Essays in Empirical Microeconomics

Heinrich Friedrich Otto Richard Karl Kögel

Munich, 2019

Essays in Empirical Microeconomics

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2019

vorgelegt von

Heinrich Friedrich Otto Richard Karl Kögel

Referent:	Prof. Dr. Joachim Winter
Korreferent:	Prof. Axel Börsch-Supan, Ph.D.
Promotionsabschlussberatung:	05. Februar 2020

Tag der mündlichen Prüfung:	23. Januar 2020
Namen der Berichterstatter:	Joachim Winter, Axel Börsch-Supan, Derya Uysal

Acknowledgments

First and foremost, I would like to thank my supervisor, Joachim Winter, for his constant support, guidance, and encouragement. My research has benefited considerably from his advice. I am grateful to have had the opportunity to learn so much from him. I would also like to express my deep gratitude to my second supervisor, Axel Börsch-Supan. His insightful feedback contributed a great deal to this dissertation. I also very much appreciate the excellent research environment that he provided at the Munich Center for the Economics of Aging. I would further like to thank Derya Uysal for kindly agreeing to serve on my dissertation committee.

I also thank my former and current colleagues at MEA and the Chair of Empirical Economic Research for numerous stimulating and helpful discussions. In particular, I want to express my gratitude to my two co-authors, Helmut Farbmacher and Martin Spindler. Working with them was a delight and has taught me a lot.

I am grateful to Lisa Berkman for inviting me to Harvard University. My research visit was a fantastic experience that led not only to fruitful extensions of my dissertation projects but also new friendships.

Finally, I thank my girlfriend, family and friends for their encouragement and support throughout every stage of my journey.

Contents

Preface **1**

1 Financial Scarcity and Health: Evidence from the Food Stamp Program **8**

1.1 Introduction 8

1.2 Background 10

1.2.1 Food Stamp Program 10

1.2.2 Financial Circumstances over the Food Stamp Cycle 12

1.3 Empirical Strategy 13

1.4 Data 18

1.4.1 American Time Use Survey 18

1.4.2 Sample and Descriptive Statistics 19

1.4.3 Randomization Checks 23

1.5 Main Results 25

1.6 Potential Mechanisms 28

1.7 Robustness Analysis and Placebo Tests 31

1.7.1 Robustness Analysis 31

1.7.2 Placebo Tests 33

1.8 Conclusion 37

Appendix A 39

A.1 Additional Tables 40

2 Heterogeneous Effects of Poverty on Cognition **42**

2.1 Introduction 42

2.2 Experiment and Data 44

2.2.1 Experiment 44

2.2.2 Sample and Descriptive Statistics 46

2.3 Methodology 50

2.4 Results 53

2.4.1 OLS Analysis 53

2.4.2 Heterogeneity Analysis 55

2.4.3 Subgroup Analysis 64

2.5 Conclusion 67

Appendix B 69

B.1 Derivation of the Causal Forest Estimator 70

B.2 Tree Example 73

B.3 Calculation of the Variable Importance Measure 73

B.4 Procedure to Set the Covariates 74

B.5 Estimates in the Vicinity of the Two Typical Individuals 75

B.6 Additional Tables 77

3 A Natural Experiment on the Role of Response Uncertainty in Household Surveys **83**

3.1 Introduction 83

3.2 Natural Experiment and Data 86

3.2.1 Natural Experiment on Response Uncertainty in the HRS 86

3.2.2 Econometric Approach 89

3.2.3 Sample and Descriptive Statistics 90

3.3 Main Results 93

3.4 Robustness Checks 98

3.5 Conclusions 102

4 Testing under a Special Form of Heteroscedasticity 103

4.1 Introduction 103

4.2 Inference Issues and Test Procedure 104

4.3 Monte Carlo Simulations 108

4.4 Empirical Illustration: Returns to Education 113

4.5 Conclusion 117

Bibliography 118

List of Figures

1.1 Distribution of the Mean Number of Days since the Last Food Stamp Receipt 22

2.1 Variable Importance Plots for the Causal Forests 56

2.2 Causal Forest Effect Estimates over an Age–Current Income Grid . . . 58

2.3 Causal Forest Effect Estimates for the Typical Younger Individual . . . 62

2.4 Causal Forest Effect Estimates for the Typical Older Individual 63

3.1 Distribution of the Days since the Last Social Security Check Delivery . 89

3.2 Fraction of Missing Check Amounts 94

3.3 Fraction of 10-Focal Responses 97

3.4 Fraction of 50-Focal Responses 97

3.5 Fraction of 100-Focal Responses 98

4.1 Elliptical Heteroscedasticity Example 107

4.2 Power Plots for the Heteroscedasticity Tests 110

4.3 Power Plots for Wald Tests Using Conventional and Robust Standard Errors 112

4.4 Replication of Figure 1, Amin (2011) 113

List of Tables

1.1 Descriptive Statistics for the Groups of Individuals Used in the Analysis 20

1.2 Balance Checks 24

1.3 Estimates for the Effect of Financial Scarcity on Self-Assessed Physical Health 26

1.4 Estimates for the Effect of Financial Scarcity on Reported Sleeplessness 29

1.5 Robustness Checks Using Alternative End-of-Month Definitions 32

1.6 Robustness Checks Using an Alternative Early States Definition and Week Dummies 34

1.7 Placebo Tests 36

A.1 Food Stamp Issuance Dates 40

2.1 Definitions and Descriptive Statistics for the Outcomes and Regressor of Interest 47

2.2 Descriptive Statistics for the Covariates 48

2.3 OLS Average Effect Estimates 54

2.4 Subgroup Average Effect Estimates 65

2.5 Subgroup Average Effect Estimates in an Independent Experiment . . . 66

B.1 Variation in Financial Resources at Payday 77

B.2 Balance Checks 78

B.3 OLS Average Effect Estimates for the Subgroups Analyzed by Carvalho et al. (2016) 80

B.4 Causal Forest Estimates for Typical Individuals in the Vicinity of the Typical Younger Individual 81

B.5 Causal Forest Estimates for Typical Individuals in the Vicinity of the Typical Older Individual 82

3.1 Descriptive Statistics for the Covariates and Social Security Check Amount Variable 91

3.2 Variable Definitions and Descriptive Statistics for the Uncertainty Measures, Regressor of Interest, and Grouping Variables 92

3.3 Change in the Fraction of Missing Check Amounts between the Early and Late Stage of the Social Security Payment Cycle 93

3.4 Change in the Fraction of Focal Responses between the Early and Late Stage of the Social Security Payment Cycle 96

3.5 Alternative Late Dummy Definition 1: Change in the Fraction of Focal Responses between the Early and Late Stage of the Social Security Payment Cycle 100

3.6 Alternative Late Dummy Definition 2: Change in the Fraction of Focal Responses between the Early and Late Stage of the Social Security Payment Cycle 101

4.1 Heteroscedasticity Test Results for the Within-Twin Pair Regressions in Table 1, Amin (2011) 115

4.2 Replication and Re-Estimation of the Within-Twin Pair Regressions in Table 1, Amin (2011) 116

Preface

Investigating causal relationships plays a central role in empirical economic research. Some of the classic relationships studied by economists include the effect of education on income, the impact of minimum wages on unemployment, and the influence of wealth on health. Indeed, uncovering causal relationships, as opposed to mere associations, is crucial from a policy point of view. To be able to design policies that lead to desired goals, it is usually vital to understand cause and effect. Consider, for example, the relation between education and income. It has generally been observed that individuals with more education tend to have higher incomes. If this relationship were the result of unobserved factors (such as greater innate ability among the more educated) rather than reflecting a causal effect of education on income, then trying to promote individual prosperity via education policies would be futile and costly.

Estimating causal effects is often challenging, however, as doing so typically requires variation in the regressor of interest while all other factors are held constant. The gold standard for achieving this is to run an experiment in which individuals are randomly assigned to a treatment of interest. Due to ethical, financial, or political reasons, however, this is often not possible in economics. For example, most people would probably find it unacceptable to exclude individuals from the education system to study how education affects income. This inability to run experiments in many cases has led to the development of a rich toolkit of methods that empirical economists use to tease out causality from observational data. Common methods include difference-in-differences approaches, regression discontinuity designs, instrumental variables, and matching techniques (see, e.g., Abadie and Cattaneo 2018). More recently, economists have started to expand the traditional toolkit by methods from machine learning. Ex-

amples include the least absolute shrinkage and selection operator (lasso) and random forests, as well as adaptations of these that are specifically designed to study causal relationships (see, e.g., Athey and Imbens 2019).

This thesis contributes to an understanding of causal relationships in a number of important areas by employing a broad set of empirical methods and using a variety of data sources. In the first two chapters, I provide evidence on the effects of poverty on health and cognition. Considering the large number of individuals who live in poverty, shedding light on its consequences is highly relevant from a policy perspective. In 2017, for example, almost 40 million people in the US lived in poverty (Fontenot et al. 2018). The third chapter moves on to the subject of survey response behavior and focuses on response uncertainty among participants in surveys. Despite the advent of other sources of data, such as financial transactions and social media, surveys continue to be fundamental to empirical research (Couper 2013). For example, surveys are not only crucial for research on poverty, as demonstrated by the use of survey data in the first two chapters, but they also play a key role in the investigation of the challenges faced by modern societies more generally (see, e.g., the books by Börsch-Supan et al. 2013a; 2019). Understanding the determinants of individuals' survey response behavior and related data quality issues is therefore very important. While the first three chapters study causal relationships, Chapter 4 deals with hypothesis testing under a particular form of heteroscedasticity. Drawing inferences about parameters of interest is a central part of empirical research. To be able to do so requires not only valid point estimates, but also correct standard errors (Cameron and Miller 2015). Chapter 4 provides a range of insights into this requirement.

The chapters in this thesis are self-contained and can thus be read independently of one another. The first two chapters are followed by an appendix. The combined references for all chapters come after Chapter 4. In the following, I provide a more detailed summary of each chapter.

Chapter 1 is entitled *Financial Scarcity and Health: Evidence from the Food Stamp Program*. The starting point for this chapter is the lack of financial resources that dominates the lives of many low-income individuals. Common consequences of these poor

financial circumstances include difficulty affording food or medical care and problems paying bills. Many studies have documented associations between the experience of such poor financial circumstances and worse health (see, e.g., Ferrie et al. 2005; Kahn and Pearlin 2006; Tucker-Seeley et al. 2009). However, while these findings are suggestive, I am not aware of any study so far that has been able to isolate a causal effect of poor financial circumstances on health. In Chapter 1 of this thesis, I provide evidence on this subject.

Based on a sample of low-income individuals who participate in the Food Stamp Program in the US, the analysis uses variation in financial circumstances over the monthly food stamp payment cycle. A number of studies have documented that the financial resources of food stamp recipients generally decrease over this monthly cycle, leading to especially poor financial circumstances at the cycle's end (e.g., Hastings and Washington 2010; Shapiro 2005). I estimate the short-run effect of this end-of-cycle financial scarcity on self-assessed physical health. My empirical strategy exploits the random assignment of individuals to their interview day in the American Time Use Survey. The main idea is to compare food stamp recipients interviewed at the end of the monthly food stamp cycle with food stamp recipients interviewed during the rest of the cycle. As a result of the interview day randomization, the food stamp recipients are randomly assigned to one of these two groups of individuals. To guard against confounding due to events that could occur simultaneously with the end-of-cycle financial scarcity, I extend the empirical strategy using a difference-in-differences approach that exploits variation in food stamp issuance periods across states.

I find that the financial scarcity experienced by food stamp recipients at the end of the monthly food stamp cycle leads to a sizable increase in the probability of reporting bad physical health. Randomization checks, robustness checks, and placebo tests support the validity of this finding. By exploiting the time use information in the American Time Use Survey, additional analyses suggest that increased stress may be one mechanism through which this detrimental effect occurs.

The findings suggest a number of policy implications. First, measures taken to alleviate poverty may simultaneously improve the health of low-income individuals, potentially reducing the expenditures of public health care programs such as Medicaid. Second,

when designing welfare programs, more subtle features, such as the timing of payments, can also be important. To mitigate especially poor financial circumstances at the end of welfare payment cycles and their consequences, for example, it could be beneficial to distribute benefits across shorter intervals to help individuals smooth their consumption.

Chapter 2 – *Heterogeneous Effects of Poverty on Cognition* – reports the results of joint work with Helmut Farbmacher and Martin Spindler. This chapter also contributes to an understanding of the effects of poor financial circumstances. The motivation for this chapter originates in the debate about why there are associations between poverty and potentially less beneficial behavior, such as smoking or playing the lottery. In a recent article, Mani et al. (2013) put forward a hypothesis that focuses on the financial circumstances of the poor and the potentially detrimental impact of these on cognition. The authors suggested that a preoccupation with monetary concerns may reduce the mental capacity of the poor.[1] Yet while Mani et al. (2013) indeed found evidence in favor of their hypothesis based on experiments in the US and India, only one other study to date has followed up on this. In an experiment, Carvalho et al. (2016) assigned a sample of low-income US individuals randomly to perform a number of cognitive tests before or after payday. The individuals surveyed before payday faced poorer financial circumstances than those surveyed after payday. However, the authors found no before-after differences in cognitive function in the full sample or selected subgroups.

The second chapter of the thesis contributes to this nascent literature. We analyze heterogeneity in the effect of financial circumstances based on data from the experiment by Carvalho et al. (2016). To do so, we use the causal forest method by Athey et al. (2019), which is specifically designed to study treatment effect heterogeneity in experiments. The method is based on the machine learning technique of random forests, and allows non-linear treatment effects to be estimated in a fully flexible way. In the heterogeneity analysis, we include a rich set of 37 pre-treatment covariates, including age, income, employment status, and measures of past financial strain.

While our estimations do not suggest that the poorer financial circumstances before payday affect cognition in our full sample, we do find harmful effects for younger and

[1] The hypothesis thus postulates an immediate effect of financial concerns on cognitive function. If the concerns were to be alleviated, this would directly free up mental capacity again.

elderly individuals who received a very low income around the time of the experiment. For these two groups of individuals, we also find detrimental cognitive effects in an additional experiment conducted by Carvalho et al. (2016), which we do not use in our heterogeneity analysis. One implication of our findings is that it could prove helpful for policy makers to take into account possible variation in cognitive capacity over payment cycles among the individuals in question. For example, public administration could try to avoid scheduling appointments with these individuals at the end of their payment cycles to prevent potentially poor decision making due to impaired cognition.

