entropy on savings is statistically significant and economically large, and robust across a number of different specifications and entropy measures based on different product and merchant categories. The direction of the effect depends, however, on the handling of categories in which individuals make no transactions in a given month. While our results give some indication as to how we can explain this result, further research is needed.

As part of a fuller understanding of said sign reversal, more work generally is needed on understanding to what extend entropy is useful in capturing meaningful information about individuals' lives, and what particular ways of calculating entropy are most appropriate and informative in different circumstances. For instance, while, in this paper, we have calculated entropy based on the number of transactions in a given category, there are a number of alternatives. 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).

Finally, work like ours inherently risks portraying low savings mainly as a result of behavioural biases, implying that if can help people overcome these biases, they will be able to save more. However, as convincingly argued in Chater and Loewenstein (2022), for problems where the main solution lies not in individual behaviour change but public policies, such an approach runs the risk of moving attention and resources away from required society-wide action. Research by the Money and Pension Service in the UK (MPS 2018) and the Consumer Financial Protection Bureau (CFPB 2017) suggests that while low levels of savings are – unsurprisingly – often concentrated among very low income individuals or households, individual behaviour still plays a large role in explaining differences in savings levels across income groups. Hence, addressing low levels of savings is a matter for both public policy and individual behaviour change. But it is important to ensure that focusing on the latter does not distract from the former.

Appendix 3.A Interpreting entropy

To see how we can interpret entropy as the predictability of a user's spendin that we discussed above behaviour, it is useful to have a more complete understanding of Equation 3.1. 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), Shannon proposed using the log of the expression. Hence, the information of event E, often called *Shannon information*, *self-information*, or just *information*, is defined as:

$$I(E) = log\left(\frac{1}{p(E)}\right) = -log(p(E)). \tag{3.7}$$

The choice of the base for the logarithm varies by application and determines the units: base 2 means that information is expressed in bits; the natural logarithm, another popular choice, expresses information in *nats*.

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:

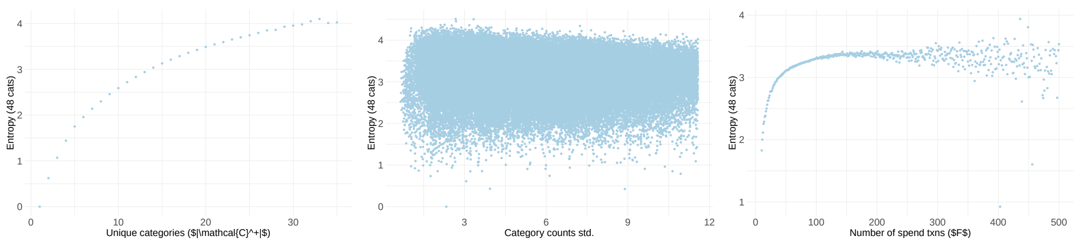
$$H(X) = -\sum_x p(x) \times log(p(x)) = \sum_x p(x)I(x) = \mathbb{E}I(x). \tag{3.8}$$

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.

Appendix 3.B Entropy components

Figure 3.4 shows the empirical relationship with our 48-categories-based unsmoothed entropy variable and these three components.[18] We can see that for the values we observe in the dataset, entropy increases monotonically in the number of unique spending categories with positive frequency counts, has no clear relationship with the standard deviation of those counts, and increases in the number of total spending transactions up to about 175 transaction, before being increasingly determined by other elements thereafter.

Figure 3.4: Correlation of entropy with its components

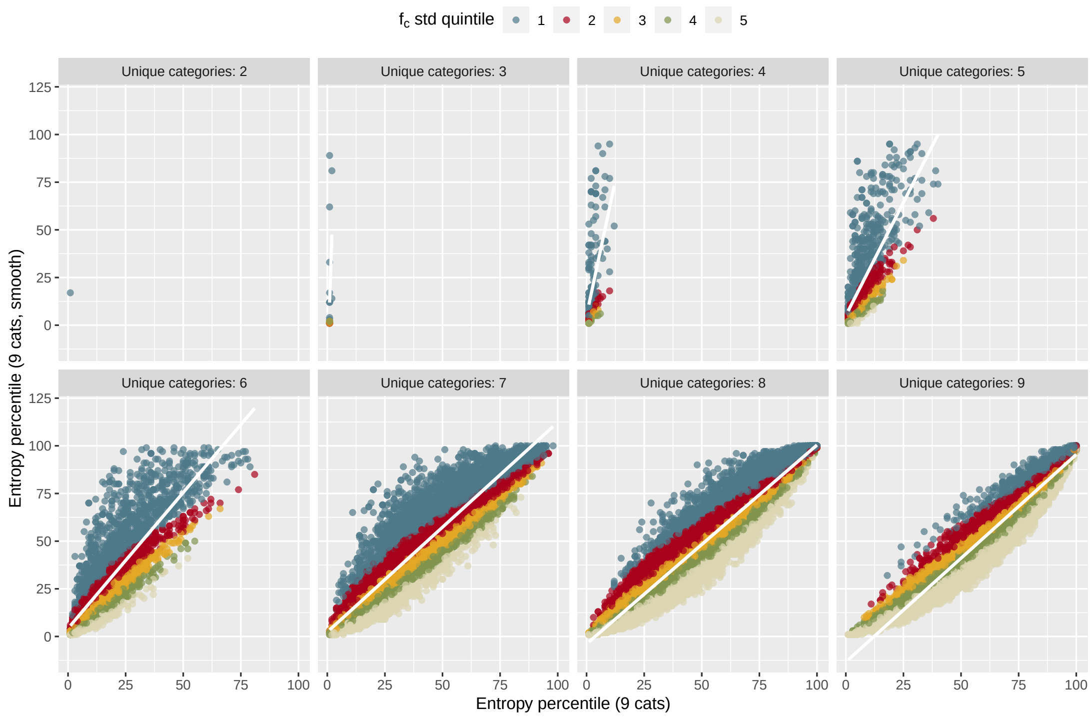


Notes: Correlation of 48-categories-based unsmoothed entropy with its three main components: the number of unique spending categories with positive frequency counts (left), the standard deviation of those frequency counts (middle), and the number of total spend transactions (left).

Appendix 3.C Effect of smoothing on entropy

[18] To highlight the main features of the relationships we have trimmed the component values at the 95th percentile.

Figure 3.5: Effect of smoothing on entropy



Notes: Percentile ranks of 9-category-based unsmoothed and smoothed entropy separated by the number of categories with positive frequency counts. White reference lines indicate equal percentile ranks. Colours indicate the quintile of the total number of spending transactions.

Chapter 4

Do Money Dashboard users spend more and save less?

4.1 Introduction

This paper evaluates whether using Money Dashboard, a UK-based financial aggregator app, is associated with a reduction in discretionary spend and an increase in *emergency savings* – short-term savings intended as a buffer against unexpected financial shocks.

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 (Phillips 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 makes it easier to monitor one's 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

[1]Well-documented behavioural biases that help explain undersaving are, among others, present bias (Laibson 1997, Ericson and Laibson 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.

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 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]

In this paper, I specifically test whether using Money Dashboard is associated with a reduction in discretionary spending and an increase in emergency 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 per month (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 emergency 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. This suggests that users either leave unspent funds to accrue in their current accounts (which would not count to my definition of savings), of allocate them across a number of different uses in small amounts I am not powered to detect.

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

[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.

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 *net effect* = *causal effect* + *"need"*($\downarrow$) + *"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 – 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 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.

