Contents

1 Introduction **1**

1.1 Theoretical context 1

1.2 Objectives and research questions 1

1.3 Chapter summaries 2

1.4 Contribution 3

1.5 Limitations and directions for further research 3

2 Conclusion **4**

2.1 Summary 4

2.2 Discussion 4

Bibliography **5**

Chapter 1

Introduction

1.1 Theoretical context

in the UK, 25 percent of adults would be unable to cover an unexpected bill of £300 (Philipps et al. 2021), while in the US, about 30 percent would be unable to cover a $400 bill (Governors of the Federal Reserve System 2022)

- Financial wellbeing is important.
- People in UK and US also don't have enough to cover unexpected outlays. (See reports). Also, Sabat and Gallagher (2019)
- This has important consequences:
  - Short-term: financial well-being (see reports)
  - Long-term (viscious cycle): scarcity hypothesis - makes it harder to focus on important things (plan for retirement, focus on healthy lifestyle, support children, ...) and might lead to vicious cycle (less savings leading to increased risk of financial hardship leading to more stress leading to less savings...)
  - Buffering agains financial hardship (Roll and Despard 2020)
- Spending and savings behaviour is important component - it's not all about lack of income.

1.2 Objectives and research questions

- I use a single dataset, provided by Money Dashboard, for all three papers.
- The data presented a number of challenges.

- Handling data of this size.
- Secure storage with easy remote access (AWS).
- Consistenc and efficient preprocessing (my process).
- Open science contribution: all work available on Github, code available for preprocessing from beginning to end.

Motivation: My experience as a researcher in well-known academic and private-sector institutions has made clear to me over the years that careful, reliable, and replicable data preprocessing is undervalued in many settings.

The result are plain data errors that overturn results (Reinhard and Rogoff) and replication issues (see Ariely controversies). Given my experience, I'd think that all we know is the tip of the iceberg.

In the vast majority of empirical research projects there is no good reason to not make the code public, even when the data is proprietary. And yet, for the vast majority of papers, there is no code available to replicate findings and check precise implementation details, which often matter but are not described in papers.

All of this harms science - the quality of it and the trust in it.

I have dedicated a lot of time during my PhD to ensure that I can rectify at least some of those issues.

The code for all my projects is available online and includes scripts that can be used to easily run the entire analysis.

1.3 Chapter summaries

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

1.4 Contribution

1.5 Limitations and directions for further research

Chapter 2

Conclusion

This thesis consists of three independent research studies on ...

In the first part of this conclusion, I will summarise the three chapters, before discussing challenges...

2.1 Summary

The two applied studies presented in Chapters 1 and 2 are purely exploratory. A natural next step to exploring their generalisability is a pre-registered replication on new data (Van den Akker et al., 2019). We did not pre-register any studies presented here since many analysis decisions were taken conditional on the data observed. It is a well-known problem that such data-dependent decisions can affect the conclusions drawn from the research (Simmons, Nelson, and Simonsohn, 2011; Gelman and Loken, 2014). The studies are thus a useful first step in exploring the questions of ethnic bias and domestic abuse and will require further, pre-registered analyses.

2.2 Discussion

Bibliography

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

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

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

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.

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

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*.

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