Chapter 3 is entitled *A Natural Experiment on the Role of Response Uncertainty in Household Surveys*. In this chapter, my co-author Joachim Winter and I focus on a determinant of response quality in household surveys: uncertainty among respondents about the quantities they are asked to report. Our analysis exploits a natural experiment that arises from the fact that Social Security checks in the US used to be delivered on the third day of each month and the notion that the interview dates in the Health and Retirement Study are quasi-randomly distributed over many weeks. We argue that these circumstances lead to exogenous variation in the time elapsed between the delivery of a participant's last Social Security check and the interview date, which can be considered a key determinant of response uncertainty in Social Security income.

Based on this natural experiment, we test the following hypotheses. Uncertainty about the amount of the Social Security payment should be greater the longer the time that has elapsed since the check was delivered because the respondents may have increasing difficulty recalling this amount. Moreover, the effect should be larger among respondents whose memory capacity is limited. In addition, we expect the effect to be even greater among those respondents who have limited memory capacity but are unaware of this limitation, as their distorted perception may lead them to putting less effort into recalling the answer to the question. As a measure of uncertainty, we use an indicator for whether an individual gave a rounded check amount as response. Rounding has been shown to be related to subjective uncertainty (see, e.g., Ruud et al. 2014). A number of studies have documented how various issues related to limited memory may lead to measurement error in economic variables, such as income or consumption expenditure (see, e.g., the literature reviewed by Browning et al. 2003; Browning et al. 2014). How-

ever, little is known about the direct role of respondents' uncertainty in determining the quality of such data. In this chapter, we provide evidence on this question.

Overall, the findings from our empirical analysis are in line with our hypotheses, even though the effects we estimate are relatively small. One may therefore ask whether our results are relevant for practical work. However, we believe that our findings are valuable considering that reporting one's Social Security income is arguably an easy enough task that one would probably not expect to find any effects at all. Our results suggest that it could be useful if survey agencies were to make metadata, such as interview dates, routinely available, as these could potentially be exploited to correct for measurement errors related to uncertainty in individuals' survey responses.

Chapter 4, entitled *Testing under a Special Form of Heteroscedasticity*, is joint work with Helmut Farbmacher. In this chapter, we deal with a special form of heteroscedasticity that leads to an upward bias in conventional, homoscedasticity-assuming standard errors. Most commonly, heteroscedasticity leads to conventional standard errors that are downward biased. When Wald tests based on these standard errors are insignificant, heteroscedasticity-robust standard errors do not change test decisions. Conversely, in situations where conventional standard errors are upward biased, using heteroscedasticity-robust standard errors may lead to different test decisions, and thus to different policy conclusions.

To be able to test for the presence of this special form of heteroscedasticity, we develop a heterocedasticity test. In Monte Carlo simulations, we show that our test is more powerful at detecting the special form of heteroscedasticity than are standard heteroscedasticity tests. This may be related to the fact that the standard test procedures test for heteroscedasticity in general, rather than the special form, and that some tests may not be well suited for detecting the non-linear nature of the heteroscedasticity of interest. In the Monte Carlo simulations, we additionally demonstrate that Wald tests using conventional standard errors lead to an actual test size that is below the given nominal level under the special form of heteroscedasticity. Conversely, Wald tests based on heteroscedasticity-robust standard errors have a correct size, and are more powerful.

In our application, we demonstrate possible consequences of the special form of het-

eroscedasticity. Amin (2011), building on work of Bonjour et al. (2003), estimated the return to education in a sample of twins. Based on conventional standard errors, his analysis did not yield a return to education estimate significantly different from zero for most of his within-twin pair regressions. Using two standard heteroscedasticity tests, we found no evidence for the presence of heteroscedasticity in Amin's (2011) data. Conversely, our proposed test does indeed suggest that the special form of heteroscedasticity is present. We then re-estimate Amin's (2011) regressions, using appropriate, heteroscedasticity-robust standard errors. Doing so yields, for most of his estimations, an estimated return to education that is significantly different from zero at conventional levels.

Chapter 1

Financial Scarcity and Health: Evidence from the Food Stamp Program

1.1 Introduction

In 2014, almost one in five Americans lived in poverty or near poverty (DeNavas-Walt and Proctor 2015; Hokayem and Heggeness 2014). A central theme in the lives of many low-income individuals is the lack of financial resources. As a result of this financial scarcity, low-income individuals often find themselves in a struggle to make ends meet, involving hardships such as problems paying bills and difficulty affording food or medical care (e.g., Barr 2012; Edin and Lein 1997; Ouellette et al. 2004). Investigating the causal consequences of such poor financial circumstances has recently become of interest to economists, who have so far focused on outcomes related to cognition and decision-making (see, e.g., Carvalho et al. 2016; Mani et al. 2013; Schilbach et al. 2016).

In epidemiology and related fields, a number of studies have documented that living in poor financial circumstances is associated with worse health. For example, poor financial circumstances have been found to be correlated with bad self-assessed health,

depression, illness symptoms, limitations in activities of daily living, serious chronic conditions, heart attacks, and mortality.[1] However, despite these suggestive findings, which are based not only on cross-sectional but also on longitudinal data, it is still not clear to what extent poor financial circumstances causally affect health. Given the large sums of money spent on the health of low-income individuals, understanding this relationship is highly relevant from a policy perspective.[2] Unfortunately, empirically isolating causal effects of financial circumstances on health is challenging. Causal effect estimates may be confounded not only by unobserved individual characteristics (such as a potentially worse health status of the financially strained in general) but also by reverse causality (i.e., health affecting financial circumstances).

This paper provides causal evidence on the effect of financial circumstances on health, based on a sample of low-income individuals who participate in the US Food Stamp Program (FSP). Previous studies have documented that the financial resources of food stamp recipients generally decrease over the monthly food stamp payment cycle, leading to especially poor financial circumstances at the cycle's end. I denote this state at the end of the monthly cycle in which many food stamp recipients' financial resources are especially scarce as financial scarcity. I estimate the short-run effect of this financial scarcity on self-assessed physical health by exploiting the random assignment of individuals to their interview day in the American Time Use Survey (ATUS).

The main idea behind the empirical strategy is to compare food stamp recipients interviewed at the end of the monthly food stamp payment cycle with food stamp recipients interviewed during the rest of the cycle. The random interview day assignment implies that the individuals are randomly assigned to one of these two groups of individuals. Balance checks confirm the success of the random assignment. To account

[1] Bad self-assessed health (Gunasekara et al. 2013; Kahn and Pearlin 2006; Wickrama et al. 2006; Stronks et al. 1998), depression (Butterworth et al. 2009; Schulz et al. 2006), illness symptoms (Kahn and Pearlin 2006; Stronks et al. 1998), limitations in activities of daily living (Szanton et al. 2010), serious chronic conditions (Kahn and Pearlin 2006), heart attacks (Ferrie et al. 2005), mortality (Tucker-Seeley et al. 2009). The studies measure poor financial circumstances typically by using indices that combine questions about whether individuals experience various hardships (such as the ones mentioned above).

[2] For example, in 2012, the US public expenditures for Medicaid amounted to approximately $432 billion (Truffer et al. 2013). This corresponds to around eight times the amount spent in the same year on the Earned Income Tax Credit scheme (Carrington et al. 2013), which is another major welfare program in the US.

for confounding events that may occur simultaneously with financial scarcity and to protect against imperfect random assignment, I extend the estimation approach using a difference-in-differences approach with two different control groups that exploits variation in food stamp issuance periods across states. Given the empirical strategy and temporary nature of the financial scarcity that I focus on, this study speaks to the short-run effect of a temporary particularly poor financial situation among a group of low-income individuals in the US.[3]

The estimations yield that the experience of the end-of-cycle financial scarcity has a detrimental effect on an individual's self-assessed physical health. Several robustness checks and placebo tests support the validity of this finding.

The remainder of this paper is structured as follows. Section 1.2 describes the Food Stamp Program and the financial circumstances over the food stamp cycle. Section 1.3 explains the empirical strategy. Section 1.4 describes the data and shows the results from randomization checks. Section 1.5 presents the main results. Section 1.6 discusses two potential mechanisms through which the effect of interest may occur. Section 1.7 presents the results from robustness checks and placebo tests. Section 1.8 concludes.

1.2 Background

1.2.1 Food Stamp Program

The Food Stamp Program is one of the central elements of the US social safety net.[4] In 2014, the FSP provided assistance to 46.5 million people at a cost of $74.2 billion (Gray and Kochhar 2015). The main goal of the FSP is to reduce food insecurity. It does so by distributing vouchers to eligible households that can be used to buy most food items at grocery stores and other authorized retailers (e.g., alcohol and prepared foods cannot be bought with food stamps). Although they are in-kind benefits, food stamp

[3]Carvalho et al. (2016) and Mani et al. (2013) also use variation in financial resources around paydays to examine the effects of financial circumstances. However, they administer their own surveys and look at outcomes related to cognition and economic decision-making.

[4]The FSP was renamed the Supplemental Nutrition Assistance Program (SNAP) in October 2008. However, I refer to the program as FSP, because the empirical analysis uses data from a time period when the program was mostly called FSP.

recipients treat food stamps similarly to cash transfers of the same amount (Hoynes and Schanzenbach 2009).

The FSP is federally funded and its rules are mostly set at the federal level. There is little variation in the program across states and its characteristics have not varied much in the last few decades (Hoynes and Schanzenbach 2016).

In contrast to other welfare programs, eligibility for the FSP is universal. It is not restricted to specific groups in the population, such as the disabled or families with children. Generally, for a household to be eligible for food stamps, it must satisfy three criteria based on its monthly gross income, net income calculated by making permitted deductions from the gross income, and its countable resources. For example, one of the criteria is that the gross monthly household income must not be greater than 130 percent of the poverty line. In addition, there are households that are categorically eligible for food stamps and, therefore, need not fulfill the three criteria.

The amount that a household receives in food stamps is calculated by subtracting 30 percent of the household's net income from a maximum benefit amount, which depends on the size of the household and is adapted annually to reflect food price changes. In 2007, the average food stamp household received $212 in monthly food stamps and consisted of 2.2 individuals. Its gross monthly income was $691, net monthly income $330, and its countable resources amounted to $144 (Wolkwitz and Leftin 2008). Thus, while the benefit amount may not seem much at $212, food stamps are still an important part of the income of the average food stamp household considering its low financial resources.

Each food stamp household in every state receives its food stamp benefits once per month. However, the timing of when the benefits are paid out within the month varies across states. Table A.1 in the appendix lists the issuance periods for all states during the sample period. While some states issue all of their food stamps on one day of the month, such as New Hampshire and Virginia, most states stagger the food stamp distribution, i.e., they distribute the benefits over a period of days. Among the states that stagger the food stamp payments, there is variation in the day of the month when a state starts its issuance period and how long the period lasts. For example, California issues food stamps between the first and tenth day of each month, whereas Mississippi distributes the ben-

efits between day 5 and day 19 each month.[5] Within the given issuance period, each state determines the food stamp delivery day for a household quasi-randomly, based on, for example, the Social Security number or case number. A household's food stamp delivery day is always the same each month.

Since 2004, all states issue food stamp benefits via an electronic system called Electronic Benefit Transfer (EBT). On the specified food stamp delivery day, the monthly benefits of a household are transferred to its EBT card, which works similar to a conventional debit card. The food stamp benefits are, therefore, immediately available on the designated delivery day and they can be used for shopping right away.

1.2.2 Financial Circumstances over the Food Stamp Cycle

A number of studies have investigated the financial circumstances and related behaviors of food stamp households over their monthly food stamp cycle, i.e., the time from one food stamp benefit payment to the next.

Using data from retailers (e.g., Castellari et al. 2017; Goldin et al. 2016; Hastings and Washington 2010) and surveys (e.g., Shapiro 2005; Wilde and Ranney 2000), research finds evidence that the expenditures of food stamp households decrease in the time since the last receipt of food stamps. For example, based on panel data containing 1.13 million observations over the period 2004–2011, Goldin et al. (2016) estimate a 27 percent drop in food expenditures between the first and last week of the monthly food stamp cycle for food stamp eligible households relative to non-food stamp households.[6]

Related to these studies, Cole and Lee (2005) examine food stamp redemption patterns using actual transaction data from the EBT system. Their analysis yields that food stamp households spend on average 80 percent of their food stamp benefits within the first 14

[5] Foley (2011) investigates what factors influence how states set their welfare payment schedules. He finds that common considerations include monthly budget processes, administrative program aspects, and requests from retailers to reduce monthly demand fluctuations by staggering welfare payments. However, he also finds that for many programs and jurisdictions the payment schedules were set a long time ago and why they were set the way they are is not documented. Foley's (2011) findings thus suggest that there are no clear systematic reasons for the variation in issuance periods across states.

[6] Studies based on other populations have also found that individual expenditure behavior is sensitive to the timing of income receipt (see, e.g., Johnson et al. 2006; Shapiro and Slemrod 1995; Stephens 2003; 2006).

days of the food stamp cycle. After 21 days, they have exhausted almost all of their food stamps, having only 9 percent left. On the last day, 97 percent of all benefits are spent. Cole and Lee (2005) also find that the food stamp redemption patterns vary very little across states, community characteristics, and household characteristics. Additionally, the patterns appear to be relatively stable over time (U.S. Department of Agriculture 2006).

Going beyond the analysis of expenditure patterns, Shapiro (2005) provides further evidence for the monthly variation in food stamp households' financial circumstances. He exploits plausibly exogenous variation in the time since the last food stamp receipt across individuals in survey data such that his results are unlikely to be driven by unobserved heterogeneity. In addition to a decline in food expenses as the food stamp cycle progresses, Shapiro (2005) estimates that the caloric intake of food stamp household members goes down by 10 to 15 percent between the beginning and end of the cycle. Furthermore, he finds evidence that the food stamp recipients' desperation for money rises over the monthly cycle: the more days that have passed since the last food stamp payment, the more likely they are to hypothetically accept less than $50 today versus $50 dollars in a month. At the same time, the smallest amount of cash that they would be willing to accept today decreases over the cycle. When asked about their willingness to accept less than $50 today versus $50 *in a week*, his analysis yields that the food stamp households have a higher probability of accepting the option of less than $50 today in the last week of the cycle compared with the rest of the food stamp cycle.