My work mainly contributes to three strands of the literature. The first is the nascent literature that studies the effect of FinTech apps on financial outcomes suggests. This literature suggests that these apps can indeed lead to improved financial outcomes through a number of different channels: providing users with information about their spending relative to peers has been found to reduce spending (D'Acunto et al. 2020); offering budgeting options, to reduce spending (Lukas and Howard 2022); goal setting, to increase savings (Gargano and Rossi 2021); and facilitating access to financial information, to reduce non-sufficient funds fees (Carlin et al. 2022) and discretionary spend (Levi and Benartzi 2020).

The two studies that are most closely related to my work are Lukas and Howard (2022) and Gargano and Rossi (2021). Lukas and Howard (2022) also use data from Money Dashboard to study the effect of budgeting on discretionary spend. In line with my finding, they find that budgeting reduces discretionary spend in the associated categories and that this effect persists throughout the six-month post-budgeting period they study. Their approach also suffers from the two main limitations of my paper discussed above: budgeting is not randomly assigned, and they cannot isolate the effect of budgeting from other MDB features that might help users reduce their spending. There are two main differences to my work: first, Lukas and Howard (2022) do not study the effect of MDB use on savings. Second, they use data from between January 2014 and December 2016, and do not select their sample to ensure that it only contains users for whom they can observe all financial accounts. As I discuss in Section 4.2.2, there is a large number of users in the raw dataset for whom we are unlikely to observe all transactions, making careful sample selection critical to ensure that observed changes in behaviour do not merely reflect a shift of transactions between observed and unobserved accounts. Gargano and Rossi (2021) study the effect goal setting on savings and find that setting savings goals does increase savings rates. Their approach does not suffer from the two limitations of my paper since they exploit the random assignment of some users into a group of beta testers that can set savings goals, and because the savings app that provides the data for their study was initially designed as a simple savings app and provides no additional features that could plausibly influence savings behaviour. However, the flip-side of being able to cleanly identify the effect of goal setting is that most apps that people use to increase their savings or reduce their spending do – like Money Dashboard – have multiple features that might alter financial behaviour, so that studying their joint effect is of interest, too.

The second strand of related literature is the very recent literature on studying interventions to help increase emergency 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, Iwry, et al. 2020, Berk et al. 2022).

Finally, my 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]

The remainder of this paper is organised as follows: Section 4.2 introduces the dataset used, discusses preprocessing and presents summary statistics; Section 4.3 introduces the empirical approach used in the analysis; Section 4.4 presents the results; and Section 4.5 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]

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

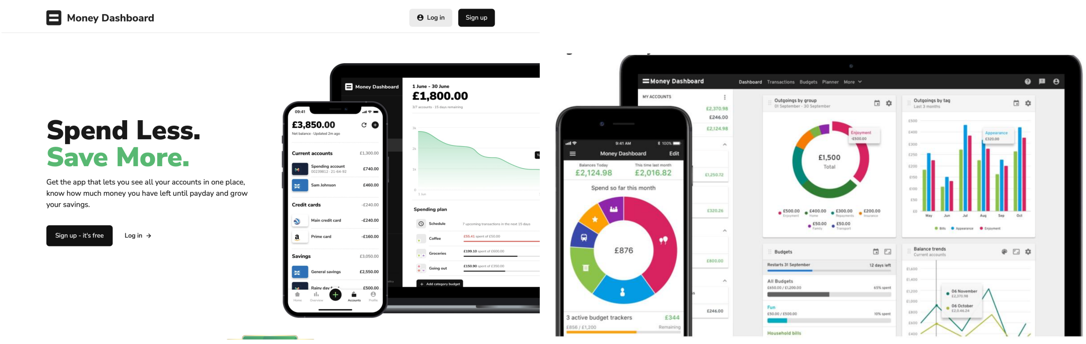
[4]The projects GitHub repo that contains all files used to produce the results can be found at https://github.com/fabiangunzinger/mdb_eval.

4.2 Data

4.2.1 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 4.1 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.

Figure 4.1: Money Dashboard website and user interface



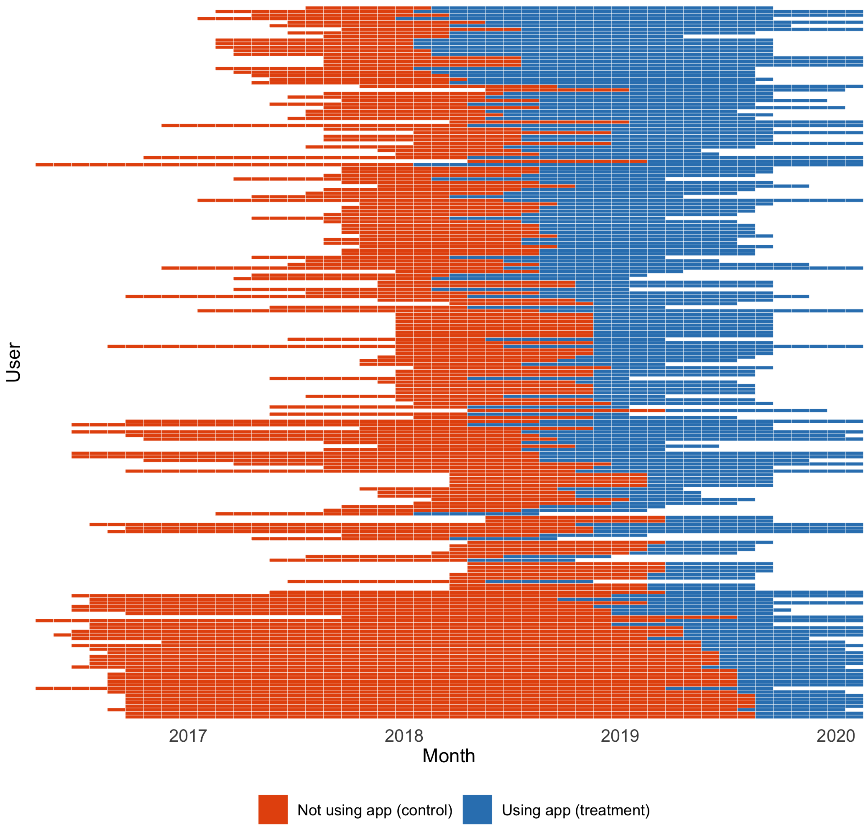
Notes: Screenshot from the top of the Money Dashboard website (left) and example of its user interface (right), illustrating some of the transaction grouping and budgeting functions.

Crucially, for this paper, MDB can access up to three years of historic data for each linked account. Figure 4.2 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 occurred 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 unlinked accounts.[5] Furthermore, data is collected automatically and

[5]For an example where debt-reduction in one set of accounts is offset by an increase in debt

Figure 4.2: 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 users. To the left of the observed period, this is because the app cannot access data before that point; to the right, because they have stopped using the app.

in real-time rather than through surveys that collect data with a time-lag and often rely on consumer'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, younger, 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 we can observe user's complete financial behaviour if they add all their financial accounts to the app, it is not

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 et al. (2020).

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 misclassified some transactions and cannot classify others altogether. I address this as part of the cleaning process documented below.

4.2.2 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 not run as part of this project.[7] Here, I briefly describe the main cleaning steps and their rationale. I drop the fewer than 1 in 1000 transactions with a missing description string because these cannot be categorised. 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 4.1 lists the precise conditions I apply to implement these criteria and their effect on sample size.[11]

[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.