In sum, the results of the studies discussed indicate that the financial resources of food stamp households decrease as the monthly food stamp cycle progresses, leading to particularly poor financial circumstances at the end of the cycle.[7]

1.3 Empirical Strategy

This section explains the empirical strategy to estimate the short-run effect of the financial scarcity at the end of the food stamp cycle on self-assessed physical health. The

[7]This notion is further supported by a large qualitative study of the lives of food stamp recipients by Edin et al. (2013).

strategy exploits the random interview day assignment in ATUS and the variation in food stamp issuance periods across states. The next section subsequently describes the data I use in more detail and reports the results from randomization checks.

ATUS does not contain the actual day when a food stamp recipient receives his or her food stamps. For this reason, I select all food stamp recipients from the states that issue all of their food stamps early in the month for the first estimation approach. I call this group of states the early states and define a state to be an early state if its food stamp issuance period starts before the fifth day of each month and lasts at most ten days.[8] Table A.1 in the appendix lists all of the early states. Due to the payment of food stamp benefits early in the month, I know that in the early states the food stamp cycle coincides approximately with the actual calendar month. Therefore, I also know that the individuals in the early states experience financial scarcity approximately at the end of the calendar month.

The first estimation approach, which I call early states approach, thus compares food stamp recipients from the early states interviewed at the end of the calendar month with food stamp recipients from the same states *not* interviewed at the end of the calendar month. I call the latter period the beginning of the month.[9] The regression equation that I estimate using a linear probability model is:

$$y_i = \alpha + \beta\, end_i + \gamma X_i + \epsilon_i, \tag{1.1}$$

where y_i is a dummy that equals one if individual i reports fair or poor physical health and zero otherwise, i.e., if i reports excellent, very good or good physical health. The dummy variable end_i is equal to one if individual i was interviewed at the end of the month, which I define in the main specification as the last ten days of the calendar month, and zero otherwise. To improve precision and assert that the randomization

[8] When defining the early states, there is a trade-off between restricting the food stamp issuance period to a smaller time window at the beginning of the month and sample size. My early states definition tries to balance this trade-off. Section 1.7 shows that the main results are robust to an alternative early states definition.

[9] Because the individuals' food stamp cycle only approximately coincides with the calendar month, there may be individuals in the end-of-month group that have not actually reached the end of their monthly cycle, and vice versa. This may bias the effect estimates towards zero.

procedure of the survey worked out well, Equation (1.1) also contains a vector of control variables X_i. Apart from standard demographic variables, X_i includes dummies for individual i's weight based on the Body Mass Index (BMI) and a dummy for whether he or she is disabled. Additionally, Vector X_i contains year$\times$month and state of residence dummies as well as a dummy for whether the interview took place on the weekend. The notes for Table 1.3 list all of the covariates. ϵ_i is the zero-mean error term. The standard errors are clustered at the state–quarter level.[10] The coefficient of interest is β, which corresponds to the effect of the end-of-cycle financial scarcity on the probability of reporting fair or poor physical health.

In general, one major threat to obtaining an unbiased estimate for the effect of interest using Equation (1.1) is selection based on unobserved individual characteristics. For example, if individuals interviewed at the end of the month are generally more pessimistic about their health or have generally worse health, inducing them to assess their health to be worse compared with those interviewed at the beginning of the month, then the effect estimate may be upward biased, suggesting a more detrimental effect of financial scarcity than is actually the case. However, because the individuals used in the analysis are randomly assigned to the interview period, resulting from the interview day randomization in ATUS, this type of selection does not threaten the estimation. The random assignment breaks all correlations between the end-of-month dummy and characteristics of individuals that may determine their self-assessed physical health apart from the experience of financial scarcity. However, if there are factors other than financial scarcity that differ systematically between the end of the month and the beginning of the month, and which affect self-assessed physical health, then the estimation based on Equation (1.1) will give misleading results. One such factor could be, for example, that individuals interviewed at the end of the month are exhausted from a long month's work, inducing them to report worse health than they would have reported otherwise if interviewed earlier in the month.

To account for such potential factors, I extend the early states approach using a difference-in-differences (DID) approach. The DID approach additionally protects against bias that could result from an imperfect random interview period assignment that is the same for

[10]The conclusions from the main analysis are robust to clustering at alternative levels, such as at the state–month level, and to using unclustered (heteroscedasticity-consistent) standard errors.

the treatment and control group. However, the randomization checks indicate that imperfect randomization is unlikely to be a concern in the estimations.

The idea of the DID approach is to compare the 'beginning of the month–end of the month' change in self-assessed physical health between the food stamp recipients from the early states and a suitable control group that is unlikely to experience scarcity at the end of the month, which I describe below. The regression equation that I estimate via OLS is:

$$y_i = \alpha + \delta\, end_i + \eta\, fsp_early_i + \beta\, end_i \times fsp_early_i + \gamma X_i + \epsilon_i, \tag{1.2}$$

where, as in the early states approach, y_i is a dummy that is equal to one if individual i reports fair or poor physical health and zero otherwise and dummy end_i equals one if individual i is interviewed in the last ten days of the month and zero otherwise. The dummy fsp_early_i takes on the value one if individual i is a food stamp recipient from the early states and zero otherwise. Vector X_i contains the same variables as for the early states approach. ϵ_i is the zero-mean error term and the standard errors are again clustered at the state–quarter level. Coefficient β is the effect of interest.

For estimations based on Equation (1.2) to yield an unbiased effect estimate, the crucial assumption that is required to hold is the parallel trends assumption. In the present case, the assumption states that the individuals from the early states would experience the same average change in self-assessed health between the beginning of the month and end of the month in the absence of scarcity as the average change in self-assessed health between the beginning of the month and end of the month for the control group (conditional on X_i). When thinking about control groups for which the parallel trends assumption may hold, two groups of individuals come especially to mind. I estimate Equation (1.2) with both of these groups separately.

The first control group consists of all of the food stamp recipients that are not from the early states. As Table A.1 in the appendix shows, many of the non-early states stagger their food stamp issuance over a longer time span than the early states. For this reason, I call this group of states the staggering states. In addition, many of the staggering states

start issuing food stamps more towards the middle of the month and thus a bit later than the early states. The greater staggering and later food stamp issuance implies that the time since the last food stamp receipt does not change as much on average between the beginning of the month and end of month for the individuals from the staggering states as for the individuals from the early states, and that in neither of the two periods of the month the food stamp cycle for the staggering states' individuals has on average reached its very end. The group of food stamp recipients from the staggering states is, therefore, on average not only at the beginning of the month but also at the end of the month unlikely to experience the financial scarcity that the group of food stamp recipients from the early states has to face at the end of the month. At the same time, it may be plausible that the parallel trends assumption holds because both groups consist of individuals who participate in the FSP.

The second control group consists of all non-food stamp recipients from the early states. As non-food stamp recipients are on average wealthier than food stamp recipients, it is unlikely that they experience the financial scarcity of the food stamp recipients. At the same time, it could be argued that the parallel trends assumption is fulfilled because both groups come from the early states.

If there truly are factors other than the experience of financial scarcity that vary systematically between the beginning of the month and end of the month, then it is a priori not clear which of the two DID estimations yields more reliable effect estimates. This is because the estimations based on the two controls groups need not necessarily give similar results and because there is uncertainty about which of the two control groups is more suitable. If there are no such alternative factors present, however, then the DID approach using either of the two control groups, and the early states approach, should yield similar and valid effect estimates. The estimations below suggest that the latter situation is the case.

1.4 Data

1.4.1 American Time Use Survey

For the empirical analysis, I use data from the American Time Use Survey. ATUS data are well suited for investigating the research question of interest. The survey contains information on food stamp receipt and self-assessed physical health (for selected years), and, unlike any other potentially suitable survey, it assigns individuals randomly to their interview day. The random interview day assignment allows me to adopt the outlined estimation strategy.

The main purpose of ATUS is to obtain nationally representative estimates of how individuals in the United States spend their time. The survey is conducted by the US Census Bureau and is sponsored by the Bureau of Labor Statistics. Since 2003, annual ATUS waves are available, in most years containing 12,000–13,000 observations.

ATUS is based on a random sample drawn from the households that have recently finished their last interview for the Current Population Survey (CPS). From every drawn household, one household member aged 15 or older is randomly selected to be questioned in ATUS. The interview is conducted via telephone and takes place 2–5 months after the last CPS interview. The interview day is randomly assigned, using a procedure that can be described as follows. For a given designated respondent, the month of the interview is randomly selected. In this month, then the interview week and subsequently the day of the week are randomly selected.[11] Each designated respondent is notified in advance about the day when the interview is scheduled to take place. If an individual is unable to do the interview on the specified date, then he or she will be contacted on the same weekday as the one of the originally planned interview day in the following up to seven weeks. Each ATUS respondent is interviewed only once.

Apart from information about how individuals use their time, ATUS includes limited demographic information. The survey can be linked to the CPS, which increases the

[11] The random assignment is performed so that the number of designated respondents is evenly spread across the weeks of the year and so that 25 percent of all respondents are allocated to a Sunday and Monday, respectively, and 10 percent to every other day of the week. The respondents are asked about their time use the day before the interview. The day of the week allocation thus implies that 50 percent of all individuals report about a weekend day.

number of variables available. Additionally, in 2006–2008, all ATUS respondents were asked to assess their physical health as part of the supplementary Eating and Health module, which contains a small number of health-related questions. The wording of the question is 'In general, would you say that your physical health is Excellent, Very Good, Good, Fair, or Poor?' The module also asked the respondents whether they or anyone else in their household received food stamp benefits in the past 30 days. I refer to individuals who answer this question with 'yes' as food stamp recipients.

1.4.2 Sample and Descriptive Statistics

The analysis sample consists of data from the ATUS waves 2006–2008, which contain the required variables, supplemented by data from the respondents' last CPS interview. For the sample, I select all individuals that belong to one of the three groups used in the empirical strategy and who do not have missing information for the analysis variables.

This selection procedure yields 1,322 food stamp recipients from the early states, 997 food stamp recipients from the staggering states, and 18,592 non-food stamp recipients from the early states.[12] In all three groups, the observations are evenly distributed across the three years.

Table 1.1 presents descriptive statistics of the variables used in the main analysis for each of the three groups of individuals separately. The table indicates that the two groups of food stamp recipients are overall very similar. The only notable differences in variable means are a nine percentage points lower share of black people and an eight percentage points higher share of individuals living in metropolitan areas for the recipients from the early states relative to the recipients from the staggering states.

Conversely, the non-food stamp recipients are quite different from the food stamp recipients. This is not surprising considering that the FSP especially targets low-income individuals. Comparing the means for the three groups shows that, for example, the food stamp recipients are overall less educated, have a higher share of disabled people and fewer of them are employed. Additionally, in food stamp households there is

[12]Due to missing values, I dropped approximately 11, 14, and 9 percent of all food stamp recipients from the early states, food stamp recipients from the staggering states, and non-food stamp recipients from the early states, respectively.

Table 1.1. Descriptive Statistics for the Groups of Individuals Used in the Analysis

	Food stamp recipients				Non-food stamp recipients	
	Early states		Staggering states		Early states	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Outcome variable						
Fair or poor physical health	0.399	0.490	0.415	0.493	0.144	0.351
Regressor of interest						
End	0.303	0.460	0.305	0.461	0.298	0.457
Individual characteristics						
Age in years	43.107	16.620	44.720	17.150	46.382	17.210
Black	0.269	0.443	0.360	0.480	0.108	0.310
Male	0.287	0.452	0.302	0.459	0.467	0.499
Employed	0.368	0.482	0.392	0.488	0.682	0.466
Retired	0.113	0.316	0.131	0.338	0.154	0.361
Disabled	0.257	0.437	0.243	0.429	0.031	0.174
Less than high school	0.319	0.466	0.332	0.471	0.138	0.345
High school	0.363	0.481	0.381	0.486	0.252	0.434
Some college	0.181	0.385	0.166	0.373	0.182	0.386
College	0.119	0.324	0.111	0.315	0.304	0.460
Advanced degree	0.018	0.134	0.009	0.095	0.123	0.329
Underweight	0.021	0.144	0.018	0.133	0.016	0.127
Normalweight	0.300	0.459	0.262	0.440	0.375	0.484
Overweight	0.305	0.461	0.306	0.461	0.360	0.480
Obese	0.374	0.484	0.414	0.493	0.249	0.433
Household characteristics						
Spouse/partner present in household	0.282	0.450	0.292	0.455	0.551	0.497
Low-income household	0.883	0.322	0.888	0.316	0.251	0.433
No. of adults in household	1.711	0.912	1.696	0.848	1.883	0.775
No. of children in household	1.356	1.394	1.345	1.420	0.891	1.113
Metropolitan area	0.739	0.439	0.814	0.389	0.814	0.389
Observations	1,322		997		18,592	

Notes: Author's calculations based on 2006–2008 ATUS data supplemented by CPS data. The outcome variable fair or poor physical health equals one if the individual reports fair or poor physical health and zero if the individual reports excellent, very good, or good physical health. The dummy end is one if the individual is interviewed in the last ten days of the month and zero otherwise. The dummy low-income household equals one if the individual lives in a household with a monthly gross income equal to or below 185 percent of the poverty line and zero otherwise. The weight dummies based on the classification of the World Health Organization (2000) take on the value one if the following conditions hold and zero otherwise: Underweight: BMI$<$18.5; normalweight: 18.5$\leq$BMI$<$25; overweight: 25$\leq$BMI$<$30; obese: BMI$\geq$30. Age in years takes on the value 80 for individuals aged 80 through 84 and the value 85 for individuals aged 85 and above. In all regression models controlling for age, I additionally include age squared as a covariate. The dummies black, male, disabled, metropolitan area, and the five education dummies were measured at the last CPS interview. All other listed variables were measured at the ATUS interview.

less often a partner or spouse present and the number of children is on average higher. There is also a marked difference in the mean for the outcome variable. The share of food stamp recipients assessing their physical health as fair or poor as opposed to good, very good, or excellent is approximately 40 percent. The corresponding share for the non-food stamp recipients is only about 14 percent.

To assess the notion that overall neither of the two groups of food stamp recipients experiences financial scarcity at the beginning of the month and that only the group of food stamp recipients from the early states experiences financial scarcity at the end of the month, Figure 1.1 displays the distribution of the mean number of days since the last food stamp receipt for both groups of food stamp recipients and both periods of the month.[13]

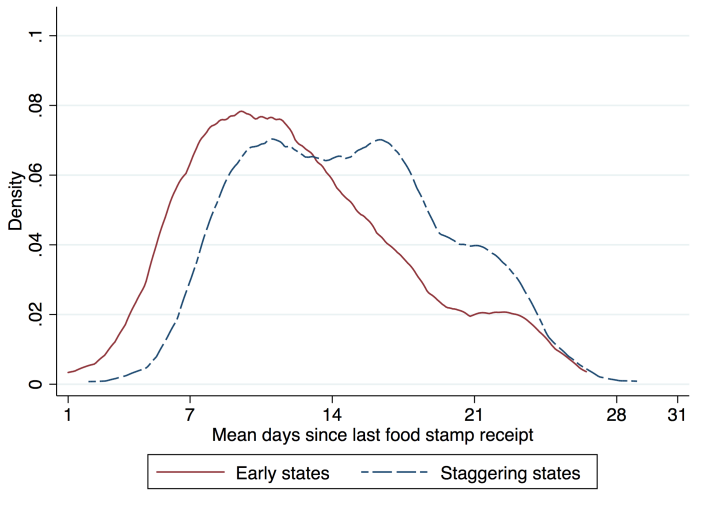
Panel A suggests that most of the food stamp recipients interviewed at the beginning of the calendar month from both groups have not yet reached the end of their monthly food stamp cycle. The mean time since the last food stamp receipt averaged over all individuals is 12.34 days for the recipients from the early states and 14.76 days for the recipients from the staggering states. Furthermore, only 9 percent of the early states' and only 14 percent of the staggering states' recipients have a mean time since their last food stamp receipt of 21 days, i.e., three weeks, or more.

Conversely, Panel B suggests that the food stamp cycle has progressed considerably further for the recipients from the early states interviewed at the end of the calendar month relative to their counterparts interviewed at the beginning of the month. The mean time since the last food stamp receipt averaged over all the early states' recipients is now 22.08 days, and 62 percent of these individuals have a mean time since the last food stamp receipt of at least 21 days. The food stamp cycle is also at a more advanced stage for the staggering states' recipients interviewed at the end of the calendar month, even though less so than for the recipients from the early states. The averaged mean time since the last food stamp receipt is 17.99 days and only for 26 percent of all individuals from the staggering states the mean time since the last food stamp receipt is at least 21

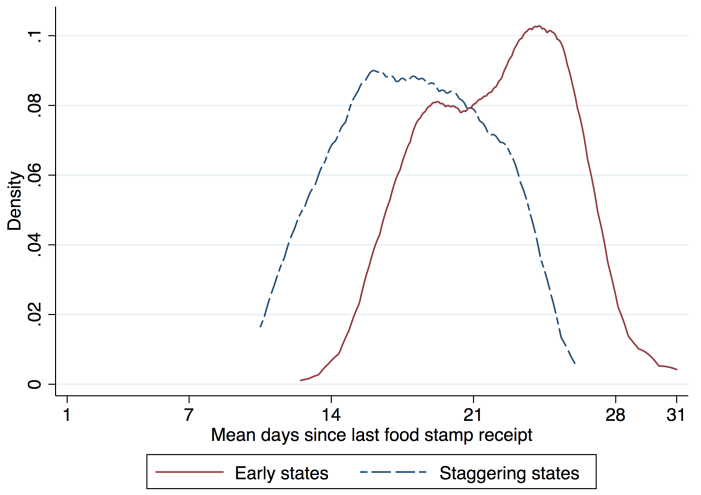
[13] I compute the mean days since the last food stamp receipt for individual i by taking the average distance between i's interview day and each possible day he or she could have received his or her last food stamps based on the food stamp issuance dates for his or her state of residence. I use this relatively imprecise measure because I do not observe the actual individual food stamp delivery days.

Figure 1.1. Distribution of the Mean Number of Days since the Last Food Stamp Receipt

Panel A. Beginning of the Month



Panel B. End of the Month



Notes: The bandwidth is selected according to Silverman's rule of thumb. All individuals interviewed in the last ten days of the calendar month belong to the group 'end of the month'. All other individuals belong to the group 'beginning of the month'. See Footnote 13 for how I compute the mean days since the last food stamp receipt for a given individual.

days.

In sum, the insights from Figure 1.1 support the notion exploited in the empirical strategy. The food stamp cycle for the group of food stamp recipients from the early states coincides approximately with the calendar month and the group of food stamp recipients from the staggering states has not yet reached the end of the food stamp cycle in either of the two periods of the month.

1.4.3 Randomization Checks

In the following, I assess whether the interview period randomization, which results from the interview day randomization in ATUS, is successful in balancing the characteristics between the individuals interviewed at the beginning of the month and at the end of the month. An imperfect interview period randomization could threaten the validity of the empirical analysis.

Table 1.2 reports means for individual and household characteristics by period of the month for the three groups of individuals used in the analysis. In addition, the table displays p-values from t-tests, testing for differences in means by period of the month for each listed variable, and F-tests, testing whether all mean differences within each group are jointly equal to zero. For each of the three groups, the table shows that there are overall only small differences in means between the individuals interviewed at the beginning of the month and at the end of the month. For 57 out of the 60 pairwise mean comparisons, the t-tests fail to reject the hypothesis of equal means at the 10 percent level.[14] Furthermore, the F-tests are far from rejecting that all mean differences within each group are jointly equal to zero at the 10 percent level, as the second to last row in the table shows. The corresponding p-values are 0.853, 0.475, and 0.652. The balance checks in Table 1.2 thus suggest that the interview period randomization successfully balances the characteristics between the two periods of the month.

[14]The three exceptions are as follows: for the food stamp recipients from the staggering states, the 5.7 percentage point difference in the share of retired people is significant at the 5 percent level; and for the non-food stamp recipients, the 0.5 year age difference and 1.4 percentage point difference in the share of individuals who have a normal weight is significant at the 10 percent level. The set of covariates in the estimations includes these three variables.

Table 1.2. Balance Checks

	Food stamp recipients						Non-food stamp recipients		
	Early states			Staggering states			Early states		
	Beginning of the month	End of the month	p-value	Beginning of the month	End of the month	p-value	Beginning of the month	End of the month	p-value
Individual characteristics									
Age	43.445	42.329	0.262	45.238	43.539	0.150	46.528	46.037	0.075
Black	0.273	0.259	0.620	0.351	0.382	0.349	0.108	0.109	0.783
Male	0.294	0.269	0.357	0.303	0.299	0.907	0.465	0.472	0.393
Employed	0.357	0.392	0.235	0.387	0.405	0.595	0.683	0.680	0.746
Retired	0.113	0.112	0.970	0.149	0.092	0.015	0.156	0.150	0.282
Disabled	0.261	0.249	0.668	0.231	0.270	0.188	0.031	0.032	0.721
Less than high school	0.322	0.312	0.700	0.323	0.352	0.375	0.138	0.139	0.788
High school	0.370	0.347	0.412	0.388	0.365	0.491	0.249	0.259	0.133
Some college	0.175	0.195	0.393	0.173	0.151	0.395	0.183	0.182	0.948
College	0.113	0.132	0.320	0.105	0.125	0.364	0.308	0.297	0.135
Advanced degree	0.020	0.015	0.567	0.010	0.007	0.589	0.123	0.123	0.921
Underweight	0.024	0.015	0.301	0.020	0.013	0.442	0.016	0.017	0.727
Normalweight	0.302	0.297	0.853	0.255	0.276	0.490	0.370	0.384	0.073
Overweight	0.296	0.324	0.314	0.316	0.283	0.297	0.362	0.354	0.306
Obese	0.378	0.364	0.635	0.408	0.428	0.570	0.251	0.245	0.332
Household characteristics									
Spouse/partner present in household	0.282	0.282	0.985	0.294	0.286	0.794	0.550	0.552	0.813
Low-income household	0.879	0.890	0.575	0.879	0.908	0.181	0.248	0.257	0.209
No. of adults in household	1.698	1.741	0.436	1.687	1.717	0.605	1.883	1.884	0.912
No. of children in household	1.325	1.429	0.211	1.303	1.441	0.159	0.885	0.907	0.215
Metropolitan area	0.745	0.726	0.466	0.820	0.803	0.526	0.814	0.814	0.894
Test for H_0: all differences in means are jointly equal to zero			0.853			0.475			0.652
Observations	921	401		693	304		13,049	5,543	

Notes: Author's calculations based on 2006–2008 ATUS data supplemented by CPS data. The table reports the means for the listed variables by interview period of the month. All individuals interviewed in the last ten days of the calendar month belong to the group 'end of the month'. All other individuals belong to the group 'beginning of the month'. The p-values for each variable are from t-tests for H_0: *no difference in means between the beginning-of-month and end-of-month group*. The test for H_0: *all differences in means jointly equal to zero* tests for each of the three groups of individuals separately whether the differences in means between the beginning-of-month and end-of-month group for all the listed variables are jointly equal to zero. This test is an F-test based on a regression of the end of-month dummy on all listed variables that tests whether the coefficients on all the included regressors are jointly equal to zero. The dummies normalweight and college are omitted due to multicollinearity. For variable definitions, see the notes for Table 1.1.

Additionally, note that the DID approach protects against bias due to imperfect random interview period assignment that is the same for the food stamp recipients from the early states and the respective control group. For this reason, especially an imperfect interview period randomization specific to one of the three groups used in the analysis could pose a threat to the validity of the estimations. For example, one could be concerned that especially food stamp recipients from the early states do not want to participate in the survey at the end of the month, due to financial scarcity. However, the means for the dummy variable end in Table 1.1 indicate that almost exactly 30 percent of all individuals in each of the three groups are interviewed at the end of the month. Group-specific interview period selection is, therefore, unlikely to threaten the validity of the analysis, considering these almost identical shares in addition to the findings of the balance checks.[15]

1.5 Main Results

Table 1.3 presents the main estimates for the short-run effect of the financial scarcity at the end of the food stamp cycle on self-assessed physical health.

Column (1) reports the effect estimate from the early states approach, which compares the food stamp recipients from the early states interviewed at the beginning and end of the calendar month, without controls. The estimation yields that the experience of financial scarcity increases the probability of reporting fair or poor physical health by 5.3 percentage points. The estimate is significant at the 10 percent level. Adding the control variables to the model in Column (2) increases the estimate moderately to 7.2 percentage points. The R^2 goes up considerably, from 0.003 to 0.342, and the effect estimate becomes more precisely estimated, being now significant at the 1 percent level.

Columns (3)–(6) report the main estimates from the DID approach, using as a control group a group of individuals that is unlikely to experience financial scarcity at the end of the calendar month.

In the DID models, the coefficient on the dummy variable end gives the change in the

[15]The shares are also close to the share of individuals assigned to the end-of-month period among all designated ATUS respondents, which is 29 percent.

Table 1.3. Estimates for the Effect of Financial Scarcity on Self-Assessed Physical Health

	Early states approach		Difference-in-differences approach			
Control group			Staggering states food stamp recipients		Early states non-food stamp recipients	
	(1)	(2)	(3)	(4)	(5)	(6)
End	0.053*	0.072***	−0.006	−0.004	−0.003	−0.002
	(0.027)	(0.022)	(0.032)	(0.025)	(0.006)	(0.005)
Fsp_early	–	–	−0.034	−0.033	0.238***	0.033**
			(0.024)	(0.021)	(0.016)	(0.014)
Fsp_early × end	–	–	0.059	0.076**	0.056*	0.067***
			(0.042)	(0.034)	(0.030)	(0.024)
Individual controls		✓		✓		✓
Household controls		✓		✓		✓
Time controls		✓		✓		✓
State controls		✓				✓
R^2	0.003	0.342	0.002	0.296	0.030	0.241
Observations	1,322	1,322	2,319	2,319	19,914	19,914

Notes: Standard errors clustered at the state–quarter level are in parentheses. The estimations are based on 2006–2008 ATUS data supplemented by CPS data. All models are estimated via OLS. The outcome variable is a dummy variable that equals one for individual i if i reports fair or poor physical health and zero otherwise. The dummy end equals one for individual i if i was interviewed in the last ten days of the calendar month and zero otherwise. Fsp_early is one for individual i if i is a food stamp recipient from the early states and zero otherwise. The set of individual controls consists of the variables listed under individual characteristics in Table 1.1. The dummies college and normalweight are omitted due to multicollinearity, and age squared is additionally added. The set of household controls consists of the variables listed under household characteristics in Table 1.1. The time controls are a full set of year×month dummies and a dummy that equals one for individual i if i's interview took place on the weekend and zero otherwise. The state controls are dummies for each state. For the food stamp recipients from the early states interviewed at the beginning of the month, the probability of reporting fair or poor physical health is 0.383.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

probability of reporting bad physical health between the beginning and the end of the month for each of the two respective control groups. For either of the control groups, with and without controls, the estimated coefficient is far from significant at the 10 percent level and is close to zero. This suggests that there are no factors other than the experience of financial scarcity that influence the individuals' self-assessed physical health at the end of the month.

The interaction term fsp_early×end gives the financial scarcity effect estimate in the DID estimations. Using the food stamp recipients from the staggering states as the control group, the estimated effect without controls is 5.9 percentage points; however this is imprecisely estimated. After adding controls, the estimate becomes significant at the 5 percent level and is now 7.6 percentage points. The DID model that uses the non-food stamp recipients from the early states as the control group yields an effect estimate of 5.6 percentage points without controls, which is significant at the 10 percent level. With control variables, the model gives an estimate of 6.7 percentage points, and is significant at the 1 percent level. The DID estimates with and without controls are thus very similar to their corresponding estimates from the early states approach. This could already be expected, considering the small estimated changes in the probability of reporting bad physical health between the beginning and end of the month for both control groups.

In summary, all of the specifications from both estimation approaches indicate a detrimental short-run effect of the end-of-cycle financial scarcity on self-assessed physical health. After adding controls, the regressions yield that the financial scarcity increases the probability of reporting fair or poor physical health by around seven percentage points. The probability of reporting fair or poor physical health for the food stamp recipients from the early states interviewed at the beginning of the month is 38.3 percent. Relative to this baseline probability, the effect appears quite sizable, corresponding to an increase of around 18 percent.