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 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 who 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 after-tax 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 – users who signed up to the app before it was publicly available – 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. In fact, while it is not surprising that only considering users that signed up after March 2017 would reduce the sample size considerably, the large reduction in users due to requiring at least one savings account, an annual after-tax income of at least £5,000, and at least 10 monthly transactions are surprising, and suggest that the raw dataset contains a large number of users that do not use Money Dashboard as intended. The drastic effect of the selection procedure on sample size also underlines how important careful sample selection is to ensure that we can be reasonably sure to observe all relevant financial transactions for all users in the sample, and thus to be confident that observed changes in financial behaviour do not reflect a shift of transactions between observable and unobservable accounts.

[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.

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

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

Table 4.1: 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.

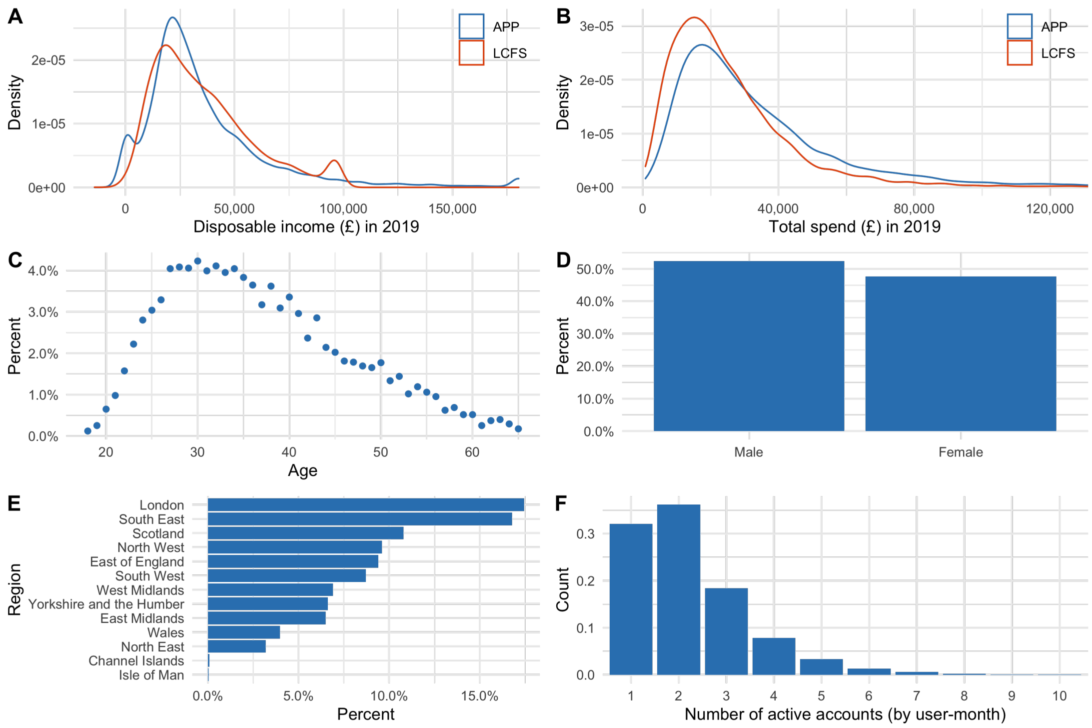
4.2.3 Summary statistics

Figure 4.3 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 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 4.2.2, 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 4.2.1, 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.

[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`.

[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.

Figure 4.3: 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 4.2 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 discretionary spend accounts for about 30 percent of total monthly spend

Table 4.2: Summary statistics

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Txn count	111.87	59.67	10	70	101	142	327
Month income (£)	2,853.49	2,495.24	0.00	1,407.21	2,217.37	3,595.92	15,027.54
Savings account inflows (£)	747.25	2,448.33	0.00	0.00	0.00	400.00	18,809.50
Savings account outflows (£)	762.40	2,422.94	0.00	0.00	0.00	400.00	18,099.38
Savings account netflows (£)	−7.39	2,883.70	−20,000.00	0.00	0.00	50.00	21,675.41
Month spend (£)	2,760.39	2,609.92	200.02	1,225.41	2,016.70	3,318.95	17,092.19
Age	37.46	9.97	18	30	36	44	65
Female dummy	0.44	0.50	0	0	0	1	1
Urban dummy	0.84	0.37	0	1	1	1	1
Active accounts	3.11	1.74	1	2	3	4	10
Discretionary spend (£)	860.63	736.08	0.00	369.34	662.98	1,118.28	4,181.71

4.3 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 recent 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 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 a generalised version of the average treatment effect of the treated (ATT) of the difference-in-differences framework, which the authors call the group-time average treatment effect: the average treatment effect at time t for the group (or cohort) of individuals first treated at time g, defined as:

[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'.

$$ATT(g,t) = \mathbb{E}[Y_{i,t}(g) - Y_{i,t}(0)|G_i = g], \tag{4.1}$$

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. G_i indicates the first period in which individual i was first exposed to the treatment.

A key feature of these group-time average treatment effects is that they allow for heterogeneous treatment effects both across time and across groups. To see this, first consider, for example, the group of individuals first treated in April 2019 (i.e. G_i = *2019-04*, for all individuals in the group) at two different points in time, June and July 2019, say, and notice that Equation 4.1 allows for $ATT(\textit{2019-04}, \textit{2019-06}) \neq ATT(\textit{2019-04}, \textit{2019-07})$: hence, $ATT(g,t)$ allows for heterogeneity across time. Similarly, consider also another group of individuals first treated in March 2018, and notice that Equation 4.1 allows for $ATT(\textit{2019-04}, \textit{2019-06}) \neq ATT(\textit{2018-03}, \textit{2019-06})$: hence, $ATT(g,t)$ allows for heterogeneity across groups.

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]

[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

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, g', that is not yet treated at time t:

$$ATT(g,t) = \mathbb{E}[Y_{i,t} - Y_{i,g-1} | G_i = g] - \mathbb{E}[Y_{i,t} - Y_{i,g-1} | G_i = g'], \text{ for any } g' > \text{t}. \quad (4.2)$$

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

$$ATT(g,t) = \mathbb{E}[Y_{i,t} - Y_{i,g-1} | G_i = g] - \mathbb{E}[Y_{i,t} - Y_{i,g-1} | G_i \in \mathcal{G}]. \quad (4.3)$$

This equation encapsulates two main results:[21] 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.[22]

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,

$$\widehat{ATT(g,t)} = \frac{1}{N_g} \sum_{i:G_i=g} [Y_{i,t} - Y_{i,g-1}] - \frac{1}{N_{\mathcal{G}}} \sum_{i:G_i \in \mathcal{G}} [Y_{i,t} - Y_{i,g-1}], \quad (4.4)$$

where N_g is the number of individuals first treated at time g (i.e. members of group g), and $N_{\mathcal{G}}$ is the number of individuals not yet treated at time t (i.e. all members of all groups in $\mathcal{G}$).

Once these building blocks are estimated, aggregating them to event-study type treatment effects that provide the (weighted) average treatment effect l

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.

[21] The results are formally proven in Section 2 of Callaway and Sant'Anna (2021).

[22] 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.

periods away from treatment adoption across different adoption groups can be achieved by calculating

$$ATT_l = \sum_g w_g ATT(g, g+l), \tag{4.5}$$

where w_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.[23]

Inference: Inference is based on a bootstrap procedure that can account for clustering and produces simultaneous (or uniform) confidence bands that account for multiple hypothesis testing.[24]

4.4 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.[25]

4.4.1 Main results

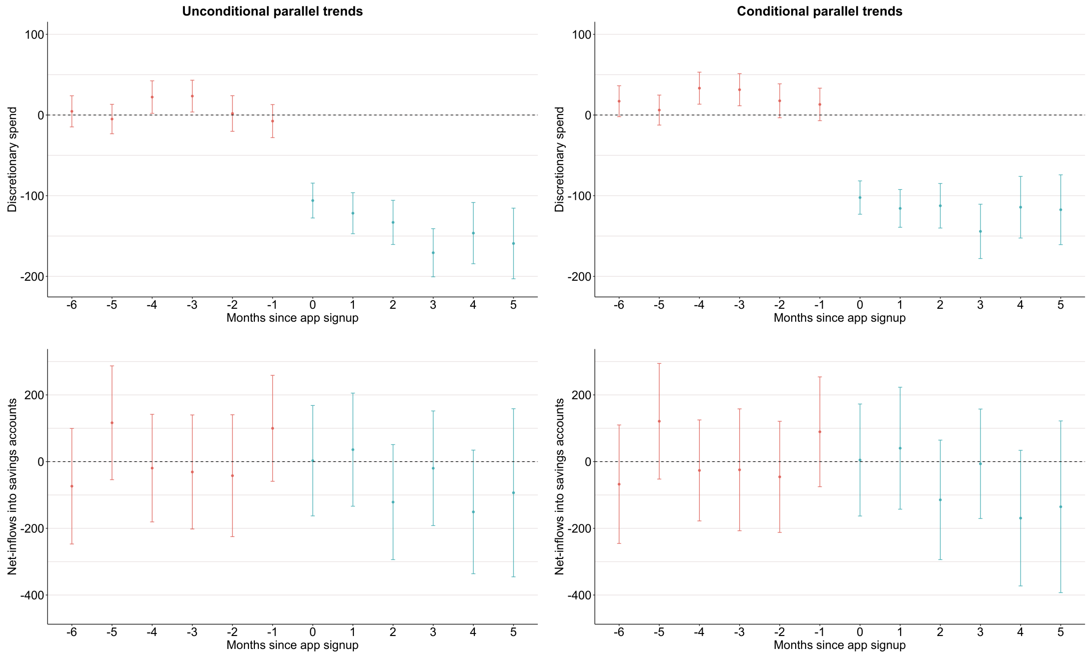
Figure 4.4 shows the effect of app use on monthly discretionary spend (top row) and monthly net-inflows into savings accounts (bottom row) under the unconditional (left column) and conditional (right column) parallel trends assumptions. As discussed in Section 4.3, the conditional parallel trends assumption states

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

[24] For more details on the inference procedure, 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.

[25] Discretionary spend is calculated solely based on card transactions, 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 4.4: 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 4.3. 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 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 4.5. 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 4.3.

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 4.2), 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.

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 who 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.[26]

We can also see that in periods -4 and -3, 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. Confidence intervals are much larger than for discretionary spend because – as can be seen in Table 4.2 – savings account inflows are often zero, which leads to a normalised standard deviation much larger than that for discretionary spend.

The absence of an increase in savings 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 emergency savings cushion or as a contribution towards specific savings goals. The absence of such transfers naturally raises the question "where does the money go?", to which I turn in the next section.

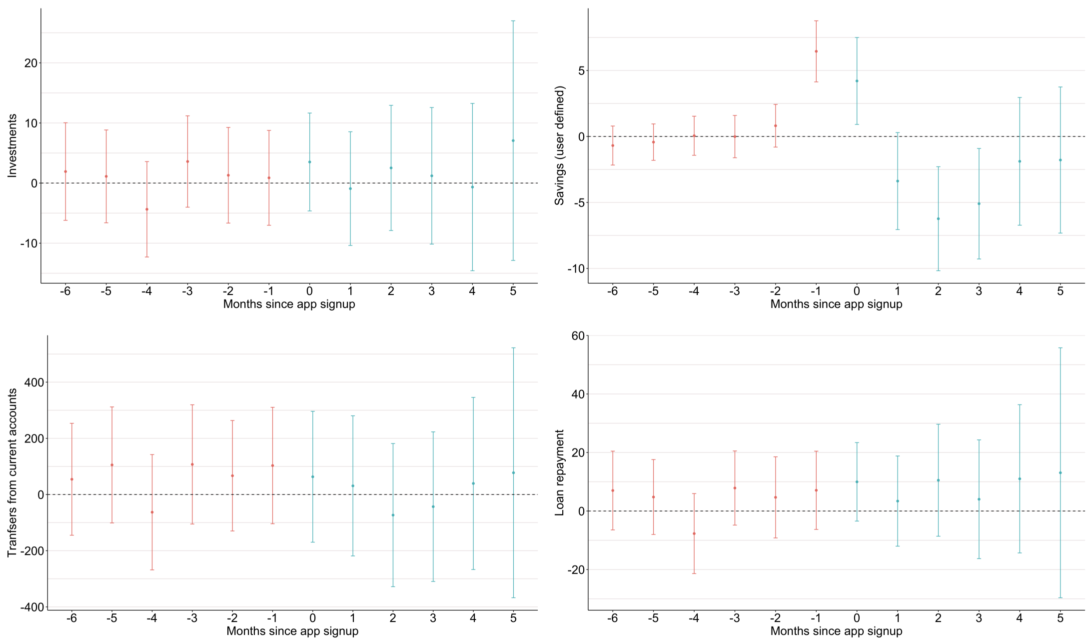
4.4.2 Where do additional funds go?

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.

[26] As discussed in Section 4.2, 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 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.

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 4.5 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.[27] 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 artefact of users only manually labelling transactions during active app use rather than also labelling historical transactions.[28]

Figure 4.5: Possible destinations of additional savings



Notes: Point estimates represent group-time average treatment effects aggregated to periods since treatment exposure, as defined in Section 4.3. 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.

Another possibility is that users use the additional funds to pay down debt.

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

[28] The pattern could be explained in two ways: either users can see the entire account history 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.

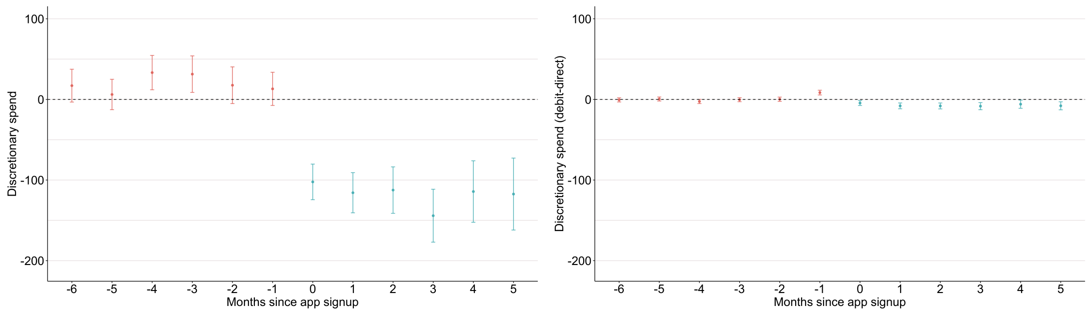
The bottom right panel in Figure 4.5 shows loan repayments, which do not change significantly during the 12-month period around app signup either.

The finding that saved funds do not flow into any of the plausible channels can be explained in three ways (and any combination thereof): (i) users might leave saved funds in their current accounts, (ii) users are homogenous and allocate saved funds across a large number of different uses in amounts I am not powered to detect, (iii) users are heterogeneous and allocate funds across a small number of uses that differ between users, so that, in the aggregate, I am not powered to detect these changes either.