1.6 Potential Mechanisms

This section discusses two potential mechanisms through which the food stamp recipients' financial scarcity may affect self-assessed physical health. One potential mechanism for the effect could be that the experience of financial scarcity increases the individuals' levels of stress. Indeed, Edin et al. (2013) found evidence that the lack of financial resources may be a stressful experience for many food stamp recipients, and a number of studies have found that stress may lead to negative health consequences, such as headaches and back pain (see Benson and Proctor 2010). Anecdotal evidence from a Washington Post article further supports the notion that stress could be a relevant mechanism (see Saslow 2013). The article is about a food stamp recipient who reports getting anxiety headaches at the end of the monthly food stamp cycle when her financial resources are exhausted. Unfortunately, ATUS does not contain any direct measure of stress to explore this mechanism. However, the survey does include the time use category 'sleeplessness', which captures the reported number of minutes that an individual was sleepless on the day before the interview. Examples for the category include lying awake and tossing and turning.[16] A number of studies have documented associations between stress and sleep difficulties (see, e.g., Åkerstedt 2006; Kahn et al. 2013). Thus, if stress is a mechanism for the financial scarcity effect, then a positive impact of financial scarcity on individuals' reported sleeplessness may seem plausible.[17]

Table 1.4 presents estimates for the effect of financial scarcity on reported sleeplessness from regressions that are analogous to the ones in the main analysis. Panel A shows that the estimated financial scarcity effect on the probability of reporting any sleeplessness is positive and significant at conventional levels in all estimations. Each regression yields a similar effect estimate between around three and four percentage points. This magnitude appears sizable, with about 49 to 66 percent relative to the 6.1 percent probability of reporting any sleeplessness for the food stamp recipients from the early states at the beginning of the month. Panel B gives estimates for the financial scarcity effect on

[16] It appears unlikely that individuals are generally able to recall the exact number of minutes that they were sleepless. Nevertheless, the reported sleeplessness may still serve as a rough measure for individuals' actual amount of time that they spent in sleeplessness.

[17] A lack of sleep has also been found to be associated with worse health (see, e.g., Dinges et al. 1997; Paiva et al. 2015). A potential effect of financial scarcity on self-assessed physical health via increased stress could, therefore, also go through sleeplessness itself.

Table 1.4. Estimates for the Effect of Financial Scarcity on Reported Sleeplessness

	Early states approach		Difference-in-differences approach			
Control group			Staggering states food stamp recipients		Early states non-food stamp recipients	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Outcome: Dummy equal to 1 if an individual reports any sleeplessness						
End	0.034**	0.040**	−0.008	−0.011	−0.003	−0.003
	(0.017)	(0.017)	(0.016)	(0.015)	(0.004)	(0.004)
Fsp_early	–	–	−0.010	−0.013	0.010	−0.013
			(0.013)	(0.014)	(0.008)	(0.009)
Fsp_early × end	–	–	0.042*	0.044**	0.037**	0.039**
			(0.023)	(0.022)	(0.017)	(0.017)
R^2	0.004	0.088	0.002	0.032	0.001	0.012
Panel B. Outcome: Number of minutes of reported sleeplessness						
End	5.704**	5.863**	−0.045	−0.150	−0.301	−0.347
	(2.367)	(2.313)	(1.839)	(1.770)	(0.427)	(0.423)
Fsp_early	–	–	−1.132	−1.272	0.729	−2.144**
			(1.440)	(1.525)	(0.873)	(1.034)
Fsp_early × end	–	–	5.748*	5.883**	6.004**	6.186**
			(2.995)	(2.928)	(2.389)	(2.373)
R^2	0.006	0.071	0.004	0.034	0.001	0.017
Individual controls		✓		✓		✓
Household controls		✓		✓		✓
Time controls		✓		✓		✓
State controls		✓				✓
Observations	1,322	1,322	2,319	2,319	19,914	19,914

Notes: Standard errors clustered at the state–quarter level are in parentheses. The estimations are based on 2006–2008 ATUS data supplemented by CPS data. All models are estimated via OLS. The dummy end equals one for individual i if i was interviewed in the last ten days of the calendar month and zero otherwise. Fsp_early is one for individual i if i is a food stamp recipient from the early states and zero otherwise. For a description of the control variables, see the notes for Table 1.3. For the food stamp recipients from the early states interviewed at the beginning of the month, the probability of reporting any sleeplessness is 0.061 and the average number of minutes of reported sleeplessness is 4.643.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

the number of minutes that an individual reports being sleepless. In all regressions, the estimated effect is approximately six minutes. This corresponds to a 130 percent increase relative to the average minutes of sleeplessness of 4.64 for the food stamp recipients from the early states at the beginning of the month. All of the estimates are significant at conventional levels. The findings in Table 1.4 thus suggest that increased stress may indeed be a mechanism through which the financial scarcity at the end of the food stamp cycle affects self-assessed physical health.

Considering that the FSP targets individuals that are threatened by food insecurity, another mechanism for the effect of interest may be that the food stamp recipients experience hunger due to not having enough food as a result of the financial scarcity. This could lead to symptoms such as stomach ache and dizziness. As discussed in Section 1.2.2, Shapiro (2005) estimates a decline in caloric intake for food stamp recipients between the beginning and end of the monthly food stamp cycle. This finding suggests that hunger could also play a role. Unfortunately, ATUS does not include any suitable measure for hunger. A closer investigation of this potential mechanism may thus provide a fruitful opportunity for future research.[18]

[18]Regressions for the effect of financial scarcity on the time use category 'eating and drinking' yield estimates that are close to zero and, except in one case, are insignificant at the 10 percent level. This finding could be interpreted as evidence against hunger being a relevant mechanism. However, this finding could, for example, also result from the situation that individuals take more time to eat a possibly smaller amount of available food, potentially still being hungry. The estimates for the time spent eating and drinking, therefore, do not appear to provide conclusive evidence on the hunger mechanism. The time use category 'eating and drinking' cannot be divided into eating and drinking separately.

1.7 Robustness Analysis and Placebo Tests

1.7.1 Robustness Analysis

All of the main estimations, using two different approaches and two different control groups, yield similar results. To increase the confidence in the findings from the main analysis further, I assess the robustness of the estimates to alternative model specifications.

One concern could be that the estimates may be sensitive to the definition of the end-of-month period. To address this issue, I re-estimate the models from the main analysis using alternative end-of-month definitions. Table 1.5 presents estimates where I increase and decrease, respectively, the end-of-month window by three days relative to the original ten day definition. The table shows that the effect estimates remain quite stable and, if anything, generally behave as one might expect. For example, Column (2) gives that the estimated effect, based on the early states approach, decreases from 7.2 percentage points in the main specification to 5.3 percentage points when using the last 13 days of the month as the end-of-month definition. This moderate decrease appears plausible, as increasing the end-of-month window from ten to 13 days likely decreases the share of individuals experiencing financial scarcity in the end-of-month period.

The only effect estimate that loses significance at conventional levels is the estimate in Column (3) from the DID approach that uses the food stamp recipients from the staggering states as the control group and the last seven days of the month as the end-of-month definition. Nevertheless, the 5.0 percentage point estimate still indicates the presence of the financial scarcity effect. The decrease in the estimate relative to the corresponding one from the main specifications also appears plausible and may be explained as follows. As the end-of-month window becomes narrower, the share of food stamp recipients from the staggering states that experience financial scarcity in the end-of-month period likely increases, which leads the DID approach to underestimate the effect of interest. This notion is supported by the increased and now positive, yet still insignificant, change in the probability of reporting bad health between the beginning and end of the month for the staggering states' food stamp recipients, which is given by the coefficient

Table 1.5. Robustness Checks Using Alternative End-of-Month Definitions

	Early states approach		Difference-in-differences approach			
Control group			Staggering states food stamp recipients		Early states non-food stamp recipients	
	(1)	(2)	(3)	(4)	(5)	(6)
Fsp_early	–	–	−0.021	−0.038*	0.041***	0.032**
			(0.021)	(0.022)	(0.014)	(0.016)
Last 7 days of month						
End7	0.071**	–	0.024	–	−0.004	–
	(0.027)		(0.031)		(0.005)	
Fsp_early × end7	–	–	0.050	–	0.063**	–
			(0.041)		(0.027)	
Last 13 days of month						
End13	–	0.053**	–	−0.013	–	−0.003
		(0.026)		(0.024)		(0.005)
Fsp_early × end13	–	–	–	0.068*	–	0.056**
				(0.035)		(0.028)
R^2	0.341	0.340	0.296	0.295	0.240	0.241
Observations	1,322	1,322	2,319	2,319	19,914	19,914

Notes: Standard errors clustered at the state–quarter level are in parentheses. The estimations are based on 2006–2008 ATUS data supplemented by CPS data. All models are estimated via OLS. The outcome variable is a dummy variable that equals one for individual i if i reports fair or poor physical health and zero otherwise. The dummy end7 (end13) equals one for individual i if i was interviewed in the last seven (13) days of the calendar month and zero otherwise. Fsp_early is one for individual i if i is a food stamp recipient from the early states and zero otherwise. All regressions include the same control variables as the full specifications in Table 1.3.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

estimate for the end7 dummy.

As a further robustness check, Columns (2), (4), and (6) in Table 1.6 present estimates from regressions where I include dummies for each week of the month instead of the end-of-month dummy. Based on the early states approach, Column (2) shows that, relative to the first week of the month, the probability of reporting fair or poor physical health rises moderately through weeks two and three, and then becomes markedly higher with 9.1 percentage points in week four, which corresponds approximately to the definition of the end of the month in the main specifications. Estimating the financial scarcity effect by comparing the first and last week of the month thus increases the effect estimate by 1.9 percentage points, relative to the corresponding estimate from the main analysis. Only the estimate for the coefficient of the week four dummy is significant at conventional levels. Similarly, the analogous DID models also yield significant effect estimates that are larger than the corresponding ones from the main analysis.

In addition to the results for the specifications using week dummies, Table 1.6 displays estimates based on an alternative early states definition in Columns (1), (3), and (6). All states whose food stamp issuance period starts on the first day of the month and lasts at most ten days are now defined as the group of early states. All other states are defined as the group of staggering states. Using this alternative early states definition does not change the effect estimates much relative to the respective main regressions and all of the estimates remain significant at conventional levels.

1.7.2 Placebo Tests

As discussed in Section 1.3, the DID approach relies on the parallel trends assumption. In the DID model, using the food stamp recipients from the staggering states as the control group, one factor that would violate the parallel trends assumption would be the presence of a trend in self-assessed physical health between the beginning and end of the month that is specific to the individuals from the early states and staggering states, respectively. Analogously, a trend in self-assessed physical health specific to the food stamp recipients and non-food stamp recipients, respectively, would violate the parallel trends assumption when using the non-food stamp recipients from the early states as the

Table 1.6. Robustness Checks Using an Alternative Early States Definition and Week Dummies

	Early states approach		Difference-in-differences approach			
Control group			Staggering states food stamp recipients		Early states non-food stamp recipients	
	(1)	(2)	(3)	(4)	(5)	(6)
Fsp_early	–	–	−0.018	−0.043	0.040***	0.022
			(0.020)	(0.036)	(0.014)	(0.027)
End	0.069***	–	0.009	–	0.000	–
	(0.023)		(0.024)		(0.006)	
Fsp_early × end	–	–	0.059*	–	0.062**	–
			(0.034)		(0.025)	
Week-of-month dummies						
Week2	–	0.021	–	0.034	–	0.006
		(0.034)		(0.037)		(0.006)
Week3	–	0.035	–	−0.010	–	0.001
		(0.043)		(0.043)		(0.006)
Week4	–	0.091***	–	0.002	–	0.000
		(0.030)		(0.028)		(0.007)
Fsp_early × week2	–	–	–	−0.010	–	0.012
				(0.049)		(0.033)
Fsp_early × week3	–	–	–	0.040	–	0.030
				(0.060)		(0.042)
Fsp_early × week4	–	–	–	0.088**	–	0.076**
				(0.042)		(0.034)
Alternative early states definition	✓		✓		✓	
R^2	0.341	0.342	0.295	0.296	0.237	0.241
Observations	1,181	1,322	2,319	2,319	17,638	19,914

Notes: Standard errors clustered at the state–quarter level are in parentheses. The estimations are based on 2006–2008 ATUS data supplemented by CPS data. All models are estimated via OLS. The outcome variable is a dummy variable that equals one for individual i if i reports fair or poor physical health and zero otherwise. The dummy end equals one for individual i if i was interviewed in the last ten days of the calendar month and zero otherwise. Fsp_early is one for individual i if i is a food stamp recipient from the early states and zero otherwise. The dummies week2, week3, week4 are equal to one for individual i if i's interview took place in week two, three, and four, respectively, of the calendar month and zero otherwise. Week4 includes all remaining days after the third week of the month. All regressions include the same control variables as the full specifications in Table 1.3. The alternative early states definition defines all states as early states that have a food stamp issuance period which starts on the first day of the month and lasts at most ten days. All other states are defined as the staggering states.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

control group in the DID approach.

To test for such group-specific trends, and thus for the validity of the parallel trends assumption, I conduct two placebo tests. In both tests, I estimate the DID approach based on two groups of individuals that are unlikely to experience financial scarcity at the end of the month, yet who would exhibit one of the two types of group-specific trends. In the presence of group-specific trends, one would expect the regressions to yield an estimate for the coefficient on the DID interaction term that is significantly different from zero.

To test for trends specific to the early states and staggering states, respectively, I estimate the DID model based on the non-food stamp recipients from the early states and staggering states. To test for food stamp-/ non-food stamp recipient-specific trends, I estimate the DID model based on the food stamp and non-food stamp recipients from the staggering states.

Table 1.7 reports the results from the placebo tests. In each regression, the coefficient on the interaction term early×end and fsp×end, respectively, is close to zero and far from significant at the 10 percent level. This suggests that there are no group-specific trends present. Therefore, the placebo tests support the notion that the parallel trends assumption is valid in the DID estimations. In addition, the estimates for the end dummy coefficient are also close to zero and insignificant at the 10 percent level in all regressions. The placebo tests thus not only suggest that the parallel trends assumption holds, but also that actually none of the three placebo test groups experiences any change in the probability of reporting bad physical health between the beginning and end of the month.[19] This increases the confidence further in the validity of the findings from the main analysis.

[19]For all of the regressions, hypothesis tests also fail to reject at the 10 percent level that the two coefficients on the variables end, early×end and end, fsp×end, respectively, are jointly equal to zero.