4.4.3 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.

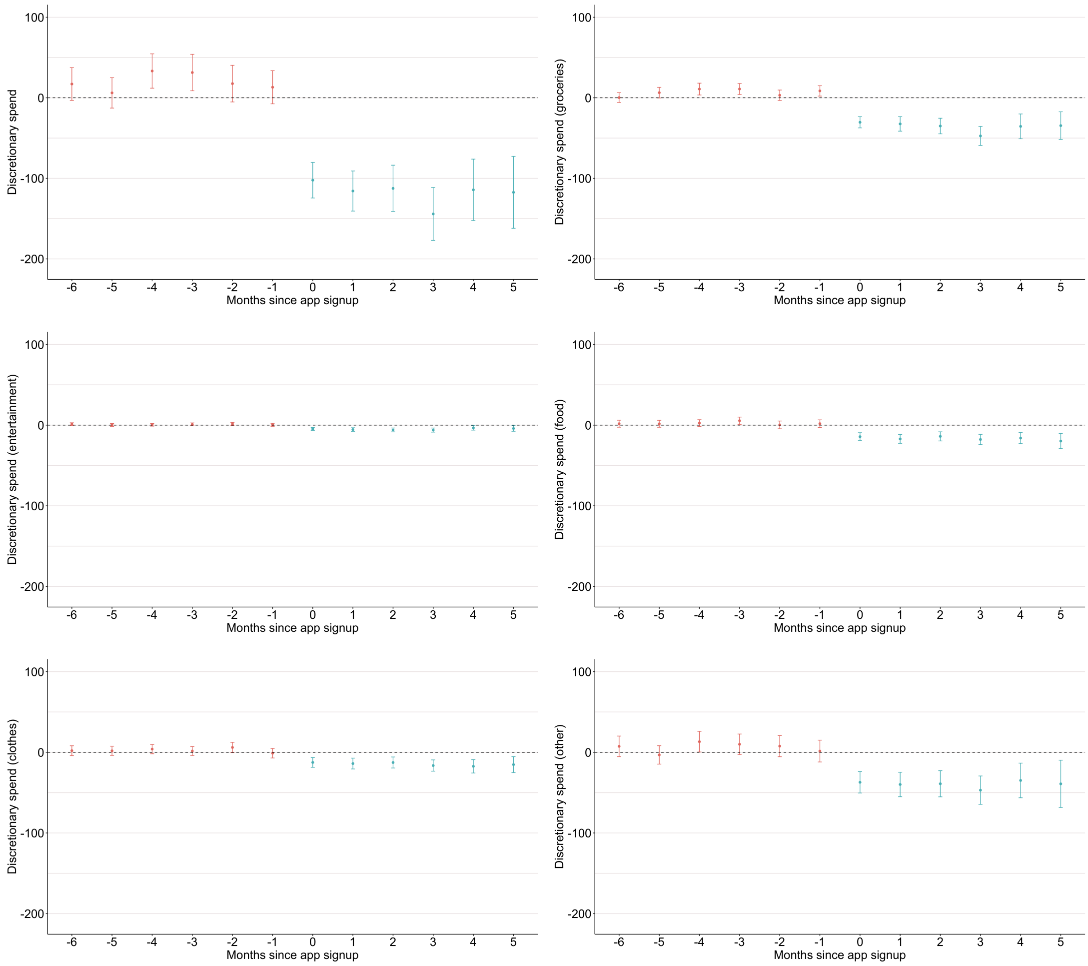
Figure 4.6: Change in direct-debit discretionary spend



Notes: Effect of app use on total discretionary spend (for reference) on the left and discretionary spend paid via debit-direct contracts on the right. Point estimates represent group-time average treatment effects aggregated to periods since treatment exposure, as defined in Section 4.3. 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.

First, users can achieve the persistent reduction in discretionary spend seen in Figure 4.4 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 reduction in spend by cancelling debit-direct contracts. Figure 4.6 shows in the left panel the reduction of discretionary spending shows in Figure 4.4 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 debit-directs.[29]

Figure 4.7: 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 4.3. 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.

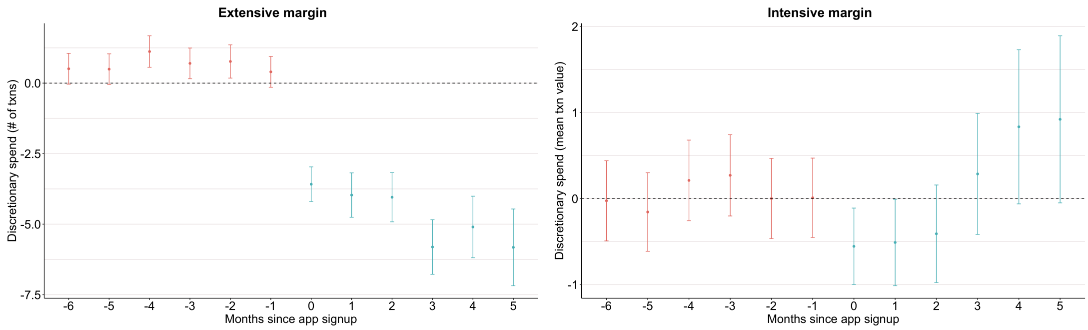
The second interesting way to disaggregate discretionary spend is into different component groups based on the type of spend; for this purpose, I group transactions into “groceries”, “clothes”, “entertainment”, “food”, which captures spending on restaurant meals and take-away, and “other”.[30] Figure 4.7 again reproduces the plot from the main results in the top-left panel for reference and

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

[30]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 purchases with Amazon, which we cannot precisely classify.

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 emerges is that these differences are quite small, and that the overall reduction in discretionary spend results from changes along all these margins.

Figure 4.8: 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 4.3. 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.

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 4.8 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 4.4.

4.5 Conclusion

In this paper, I test whether using Money Dashboard is associated with a reduction in discretionary spending and an increase in emergency 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 emergency 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 future 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 effect of app use might depend on a "dosage effect", so that the behaviour of more active users is effected more. However, work testing such an effect will have to isolate dosage of exposure from personal characteristics that make certain users engage more with the app. Third, 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, Mahoney, 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.[31] 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.

Appendix 4.A 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 4.9 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.

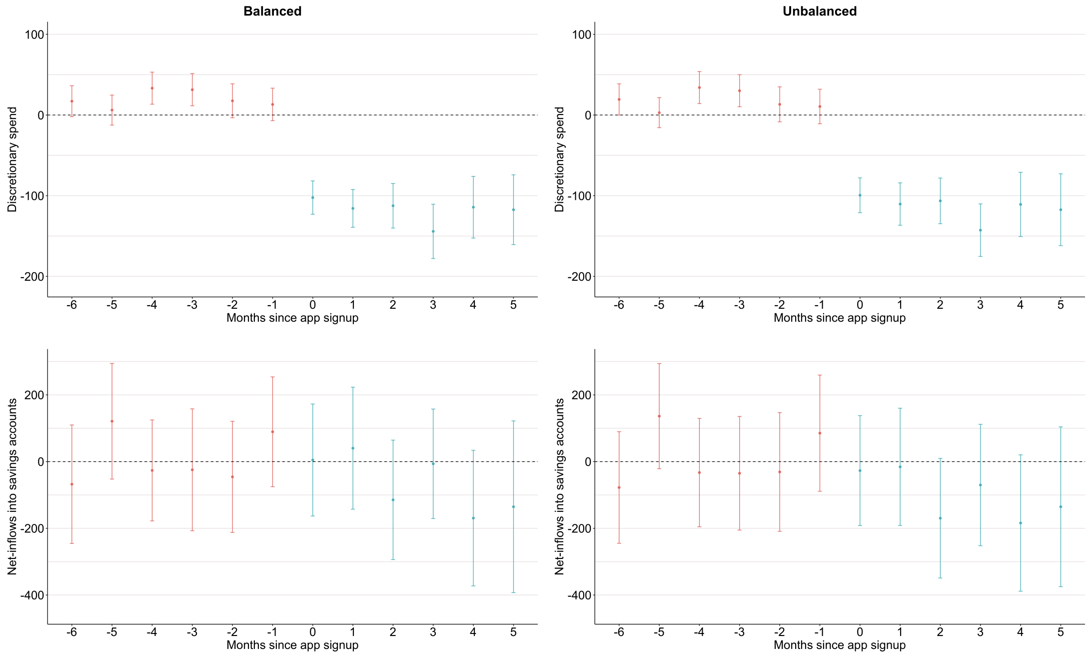
As expected, using the full data reduces the size of the confidence intervals. But the results are otherwise very similar.