Table 1.7. Placebo Tests

Comparison	Staggering states		Non-food stamp recipients	
	Food stamp recipients – non-food stamp recipients		Early states – staggering States	
	(1)	(2)	(3)	(4)
End	-0.009	-0.010	-0.009	-0.009
	(0.007)	(0.006)	(0.007)	(0.006)
Fsp	0.260^{***}	0.069^{***}	–	–
	(0.019)	(0.019)		
Fsp × end	0.003	-0.010	–	–
	(0.032)	(0.026)		
Early	–	–	-0.012	-0.001
			(0.007)	(0.005)
Early × end	–	–	0.006	0.007
			(0.009)	(0.008)
Individual controls		✓		✓
Household controls		✓		✓
Time controls		✓		✓
State controls		✓		
R^2	0.032	0.237	0.000	0.204
Observations	14,040	14,040	31,635	31,635

Notes: Standard errors clustered at the state–quarter level are in parentheses. The estimations are based on 2006–2008 ATUS data supplemented by CPS data. All models are estimated via OLS. The outcome variable is a dummy variable that equals one for individual i if i reports fair or poor physical health and zero otherwise. The dummy end equals one for individual i if i was interviewed in the last ten days of the calendar month and zero otherwise. The dummy fsp is one for individual i if i is a food stamp recipient and zero otherwise. The dummy early is one for individual i if i is from the early states and zero otherwise. For a description of the control variables, see the notes for Table 1.3.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

1.8 Conclusion

In this paper, I investigate the short-run effect of poor financial circumstances on health, using the variation in financial resources over the monthly food stamp cycle in a sample of food stamp recipients. To isolate the causal effect of interest, I exploit the random interview day assignment in the American Time Use Survey and the variation in food stamp issuance periods across states.

The empirical analysis suggests that poor financial circumstances can indeed have negative health consequences. I find that the financial scarcity experienced by food stamp recipients at the end of the monthly food stamp cycle increases the probability of reporting bad physical health by a considerable 18 percent relative to the baseline probability. Randomization checks, robustness checks, and placebo tests support the validity of this finding. By exploiting the time use information in ATUS, I find suggestive evidence that increased stress may be one mechanism through which this effect occurs.

From a policy perspective, the results suggest that measures taken to alleviate poverty may simultaneously improve the health of low-income individuals, potentially reducing the expenditures of public health care programs such as Medicaid. Furthermore, the results suggest that in the design of welfare programs, not only salient aspects (such as the benefit amount) but also more subtle features (such as the timing of payments) can be important. To mitigate particularly poor financial circumstances at the end of welfare payment cycles and their consequences, for example, it may be beneficial to distribute welfare payments in shorter time intervals (such as bi-weekly instead of monthly) to help individuals smooth their consumption. In the case of the FSP, this could be a viable option because the program issues its benefits via electronic cards. In addition, providing assistance to welfare recipients in managing their finances could also prove helpful to mitigate especially poor end-of-cycle financial situations. This may include informing individuals explicitly about the exact purpose of a given welfare program to avoid a potential misjudgment of the benefit amount. In the FSP, for example, many households believe that their food stamps are meant to cover all of the monthly food expenditures even though this is generally not the case (Edin et al. 2013).

The findings of this study suggest a number of avenues for future research. First, the

effect identified in this paper corresponds to the short-run response to a temporary particularly poor financial situation. To gain a broader understanding of the link between financial circumstances and health, it would also be important to examine the consequences of more permanent poor financial circumstances and to investigate longer-run responses. Second, it would be instructive to explore further the mechanisms behind the estimated effect, building on the first evidence presented in this study. Third, related to this, examining the extent to which the effect is driven by changes in more objective health measures or changes in health perceptions, using additional health data, would also be a fruitful subject for further research.

Appendix A

A.1 Additional Tables

Table A.1. Food Stamp Issuance Dates

State	Monthly issuance day(s)
Early states group	
Alaska	1
Arkansas	4, 5, 8, 9, 10, 11, 12, 13
California	1–10
Colorado	1–10
Connecticut	1–3
District of Columbia	1–10
Hawaii	1, 3, 5[a]
Idaho	1–5
Indiana	1–10
Iowa	1–10
Kansas	1–10
Kentucky	1–10
Michigan	1–9
Minnesota	4–13
Montana	2–6
Nebraska	1–5
Nevada	1
New Jersey	1–5
New York	1–9[b]
North Carolina	3–12
North Dakota	1
Ohio	1–10
Oklahoma	1
Oregon	1–9
Rhode Island	1
South Carolina	1–10
Tennessee	1–10
Vermont	1
Virginia	1
Washington	1–10
West Virginia	1–9
Wyoming	1–4

Notes: The table continues on the next page.

Table A.1. Continued

State	Monthly issuance day(s)
Staggering states group	
Alabama	4–18
Arizona	1–13
Delaware	5–11
Florida	1–15
Georgia	5–14
Illinois	1, 3, 8, 11, 14, 17, 19, 21, 23[a]
Louisiana	5–14
Maine	10–14
Maryland	6–15
Massachusetts	1–14
Mississippi	5–19
Missouri	1–22
New Hampshire	5
New Mexico	1–20
Pennsylvania	1-17[c]
South Dakota	10
Texas	1–15
Utah	5, 11, 15
Wisconsin	2, 3, 5, 6, 8, 9, 11, 12, 14, 15

Notes: The issuance dates are from Hamrick and Andrews (2016), who obtained the dates from the US Department of Agriculture. All of the dates are the actual issuance days for the years 2006–2008. A state belongs to the early states group if its food stamp issuance period starts before the fifth day of each month and lasts at most ten days. A state belongs to the staggering states group if it does not belong to the early states group. For further details on these definitions, see Section 1.3.

[a] There is uncertainty in the historical records about the exact dates.

[b] Weighted average issuance period for NY upstate and NY City, which have differing issuance days.

[c] Issuance days depend on the specific month.

Chapter 2

Heterogeneous Effects of Poverty on Cognition

2.1 Introduction

Many studies have documented associations between poverty and less beneficial behavior. For example, the poor are less likely than those with higher incomes to make use of preventive health services, and more likely to smoke cigarettes, play the lottery, and borrow more often at high cost.[1] Despite long-standing debates in economics and other disciplines, the reasons for such behavior remain unclear and the topic itself controversial. One recent hypothesis has focused on the financial circumstances of the poor and the potentially detrimental impact of these on cognition: In a sample of farmers from India, Mani et al. (2013) found that participants showed reduced cognitive performance before harvest, when poor, compared to after harvest, when rich. The authors suggested that a preoccupation with monetary concerns may leave the farmers before harvest with fewer mental resources available for other processes.[2]

In the only other study to have investigated this hypothesis empirically to date, Carvalho et al. (2016) assigned a sample of low-income US individuals randomly to perform a

[1] Use preventive health services (Ross et al. 2007), smoke cigarettes (Dube et al. 2009), play the lottery (Clotfelter et al. 1999), borrow at high cost (Bourke et al. 2012).

[2] See Bertrand et al. (2004; 2006) for a discussion of alternative views on the behavior of the poor.

number of cognitive tests before or after payday. The individuals surveyed before payday faced poorer financial circumstances than those surveyed after payday. However, the authors found no before-after differences in cognitive function in the full sample or selected subgroups. These mixed empirical findings, and the dearth of studies on this hypothesis in general, highlight the need to identify, at a more detailed level, the groups of individuals in which poor financial circumstances might have detrimental effects on cognitive function.

To contribute to this area of study, we therefore analyze heterogeneity in the effect of financial circumstances on cognition, focusing on identifying individuals in whom poorer financial circumstances have negative effects. To do so, we use data from the experiment conducted by Caravalho et al. (2016). For our heterogeneity analysis, we use the causal forest method by Athey et al. (2019), which was developed specifically to explore heterogeneous treatment effects in experiments. The method can be described as an adaptive nearest-neighbors approach that exploits ideas from the random forest machine learning literature to determine the relevant neighborhoods for estimating conditional average treatment effects at given points in the covariate space. Compared with traditional ordinary least squares (OLS) subgroup analyses, the causal forest method allows non-linear treatment effects to be estimated in a fully flexible way and circumvents the need to specify an interacted model, which may not always be straightforward (especially when the number of covariates is large). We examine effect heterogeneity using a rich set of 37 policy-relevant, pre-treatment covariates, including age, income, employment status, and measures of financial strain in the past. Our causal forest analysis proceeds in the following steps: First, we investigate which covariates are particularly relevant for heterogeneity in the treatment effect. Next, we examine how the effect varies across the most important variables. Subsequently, we study, in greater detail, the effect heterogeneity in regions of the covariate space where the previous step indicates particularly detrimental effects.

The results of our analysis suggest that there is strong effect heterogeneity in the two covariates age and income. For younger and elderly individuals who received a very low income around the time of the experiment, we find that the poorer financial circumstances before payday had detrimental cognitive effects. We verify this finding using

a second, independent, experiment conducted by Carvalho et al. (2016). Our results provide further evidence that there may be a causal effect of poverty on cognition. They also demonstrate the benefit of using the causal forest method to identify treatment effect heterogeneity that may have been overlooked in traditional subgroup analyses.

The remainder of this paper is structured as follows. Section 2.2 describes the experiment and our analysis sample. Section 2.3 explains the causal forest method. Section 2.4 presents average effect estimates for the full sample, the results of our heterogeneity analysis, and investigates the findings of our heterogeneity analysis in an independent experiment. Section 2.5 concludes.

2.2 Experiment and Data

2.2.1 Experiment

Carvalho et al. (2016) conducted their experiment twice, once among members of the RAND American Life Panel and then again among members of the GfK KnowledgePanel. Both are ongoing online panels with individuals aged 18 and over living in the United States. The authors restricted the sample for each experiment to individuals with an annual household income of $40,000 or less. For our analysis, we use the data from the GfK KnowledgePanel because it had the larger sample size, and because its share of compliers, i.e. the proportion of individuals who actually completed the survey before payday out of all individuals assigned to the before-payday group, was much higher. The following descriptions therefore pertain to the GfK KnowledgePanel.

The experiment consisted of a baseline survey and a follow-up survey, the former of which was used to determine individuals' paydays and the latter of which was used to administer the cognitive test. Individuals were randomly assigned to receive the survey with the cognitive test before or after payday.

In the baseline survey, individuals were asked to state all of the dates and amounts of payments that they (and their spouse) expected to receive during a reference period from 21 November to 20 December 2014. All individuals who did not give full information

about the number and dates of expected payments, or who reported expected payments for more than two different dates, were dropped from the sample.[3] Using this payment information, Carvalho et al. (2016) defined each individual's payday as follows: For individuals whose largest payment arrived at least two weeks after the previous payment, the date of the largest payment was set as the payday. For all other individuals, the payday was determined to be the payment date after the longest period without payment. If an individual's payments were fewer than two weeks apart, he or she was also excluded from the experiment.

The follow-up survey opened one week before payday for individuals assigned to the before-payday group and one day after payday for individuals assigned to the after-payday group. Carvalho et al. (2016) found that 98 percent of all individuals assigned to be surveyed before payday actually completed the survey before payday. Despite this high compliance rate, we follow Carvalho et al. (2016) in our analysis and estimate intention-to-treat effects, using the random assignment to the before-payday group as the regressor of interest.

The cognitive test in the follow-up survey was a version of the numerical Stroop task, which measures cognitive control. Participants are shown a number that consists of a repeated digit (e.g., 555). Subsequently, they must state, as quickly as possible, how many times the digit is repeated in the number rather than stating the digit itself – the correct answer in the example being three rather than five. The experiment by Carvalho et al. (2016) ran the Stroop task with 48 trials, and per trial each individual had, at most, five seconds to respond – otherwise the answer to the trial was coded as incorrect.

To confirm that the individuals actually experienced poorer financial circumstances before payday than they did after payday, the follow-up survey also collected information on individuals' cash holdings, checking and savings accounts balances, and total expenditures over the past seven days. Based on these measures, Carvalho et al. (2016) showed that the experiment had indeed created substantial variation in financial circumstances.[4] Table B.1 in the appendix presents results from our estimations that are

[3]The latter restriction was imposed to remove individuals for whom consumption smoothing may be easier.

[4]This finding is in line with previous research, which documented a sharp increase in caloric intake and expenditures at payday for certain groups of individuals (see, e.g., Mastrobuoni and Weinberg 2009;

analogous to Carvalho et al.'s (2016) for financial circumstances. These estimations yield very similar variation in financial circumstances in our sample, which is slightly smaller than Carvalho et al.'s (2016) sample, as explained in the next section.

2.2.2 Sample and Descriptive Statistics

For our analysis sample, we select all of the 2,723 individuals who were in Carvalho et al.'s (2016) full KnowledgePanel sample and subsequently drop all observations that are missing information on any of our analysis variables.[5] This selection procedure yields a sample of 2,480 individuals.

Table 2.1 presents the definitions and descriptive statistics for the cognition outcomes and treatment indicator. Our main outcome of interest is the number of correct answers per second that individual i gave over the entire Stroop task. This outcome captures the essence of the Stroop task's goal, which is to give correct answers to all trials as quickly as possible. Moreover, to gain an understanding of where the effect on our main outcome comes from, we include the numerator and denominator of our main outcome as additional outcomes: the number of correct answers over all 48 trials and the total time it took individual i to complete the entire Stroop task.[6] Table 2.1 shows that, on average, the individuals in our sample gave approximately 0.45 correct answers per second, provided about 43 correct answers in total (thereby responding correctly to most of the trials), and took approximately 100 seconds to finish the whole Stroop task. The mean for our regressor of interest, which is a dummy that is equal to one if an individual was randomly assigned to be surveyed before payday and zero otherwise, is almost exactly 50 percent. This is as expected considering the experiment's random assignment of individuals to the before-payday or after-payday group.

Table 2.2 reports descriptive statistics for the 37 covariates that we include in our heterogeneity analysis. All of these were collected before the follow-up survey, in which

Shapiro 2005; Stephens 2003; 2006).

[5] Additionally, we drop all individuals who were above the 0.99 quantile of the current income distribution in our full sample to remove potentially erroneous values. Given the definitions of our outcomes below, we also drop individuals who have missing information for any of the Stroop task's trials, i.e., who did not participate in all 48 trials of the task.

[6] Carvalho et al. (2016) conducted their Stroop task analysis at the individual$\times$trial level, using the outcomes response time per trial and a dummy which is one if an individual answered a trial correctly.