Appendix 4.B Inflows and outflows

The leftmost panel in Figure 4.10 reproduces the results from Figure 4.4 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.

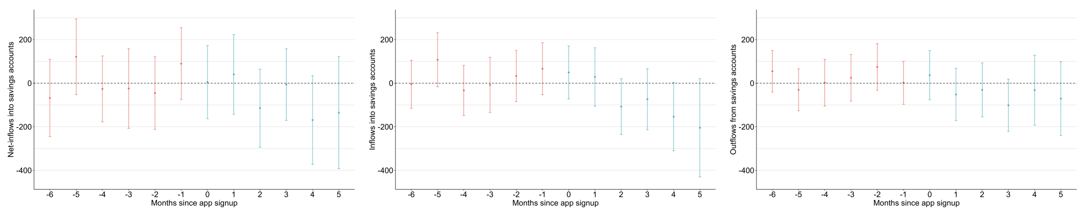
[31] In the UK, for instance, the first £4000 paid into a Lifetime ISA each year are matched by the government by payment of £1000, corresponding to a risk-free interest rate of 25 percent.

Figure 4.9: 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 4.3. 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 4.10: 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 4.3. 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.

Chapter 5

Conclusion

This thesis consists of three stand-alone chapters, each of which addresses a different aspect of my overall aim of using large-scale transaction level data, insights from behavioural science, and methods from econometrics and machine learning to study how individuals spend money and save for the short-term.

Chapter 2, which is joint work with Neil Stewart, introduces 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 instalments, and swapping car and home insurer frequently or infrequently. Each of these choices 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.

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 where 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. One 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.

We find, however, that propensities for patience based on different choice scenarios are uncorrelated, suggesting that in our particular application, our approach is unsuccessful at eliciting underlying preferences. There are three ways in which our work could be refined in future research. First, 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, it is possible that people the way people actually make intertemporal choices is not well captured, even approximately, by the discounting models on which our elicitation approach builds. 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.

Chapter 3, also join work with Neil Stewart, calculates simple summary statistics of individuals' spending profiles and tests whether differences in spending profiles are predictive of the frequency with which individuals make transactions into their savings accounts. To summarise spending profiles, we calculate the entropy of an individual's spending profile in a given period, which captures the predictability of spending transactions in said period.

We focus on spending profiles for three reasons: first, our understanding of how individuals spend their money is largely based on survey data from a relatively small number of individuals. Large-scale transaction data of the type used in this paper has become available to researchers only in recent years and has not, to the best of our knowledge, been used to explore patterns in consumer spending behaviour. Second, research suggests that spending profiles might reflect circumstances in an individual's life that are also related to their saving behaviour. Spending entropy, defined similarly as in this chapter, has been found to predict financial distress, calorie intake, and the frequency of and amount spend during supermarket visits. Finally, we think of our focus on spending profile as a proof of concept for the broader agenda of testing whether we can extract simple summary statistics from large-scale transaction data that capture relevant information about individuals' life circumstances. If successful, such information could be used in the design of financial management apps to provide timely and customised assistance to users when they need it most.

Our results show that spending entropy is predictive of the frequency of savings transactions, with an effect size similar to that of an increase in monthly income income of between £1,000 and £2,000. This holds true if we control for simple component parts of entropy, suggesting that the non-linear way in which entropy combines these components captures something about spending profiles that is of relevance. The results are also largely consistent across entropy measures calculated based on different product and merchant categories. However, they direction of the effect changes for different types of entropy that differ in the way in which we treat product or merchant categories in which an individual makes no transactions in a given period. Exploratory research suggests that the number of such zero count categories, together with the variation of counts for categories with a positive number of transactions, goes some way in explaining the outcome. But more work is necessary to more fully understand this outcome, which we leave for future research.

Chapter 4, tests whether using Money Dashboard is associated with a reduction in discretionary spending and an increase in emergency 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 emergency 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, nor additional loan repayments, which suggests that people allocate their savings to a number of different purposes that we are not powered to detect in the data.

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 who probably do have some unobserved motivation to reduce their savings actually achieve this.

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 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. 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.

Bibliography

Abadie, Alberto and Sebastien Gay (2006). "The impact of presumed consent legislation on cadaveric organ donation: a cross-country study". In: *Journal of health economics* 25.4, pp. 599–620.

Agarwal, Sumit, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, and Douglas D Evanoff (2010). "Learning to cope: Voluntary financial education and loan performance during a housing crisis". In: *American Economic Review* 100.2, pp. 495–500.

Agarwal, Sumit, Souphala Chomsisengphet, and Cheryl Lim (2017). "What shapes consumer choice and financial products? A Review". In: *Annual Review of Financial Economics* 9, pp. 127–146.

Agarwal, Sumit, Souphala Chomsisengphet, Chunlin Liu, and Nicholas S Souleles (2015). "Do consumers choose the right credit contracts?" In: *The Review of Corporate Finance Studies* 4.2, pp. 239–257.

Agarwal, Sumit, John C Driscoll, Xavier Gabaix, and David Laibson (2009). "The age of reason: Financial decisions over the life cycle and implications for regulation". In: *Brookings papers on Economic activity* 2009.2, pp. 51–117.

Agarwal, Sumit and Bhashkar Mazumder (2013). "Cognitive abilities and household financial decision making". In: *American Economic Journal: Applied Economics* 5.1, pp. 193–207.

Agarwal, Sumit, Paige Marta Skiba, and Jeremy Tobacman (2009). "Payday loans and credit cards: New liquidity and credit scoring puzzles?" In: *American Economic Review* 99.2, pp. 412–17.

Aiken, Emily, Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua E Blumenstock (2022). "Machine learning and phone data can improve targeting of humanitarian aid". In: *Nature* 603.7903, pp. 864–870.

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 (2011). "Social norms and energy conservation". In: *Journal of public Economics* 95.9-10, pp. 1082–1095.

Allcott, Hunt and Todd Rogers (2014). "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation". In: *American Economic Review* 104.10, pp. 3003–37.

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, and John Leahy (2003). "Wealth accumulation and the propensity to plan". In: *The Quarterly Journal of Economics* 118.3, pp. 1007–1047.

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

Ashraf, Nava, Dean Karlan, and Wesley Yin (2006). "Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines". In: *The Quarterly Journal of Economics* 121.2, pp. 635–672.

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.