Table 2.1. Definitions and Descriptive Statistics for the Outcomes and Regressor of Interest

	Definition	Mean	Standard deviation
Outcomes			
Correct answers per second	Number of correct answers that individual i gave across all 48 Stroop task trials divided by the total time in seconds that it took i to complete the entire Stroop task.	0.446	0.143
Number of correct answers	Number of correct answers that individual i gave across all 48 trials of the Stroop task.	42.899	10.565
Total response time in seconds	Total time in seconds that it took individual i to complete the entire Stroop task.	100.476	22.816
Regressor of interest			
Before payday	= 1 if individual i was assigned to be surveyed before payday.	0.509	0.500

Notes: $N = 2,480$. The data are from the KnowledgePanel experiment by Carvalho et al. (2016).

the Stroop task was administered.[7] These covariates give information on many policy-relevant characteristics, such as an individual's race, education, employment status, and financial strain in the past. In addition to the annual household income at the time of the baseline survey, we include a measure of the (household) income that an individual received around the time of the experiment. We call this measure current income and construct it as the sum of all payments that an individual (and his or her spouse) expected to receive during the experiment's reference period (21 November to 20 December 2014).

Overall, Table 2.2 suggests that many individuals in the sample were of low socioeconomic status. For example, 41.4 percent of them had experienced financial hardship in the past 12 months, and almost half stated that they were living from paycheck to paycheck. Also, the annual household income dummies show that 41.1 percent of all

[7]Table B.2 in the appendix shows that the experiment's randomization procedure was successful in balancing the analysis covariates between the individuals interviewed before and after payday.

individuals had an annual household income of less than $20,000, and an average current income of approximately $1738.

Table 2.2. Descriptive Statistics for the Covariates

	Mean	Standard deviation
Age	55.947	17.423
Male	0.334	0.472
Household size	1.944	1.192
Household head	0.846	0.361
Children in household	0.167	0.373
Metropolitan area	0.804	0.397
Current income	1737.987	1321.136
Share of payday pay amount relative to current income	0.762	0.278
Financial strain		
Live from paycheck to paycheck	0.489	0.500
Caloric crunch	0.470	0.499
Liquidity constrained	0.503	0.500
Financial hardship	0.414	0.493
Marital status		
Married	0.335	0.472
Divorced	0.276	0.447
Widowed	0.139	0.346
Never married	0.250	0.433
Race		
White	0.761	0.426
Black	0.100	0.300
Hispanic	0.082	0.274
Other race	0.057	0.232

Notes: $N = 2,480$. The table continues on the next page.

Table 2.2 Continued

	Mean	Standard deviation
Employment status		
Working	0.287	0.452
Unemployed	0.063	0.244
Disabled	0.199	0.399
Retired	0.388	0.487
Other employment status	0.062	0.242
Education		
Less than high school	0.063	0.244
High school	0.254	0.435
Some college	0.417	0.493
College	0.266	0.442
Annual household income		
Less than $5,000	0.048	0.215
Between $5,000 and $10,000	0.100	0.300
Between $10,000 and $15,000	0.143	0.350
Between $15,000 and $20,000	0.120	0.325
Between $20,000 and $25,000	0.149	0.356
Between $25,000 and $30,000	0.143	0.350
Between $30,000 and $35,000	0.140	0.347
Between $35,000 and $40,000	0.156	0.363

Notes: $N = 2,480$. The data are from the KnowledgePanel experiment by Carvalho et al. (2016). The dummy category other race also includes individuals of mixed ethnicity; unemployed also includes temporarily laid off individuals, and working also includes self-employed individuals. In the order of the four financial strain variables listed, each respective dummy equals one if an individual i) agrees or strongly agrees with the statement 'I live from paycheck to paycheck', ii) had to reduce consumption at the end of a pay cycle, iii) could not, or would have to do something drastic to, raise $2,000 in one week for an emergency, iv) experienced at least one out of ten hardships related to not having enough money in the past 12 months. For the ten hardships, see Table C4 in the online appendix of Carvalho et al. (2016).

2.3 Methodology

The goal of our analysis is to study heterogeneity in the effect on cognition of poorer financial circumstances before payday. To do so, we estimate conditional average treatment effects using the causal forest method, which is based on the generalized random forest framework by Athey et al. (2019). The method is designed for studying treatment effect heterogeneity in experiments and can be described as an adaptive nearest-neighbors approach that uses a type of random forest technique to determine the weighting of observations in the estimation procedure.[8] This section describes the main idea of the causal forest. For technical details, see Athey et al. (2019).

To fix ideas, assume the following random effects model for individual i, $i = 1, \ldots, n$:

$$Y_i = \tau_i D_i + \epsilon_i, \tag{2.1}$$

where Y_i is one of our cognition outcomes, ϵ_i is i's outcome when assigned to be surveyed after payday, D_i is a dummy that equals one if individual i was assigned to be surveyed before payday, and τ_i corresponds to the effect of the financial circumstances before payday for individual i. Due to the random assignment of individuals to the before-payday or after-payday group, it further holds that D_i is independent of τ_i and ϵ_i.

Our quantity of interest is the conditional average treatment effect $\tau(x) = E(\tau_i | X_i = x)$, which in our case is the average effect of the financial circumstances before payday on cognition at a point x of the covariate vector X_i. For the estimation of $\tau(x)$, the causal forest method exploits the independence assumption of D_i and sets up two local moment equations. In the next step, the method obtains an estimate for $\tau(x)$ by fitting an empirical version of the local moment equations.[9] This procedure yields the causal forest estimator $\hat{\tau}$(x), which can be written as:

[8] For an introduction to random forests, see, for example, Hastie et al. (2009).
[9] See Appendix B.1 for details.

$$\hat{\tau}(x) = \sum_{\{i:D_i=1\}} \frac{\alpha_i(x)}{\sum_{\{i:D_i=1\}} \alpha_i(x)} Y_i - \sum_{\{i:D_i=0\}} \frac{\alpha_i(x)}{\sum_{\{i:D_i=0\}} \alpha_i(x)} Y_i, \tag{2.2}$$

where $\alpha_i(x)$ is a type of similarity weight, measuring individual i's relevance in the estimation of $\tau(x)$. Thus, the causal forest estimator estimates $\tau(x)$ by taking the difference in weighted average outcomes between the treated and untreated individuals.

To determine the weights $\alpha_i(x)$, the causal forest algorithm uses an approach that is based on the random forest method. The goal of Breiman's (2001) original random forest is to predict an outcome Y_i using covariates X_i by averaging over predictions from an ensemble of trees. Each tree is constructed by recursively splitting the covariate space into axis-aligned partitions, whereby at every step the split is chosen to maximize the tree's prediction accuracy. The prediction accuracy is typically evaluated using the mean squared error. After a stopping criterion has been reached, a single tree thus yields a partitioning of the covariate space into disjoint regions, or leaves, and its prediction for Y_i at point $X_i = x$ is calculated as the average Y_i over all observations that fall into the same leaf, based on their values in X_i, as the point x. For the construction of each tree, a different bootstrap sample of the data is used, and at every step only a random subset of all covariates is made available for splitting. Appendix B.2 shows an example of a single tree.

Now, for obtaining the weights $\alpha_i(x)$, the causal forest also grows an ensemble of trees using recursive partitioning. However, rather than averaging over predictions from the trees, the causal forest counts how many times individual i is in the same leaf as point x across all constructed trees, and derives $\alpha_i(x)$ based on this number. Specifically, for a set of trees $b = 1, \ldots, B$, the weight $\alpha_i(x)$ for individual i is computed as follows:

$$\alpha_i(x) = \frac{1}{B} \sum_{b=1}^{B} \frac{1\{i \in I_b(x)\}}{n_b(x)}, \tag{2.3}$$

where $I_b(x)$ is the set of all indices for the individuals that are in the same leaf as point x in tree b, and $n_b(x)$ is the number of individuals that fall into the same leaf as x in tree b. Thus, the more often individual i is in the same leaf as point x, the more weight i receives in estimating $\tau(x)$.

Compared with the random forest algorithm described above, the causal forest also uses a different splitting criterion for constructing the trees. The causal forest criterion is based on treatment effect estimates within the covariate space partitions, and, at a high level, implies that the algorithm seeks to maximize the treatment effect heterogeneity across partitions at every tree-splitting step. Athey et al. (2019) show that maximizing this criterion is related to improving the tree's expected accuracy in predicting treatment effects (rather than the outcome Y_i) at every step of the splitting procedure.

The causal forest also only allows splitting at every step based on a random subset of the covariates. In addition, the algorithm grows its trees on random subsamples of the data and implements a subsample splitting technique Athey et al. (2019) call honesty.[10] The idea behind the honest approach is to split a given subsample randomly into two roughly equally sized parts. The tree structure is subsequently grown on one of the two subsample parts, and the resulting structure is used to determine which individuals in the other subsample part are in the relevant neighborhood for estimating $\tau(x)$. Intuitively, the approach implies that observation i's outcome Y_i is not able to influence the construction of its weight $\alpha_i(x)$. This guards against spuriously extreme Y_i values obtaining unduly large influence in the data-driven weight calculation and thereby confounding the estimate for $\tau(x)$.

Athey et al. (2019) show that the causal forest estimates are consistent and asymptotically normally distributed, and derive bootstrap standard errors that allow for constructing valid confidence intervals.

We conduct our analysis in R, using the package grf by Tibshirani et al. (2018). The package implements the causal forest estimator in the function causal_forest, and also includes the bootstrap standard errors.[11] We estimate three causal forests, i.e., one for

[10] See Athey and Imbens (2016) and Wager and Athey (2018) for discussions of honesty.

[11] The function optimizes an approximation of the theoretically motivated tree-splitting criterion to increase computational efficiency. See Athey et al. (2019) for details.

each of our three outcomes. We grow each forest using 10,000 trees with at least two observations per leaf. Following the function's default values, we build each tree on a 50 percent subsample of our analysis sample, using the honest approach, and allow 27 of our 37 covariates as tree-splitting candidates at each step.[12]

2.4 Results

Section 2.4.1 describes the OLS average effect estimates for the full sample. Section 2.4.2 subsequently presents the results of our heterogeneity analysis, and Section 2.4.3 gives the estimates for our subgroup analysis based on the insights from the heterogeneity analysis, using our main analysis sample and an additional, independent, sample by Carvalho et al. (2016).

2.4.1 OLS Analysis

Table 2.3 displays the OLS estimates for the average effect of the financial circumstances before payday on the main outcome – i.e., the number of correct answers per second – and the two additional outcomes: number of correct answers and total response time. As can be seen in Column (1), the estimated effect on the number of correct answers per second is statistically insignificant at the 10 percent level, and the point estimate's magnitude of 0.007 appears small relative to the average number of correct answers per second for the after-payday group, which is 0.443. In addition, the sign of the effect point estimate goes in the direction opposite to that which one would expect if the poorer financial circumstances before payday were to impede cognitive function: on average, the individuals assigned to the before-payday group gave a greater number of correct answers per second than did the individuals assigned to the after-payday group. Similar to the results in Column (1), the estimations for the other two outcomes, shown in Columns (2) and (3), also yield effect estimates that are insignificant at the 10 percent level, small in magnitude, and whose signs go in the direction opposite to that which is

[12]Because we use the honest approach in our estimation, effectively a 25 percent subsample is used for growing each tree. For the other parameters that need to be specified in the causal_forest function, we also use the function's default values, and we enable the local centering feature of the algorithm.

expected.

In short, the estimates in Table 2.3 do not suggest that, on average, the poorer financial circumstances before payday have a detrimental effect on cognition in the full sample. This finding is in line with Carvalho et al.'s (2016) results.[13]

Table 2.3. OLS Average Effect Estimates

Outcome	Correct answers per second	Number of correct answers	Total response time (in seconds)
	(1)	(2)	(3)
Before payday	0.007	0.183	−1.062
	(0.006)	(0.425)	(0.916)
Constant	0.443***	42.805***	101.017***
	(0.004)	(0.305)	(0.643)

Notes: $N = 2,480$. The data are from the KnowledgePanel experiment by Carvalho et al. (2016). Heteroscedasticity-robust standard errors are in parentheses. For variable definitions, see Table 2.1.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

[13]Table B.3 in the appendix additionally shows our effect estimates for the subgroups analyzed by Carvalho et al. (2016). Also in line with their results, our estimations yield effect estimates that are insignificant at the 10 percent level and small in magnitude for all subgroups.

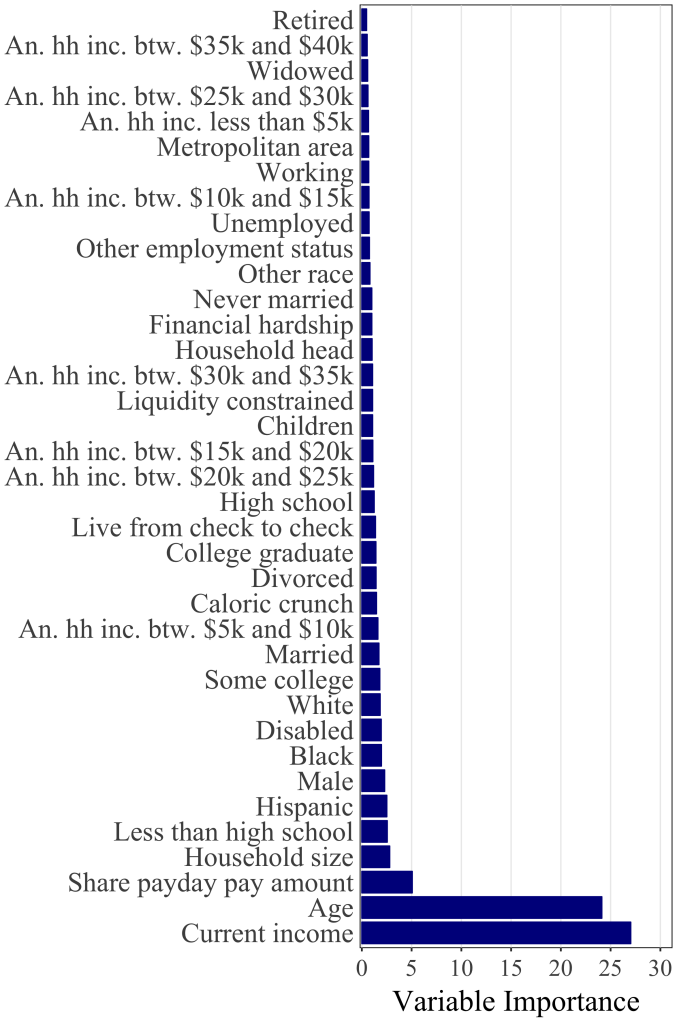
2.4.2 Heterogeneity Analysis

Our heterogeneity analysis proceeds in three steps. First, we calculate a variable importance measure for our three causal forests to identify which of the 37 covariates may be especially important for heterogeneity in our effects of interest. Next, based on these insights, we investigate in heatmaps how the conditional average treatment effects vary over the two most important variables. Subsequently, we estimate effects for two 'typical' individuals in two regions in which the heatmaps suggest particularly detrimental effects, and study how the effect estimates change when we vary the values of the 35 remaining covariates.