Ausubel, Lawrence M (1991). "The failure of competition in the credit card market". In: *The American Economic Review*, pp. 50–81.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel (2018). "The economic effects of social networks: Evidence from the housing market". In: *Journal of Political Economy* 126.6, pp. 2224–2276.

Bailey, Michael, Eduardo Dávila, Theresa Kuchler, and Johannes Stroebel (2019). "House price beliefs and mortgage leverage choice". In: *The Review of Economic Studies* 86.6, pp. 2403–2452.

Bailey, Michael, Patrick Farrell, Theresa Kuchler, and Johannes Stroebel (2020). "Social connectedness in urban areas". In: *Journal of Urban Economics* 118, p. 103264.

Bailey, Michael, Drew M Johnston, Martin Koenen, Theresa Kuchler, Dominic Russel, and Johannes Stroebel (2020). "Social networks shape beliefs and behavior: Evidence from social distancing during the covid-19 pandemic". Tech. rep. National Bureau of Economic Research.

Bailey, Michael, Drew M Johnston, Theresa Kuchler, Johannes Stroebel, and Arlene Wong (2019). "Peer effects in product adoption". Tech. rep. National Bureau of Economic Research.

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, James J Choi, David Laibson, and Brigitte C Madrian (2009). "The importance of default options for retirement saving outcomes: Evidence from the United States". In: *Social security policy in a changing environment*. University of Chicago Press, pp. 167–195.

— (2018). "Behavioral household finance". In: *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 1. Elsevier, pp. 177–276.

Beshears, John, James J Choi, David Laibson, Brigitte C Madrian, and Katherine L Milkman (2015). "The effect of providing peer information on retirement savings decisions". In: *The Journal of finance* 70.3, pp. 1161–1201.

Beshears, John, James J Choi, David Laibson, Brigitte C Madrian, and William L Skimmyhorn (2022). "Borrowing to save? The impact of automatic enrollment on debt". In: *The Journal of Finance* 77.1, pp. 403–447.

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.

Blumenstock, Joshua (2018). *Don't forget people in the use of big data for development.*

Blumenstock, Joshua Evan, Guanghua Chi, and Xu Tan (2019). "Migration and the value of social networks". In:

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On (2015). "Predicting poverty and wealth from mobile phone metadata". In: *Science* 350.6264, pp. 1073–1076.

Blumenstock, Joshua, Michael Callen, and Tarek Ghani (2018). "Why do defaults affect behavior? Experimental evidence from Afghanistan". In: *American Economic Review* 108.10, pp. 2868–2901.

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.*

Brandon, Alec, Paul Ferraro, John A List, Robert Metcalfe, Michael Price, and Florian Rundhammer (2017). "Do the effects of social nudges persist? Theory and evidence from 38 natural field experiments". In: *NBER Working Paper* w23277.

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.

Bryan, Gharad, Dean Karlan, Scott Nelson, et al. (2010). "Commitment devices". In: *Annual review of Economics* 2.1, pp. 671–698.

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.

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.*

Carroll, Gabriel D, James J Choi, David Laibson, Brigitte C Madrian, and Andrew Metrick (2009). "Optimal defaults and active decisions". In: *The quarterly journal of economics* 124.4, pp. 1639–1674.

CFPB, Consumer Financial Protection Bureau (2015). "Financial well-being: the goal of financial educaion". Tech. rep. URL: https://files.consumerfinance.gov/f/201501_cfpb_report_financial-well-being.pdf.

— (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.

Chase, J.P. Morgan (2019). "Weathering Volatility 2.0". Tech. rep. URL: https://www.jpmorganchase.com/content/dam/jpmc/jpmorgan-chase-and-co/institute/pdf/institute-volatility-cash-buffer-report.pdf.

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*.

Chetty, Raj, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, et al. (2022a). "Social capital I: measurement and associations with economic mobility". In: *Nature* 608.7921, pp. 108–121.

— (2022b). "Social capital II: determinants of economic connectedness". In: *Nature* 608.7921, pp. 122–134.

Chi, Guanghua, Han Fang, Sourav Chatterjee, and Joshua E Blumenstock (2022). "Microestimates of wealth for all low-and middle-income countries". In: *Proceedings of the National Academy of Sciences* 119.3, e2113658119.

Choi, James J, David Laibson, Brigitte C Madrian, and Andrew Metrick (2002). "Defined contribution pensions: Plan rules, participant choices, and the path of least resistance". In: *Tax policy and the economy* 16, pp. 67–113.

— (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.

D'Acunto, Francesco and Alberto G Rossi (2021). "New frontiers of robo-advising: Consumption, saving, debt management, and taxes". In: *Saving, Debt Management, and Taxes (February 1, 2021)*.

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

De Mel, Suresh, Craig McIntosh, and Christopher Woodruff (2013). "Deposit collecting: Unbundling the role of frequency, salience, and habit formation in generating savings". In: *American Economic Review* 103.3, pp. 387–92.

De Montjoye, Yves-Alexandre, Laura Radaelli, Vivek Kumar Singh, and Alex "Sandy" Pentland (2015). "Unique in the shopping mall: On the reidentifiability of credit card metadata". In: *Science* 347.6221, pp. 536–539.

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.

Duckworth, Angela L, Katherine L Milkman, and David Laibson (2018). "Beyond willpower: Strategies for reducing failures of self-control". In: *Psychological Science in the Public Interest* 19.3, pp. 102–129.

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

Ericson, Keith Marzilli (2017). "On the interaction of memory and procrastination: Implications for reminders, deadlines, and empirical estimation". In: *Journal of the European Economic Association* 15.3, pp. 692–719.

Ericson, Keith Marzilli and David Laibson (2019). "Intertemporal choice". In: *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 2. Elsevier, pp. 1–67.

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.

FCA, Financial Conduct Authority (2016). "Credit card market study - final findings report". Tech. rep. URL: https://www.fca.org.uk/publication/market-studies/ms14-6-3-credit-card-market-study-final-findings-report.pdf.

Fernandes, Daniel, John G Lynch Jr, and Richard G Netemeyer (2014). "Financial literacy, financial education, and downstream financial behaviors". In: *Management Science* 60.8, pp. 1861–1883.

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.

Gathergood, John and Arna Olafsson (2020). "The co-holding puzzle: New evidence from transaction-level data". In:

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.

Gimbel, Ronald W, Martin A Strosberg, Susan E Lehrman, Eugenijus Gefenas, and Frank Taft (2003). "Presumed consent and other predictors of cadaveric organ donation in Europe". In: *Progress in Transplantation* 13.1, pp. 17–23.

Giné, Xavier, Dean Karlan, and Jonathan Zinman (2010). "Put your money where your butt is: a commitment contract for smoking cessation". In: *American Economic Journal: Applied Economics* 2.4, pp. 213–35.

Gollwitzer, Peter M and Paschal Sheeran (2006). "Implementation intentions and goal achievement: A meta-analysis of effects and processes". In: *Advances in experimental social psychology* 38, pp. 69–119.

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.

Greenberg, Adam Eric and Hal Hershfield (2019). "Financial decision making". In: *Consumer Psychology Review* 2.1, pp. 17–29.

Gross, David B and Nicholas S Souleles (2002). "Do liquidity constraints and interest rates matter for consumer behavior? Evidence from credit card data". In: *The Quarterly journal of economics* 117.1, pp. 149–185.

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.

Guttman-Kenney, Benedict, Paul Adams, Stefan Hunt, David Laibson, and Neil Stewart (2021). "The semblance of success in nudging consumers to pay down credit card debt". Tech. rep. working paper, University of Chicago.

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.

Haushofer, Johannes and Ernst Fehr (2014). "On the psychology of poverty". In: *Science* 344.6186, pp. 862–867.

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.

Imbens, Guido W (2021). "Statistical significance, p-values, and the reporting of uncertainty". In: *Journal of Economic Perspectives* 35.3, pp. 157–74.

Jaffe, Eric (2014). "What Big Data Means For Psychological Science". In: *APS Observer* 27.

Johnson, Eric J and Daniel Goldstein (2003). *Do defaults save lives?*

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

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

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk". In: *Econometrica* 47.2, pp. 263–91.

Kaur, Supreet, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach (2021). "Do Financial Concerns Make Workers Less Productive?" Working Paper 28338. National Bureau of Economic Research. DOI: 10.3386/w28338. URL: http://www.nber.org/papers/w28338.

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.

Korniotis, George M and Alok Kumar (2011). "Do older investors make better investment decisions?" In: *The Review of Economics and Statistics* 93.1, pp. 244–265.

Kosinski, Michal, Sandra C Matz, Samuel D Gosling, Vesselin Popov, and David Stillwell (2015). "Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines." In: *American psychologist* 70.6, p. 543.

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.*

Kuchler, Theresa and Johannes Stroebel (2021). "Social finance". In: *Annual Review of Financial Economics* 13, pp. 37–55.

Kumar, Alok (2009). "Who gambles in the stock market?" In: *The Journal of Finance* 64.4, pp. 1889–1933.

Kumar, Alok and Sonya Seongyeon Lim (2008). "How do decision frames influence the stock investment choices of individual investors?" In: *Management science* 54.6, pp. 1052–1064.

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

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

— (2015). "Why don't present-biased agents make commitments?" In: *American Economic Review* 105.5, pp. 267–72.

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*.

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.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao (2013). "Poverty impedes cognitive function". In: *science* 341.6149, pp. 976–980.

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.

McKenzie, Craig RM and Michael J Liersch (2011). "Misunderstanding savings growth: Implications for retirement savings behavior". In: *Journal of Marketing Research* 48.SPL, S1–S13.

McKenzie, Craig RM, Michael J Liersch, and Stacey R Finkelstein (2006). "Recommendations implicit in policy defaults". In: *Psychological Science* 17.5, pp. 414–420.

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).

Montjoye, Yves-Alexandre de, Jordi Quoidbach, Florent Robic, and Alex Sandy Pentland (2013). "Predicting personality using novel mobile phone-based metrics". In: *International conference on social computing, behavioral-cultural modeling, and prediction*. Springer, pp. 48–55.

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.

Mullainathan, Sendhil and Eldar Shafir (2013). *Scarcity: Why having too little means so much*. Macmillan.

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.02.co.uk/help/account-and-billing/first-bill (visited on 01/25/2020).

Oettingen, Gabriele and Peter Gollwitzer (2010). *Strategies of setting and implementing goals: Mental contrasting and implementation intentions.*

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.

Perry, Vanessa G and Marlene D Morris (2005). "Who is in control? The role of self-perception, knowledge, and income in explaining consumer financial behavior". In: *Journal of consumer affairs* 39.2, pp. 299–313.

Philippon, Thomas (2019). "On fintech and financial inclusion". Tech. rep. National Bureau of Economic Research.

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

Puri, Manju and David T Robinson (2007). "Optimism and economic choice". In: *Journal of financial economics* 86.1, pp. 71–99.

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.

Rick, Scott I, Cynthia E Cryder, and George Loewenstein (2008). "Tightwads and spendthrifts". In: *Journal of consumer research* 34.6, pp. 767–782.

Rick, Scott I, Deborah A Small, and Eli J Finkel (2011). "Fatal (fiscal) attraction: Spendthrifts and tightwads in marriage". In: *Journal of Marketing Research* 48.2, pp. 228–237.

Robinson, Carly D, Gonzalo A Pons, Angela L Duckworth, and Todd Rogers (2018). "Some middle school students want behavior commitment devices (but take-up does not affect their behavior)". In: *Frontiers in Psychology* 9, p. 206.

Rogers, Todd and Avi Feller (2016). "Discouraged by peer excellence: Exposure to exemplary peer performance causes quitting". In: *Psychological science* 27.3, pp. 365–374.

Rogers, Todd, Katherine L Milkman, Leslie K John, and Michael I Norton (2015). "Beyond good intentions: Prompting people to make plans improves follow-through on important tasks". In: *Behavioral Science & Policy* 1.2, pp. 33–41.

Roll, Stephen and Mathieu Despard (2020). "Income loss and financial distress during COVID-19: The protective role of liquid assets". In: *Available at SSRN 3733862*.

Romer, David (2020). "In praise of confidence intervals". In: *AEA Papers and Proceedings*. Vol. 110, pp. 55–60.

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*.

Royer, Heather, Mark Stehr, and Justin Sydnor (2015). "Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune-500 company". In: *American Economic Journal: Applied Economics* 7.3, pp. 51–84.

Sakaguchi, Hiroaki, Neil Stewart, John Gathergood, Paul Adams, Benedict Guttman-Kenney, Lucy Hayes, and Stefan Hunt (2022). "Default effects of credit card minimum payments". In: *Journal of Marketing Research*, p. 00222437211070589.

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

Samuelson, William and Richard Zeckhauser (1988). "Status quo bias in decision making". In: *Journal of risk and uncertainty* 1.1, pp. 7–59.

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

Schaner, Simone (2018). "The persistent power of behavioral change: Long-run impacts of temporary savings subsidies for the poor". In: *American Economic Journal: Applied Economics* 10.3, pp. 67–100.

Schultz, P Wesley, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, and Vladas Griskevicius (2007). “The constructive, destructive, and reconstructive power of social norms”. In: *Psychological science* 18.5, pp. 429–434.

Schwartz, Janet, Daniel Mochon, Lauren Wyper, Josiase Maroba, Deepak Patel, and Dan Ariely (2014). “Healthier by precommitment”. In: *Psychological science* 25.2, pp. 538–546.

Shah, Anuj K, Sendhil Mullainathan, and Eldar Shafir (2012). “Some consequences of having too little”. In: *Science* 338.6107, pp. 682–685.

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.

Shui, Haiyan and Lawrence M Ausubel (2004). “Time inconsistency in the credit card market”. In: *Available at SSRN 586622*.

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:

Soman, Dilip and Amar Cheema (2011). “Earmarking and partitioning: Increasing saving by low-income households”. In: *Journal of Marketing Research* 48.SPL, S14–S22.

Stachl, Clemens, Quay Au, Ramona Schoedel, Samuel D Gosling, Gabriella M Harari, Daniel Buschek, Sarah Theres Völkel, Tobias Schuwerk, Michelle Oldemeier, Theresa Ullmann, et al. (2020). “Predicting personality from patterns of behavior collected with smartphones”. In: *Proceedings of the National Academy of Sciences* 117.30, pp. 17680–17687.

Step Change (2017). “Strengthening the saving safety net”. Tech. rep. URL: https://www.stepchange.org/Portals/0/documents/media/reports/strengthening-the-saving-safety-net-stepchange-debt-charity-response-paper-to-the-pensions-policy-institute.pdf.

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™: 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.

Youyou, Wu, Michal Kosinski, and David Stillwell (2015). "Computer-based personality judgments are more accurate than those made by humans". In: *Proceedings of the National Academy of Sciences* 112.4, pp. 1036–1040.

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).