To assess variable importance in our estimated causal forests, we use a measure implemented in the grf R package. For variable X_k, the variable importance measure essentially captures the relative frequency with which a forest split on X_k across all grown trees. The measure, therefore, gives an indication over which variables the conditional average treatment effect may vary the most. For X_k, the measure ranges from 0, if the forest never split on X_k, to 100, if the forest always split on X_k.[14] Panel A in Figure 2.1 shows the variable importance plot for the causal forest using the number of correct answers per second as the outcome. The panel yields that by far the two most important variables in the tree-splitting procedure are the covariates age and current income. Both have a variable importance value of approximately 25. All other covariates have a value of around five at most. Similarly, for the two causal forests using the outcomes number of correct answers and total response time, Panels B and C in Figure 2.1 also suggest that age and current income are by far the most important variables.

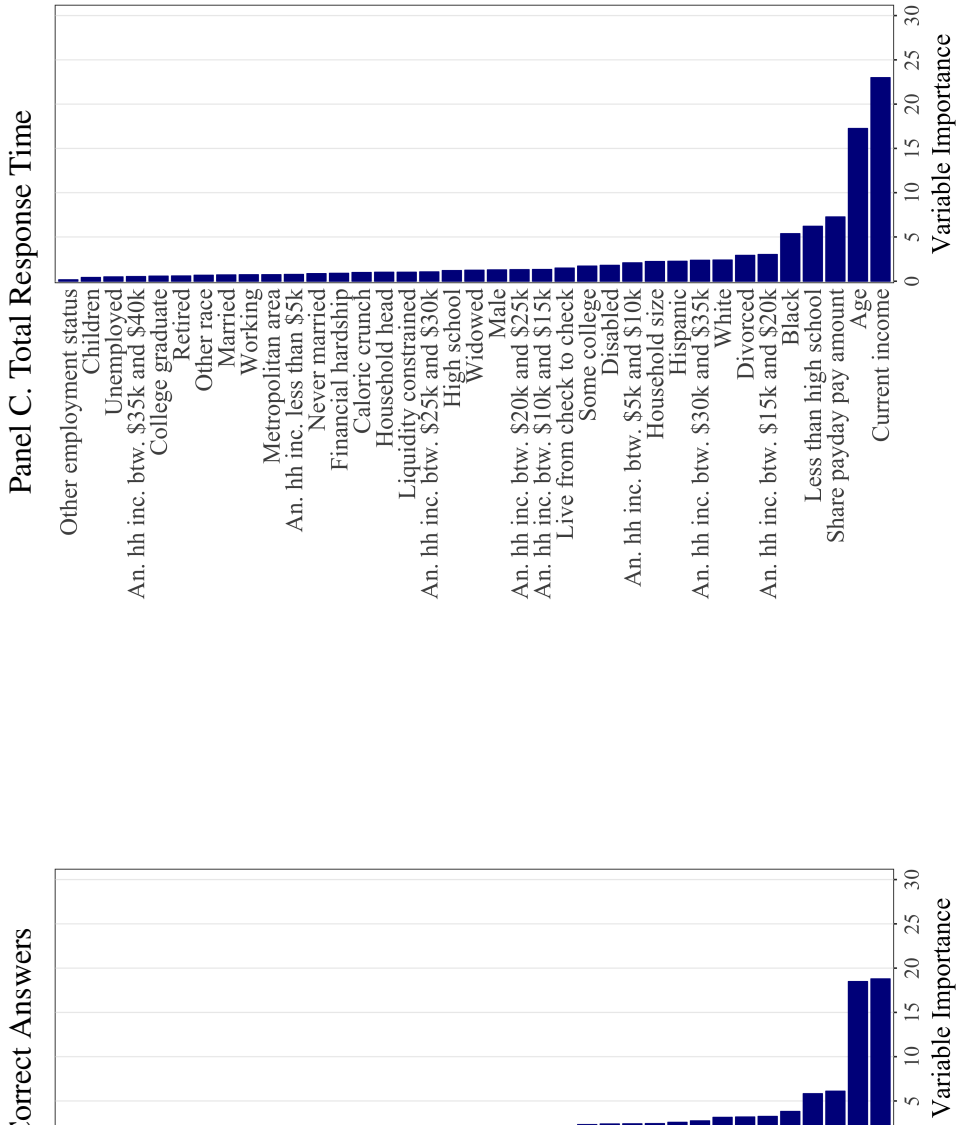
[14] See Appendix B.3 for details.

Figure 2.1. Variable Importance Plots for the Causal Forests
Panel A. Correct Answers per Second



Notes: N=2,480. The figure continues on the next page.

Figure 2.1 Continued



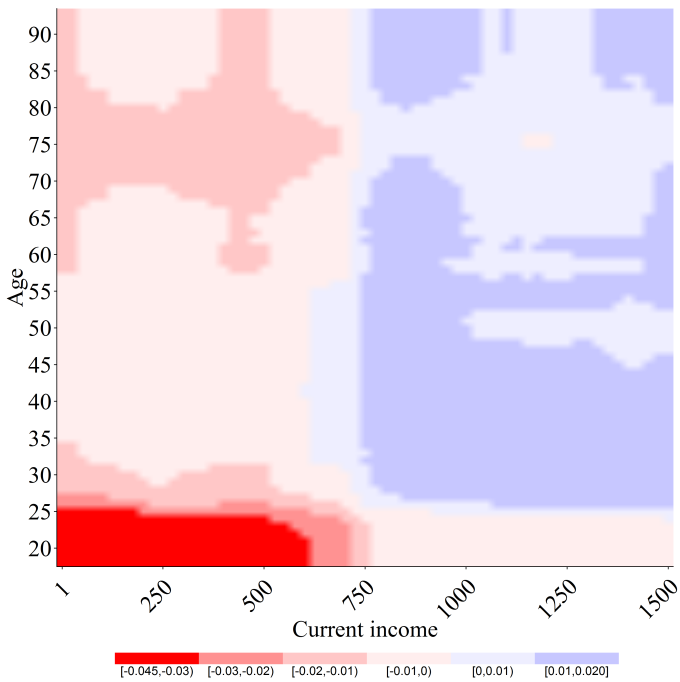
Notes: N=2,480. The The data are from the KnowledgePanel experiment by Carvalho et al. (2016). For the definitions of the outcome variables, see Table 2.1. The variable importance measures are calculated based on causal forest estimations.

Next, to explore how the effects vary in age and current income, Figure 2.2 displays heatmaps, plotting effect estimates over an age–current income grid. The maximum value on the x-axis of \$1500 corresponds to the median current income in our sample. For estimating the effects, we set all other continuous and categorical covariates to their full sample median, and all dummy covariates according to the most frequently occurring characteristics in the full sample. For example, 76.1 percent of all individuals in the sample are white. Therefore, we set the dummy white equal to one, and all other race dummies to zero.[15] Red regions indicate effect estimates that are detrimental and blue regions indicate effect estimates that are not detrimental.

Panel A in Figure 2.2 displays the estimated effects for the number of correct answers per second. The panel shows that the causal forest estimates negative effects especially for individuals who have a current income below approximately \$750 and whose age is either up to approximately 30 years or between around 70 and 80 years. A current income of \$750 appears rather low, corresponding to the 0.16 quantile of our sample's

Figure 2.2. Causal Forest Effect Estimates over an Age–Current Income Grid

Panel A. Correct Answers per Second

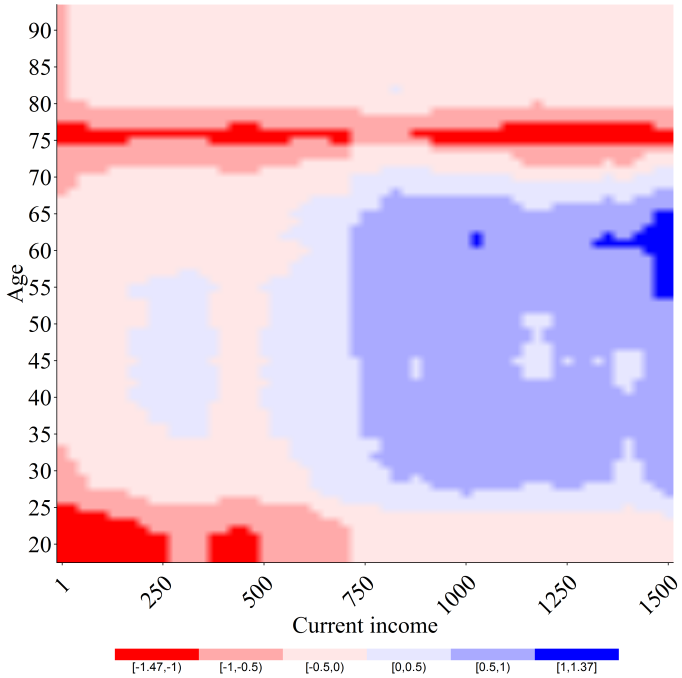


Notes: N=2,480. The figure continues on the next page.

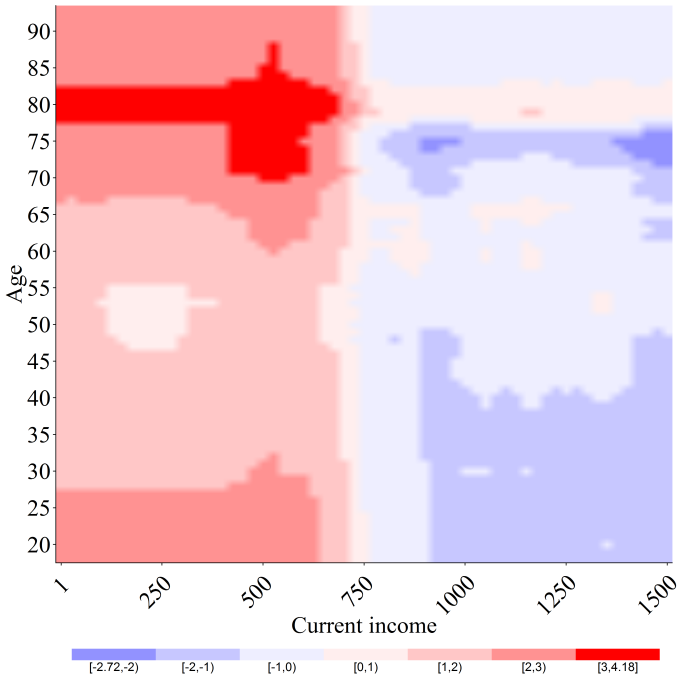
[15] See Appendix B.4 for further details.

Figure 2.2 Continued

Panel B. Number of Correct Answers



Panel C. Total Response Time



Notes: N=2,480. The data are from the KnowledgePanel experiment by Carvalho et al. (2016). For the definitions of the outcome variables, see Table 2.1. The heatmaps show conditional average treatment effect estimates obtained using the causal forest method.

current income distribution. For the younger individuals with a lower current income, the estimated effects are mostly in the range -0.02 to -0.045. The latter value corresponds to approximately 31 percent of the outcome's standard deviation and suggests that the financial circumstances before payday led to 0.045 fewer correct answers per second in the Stroop task. For the older individuals with a lower current income, the effect estimates are between -0.01 and -0.02. Similar to Panel A, Panel B shows that the causal forest using the number of correct answers as the outcome also estimates particularly detrimental effects for individuals with a current income of at most around $750, and who are either younger or older. For the older individuals, the especially detrimental effect estimates are again concentrated in the approximate age range 70 to 80 years. However, they now actually also exceed the $750 threshold. The most detrimental effect estimate in the Panel B heatmap equals -1.47, which corresponds to approximately 14 percent of the standard deviation of the outcome. Panel C displays the estimated effects for the outcome total response time. Similar to the other two panels, the heatmap also yields detrimental effects for individuals whose current income is below $750, and among the lower current income individuals, the causal forest again estimates particularly detrimental effects for younger individuals (up to around 27 years) and older individuals (approximately above age 67). The estimated effects in the most detrimental category are located at the ages 78 to 82 years for current income levels of up to $425, and then at the ages between around 70 and 83 years for current income between approximately $425 and $750. In this category, the causal forest gives effect estimates on the total response time of up to 4.18 seconds, or 18 percent of the outcome's standard deviation. Thus, the heatmaps in Figure 2.2 suggest that the poorer financial circumstances before payday impede, in particular, the cognition of younger and older individuals with a lower current income. The negative effect on the number of correct answers per second appears to result not only from fewer correct answers given but also a slower total response time.

To gain a deeper understanding of the detrimental effects of the financial circumstances before payday, we next zoom in on two regions in which the heatmaps indicate particularly harmful effects. Specifically, we estimate effects for a typical younger individual aged 20 and a typical older individual aged 75, who both have a current income

of $450. We refer to these individuals as typical because we set all other 35 covariates for estimating the effects according to the characteristics in a neighborhood of a given age-current income combination: that is, we construct a five-year age and $250 current-income window centered at the respective age–current income combination and determine the covariate values within this window using the same procedure as for creating the heatmaps above.[16] The first row in the panels of Figures 2.3 and 2.4 gives the estimates for the two typical individuals and all three outcomes. We call these estimates the typical individual baseline estimates. To study how changing the other 35 covariates affects the effect estimates, the panels then show, in the rows below the first row, estimates for which we change one characteristic of a given typical individual at a time, leaving all other variables constant. The empty rows indicate how the covariates are set for a typical individual. For example, for the younger individual in Figure 2.3, the row labeled 'Male = 0' is empty. This indicates that the younger typical individual is female. The row labeled 'Male = 1' then shows the effect estimate when we change the typical individual's gender from female to male. Similarly, the row labeled 'Unemployed = 1' gives the effect estimate when we change the individual's employment status from working to unemployed (every time leaving all other covariates unchanged). In both figures, the horizontal bars indicate 90 percent confidence intervals.

The first row of Panel A in Figures 2.3 and 2.4 shows that the causal forest estimates a negative effect of the financial circumstances before payday on the number of correct answers per second for the younger and older typical individuals. For the younger individual, the estimated effect is -0.0477, and significant at the 1 percent level. The effect size corresponds to approximately one third of the outcome's standard deviation. For the older individual, the effect estimate is -0.0370, or approximately 26 percent of the standard deviation of the outcome. The estimate is significant at the 5 percent level.[17]

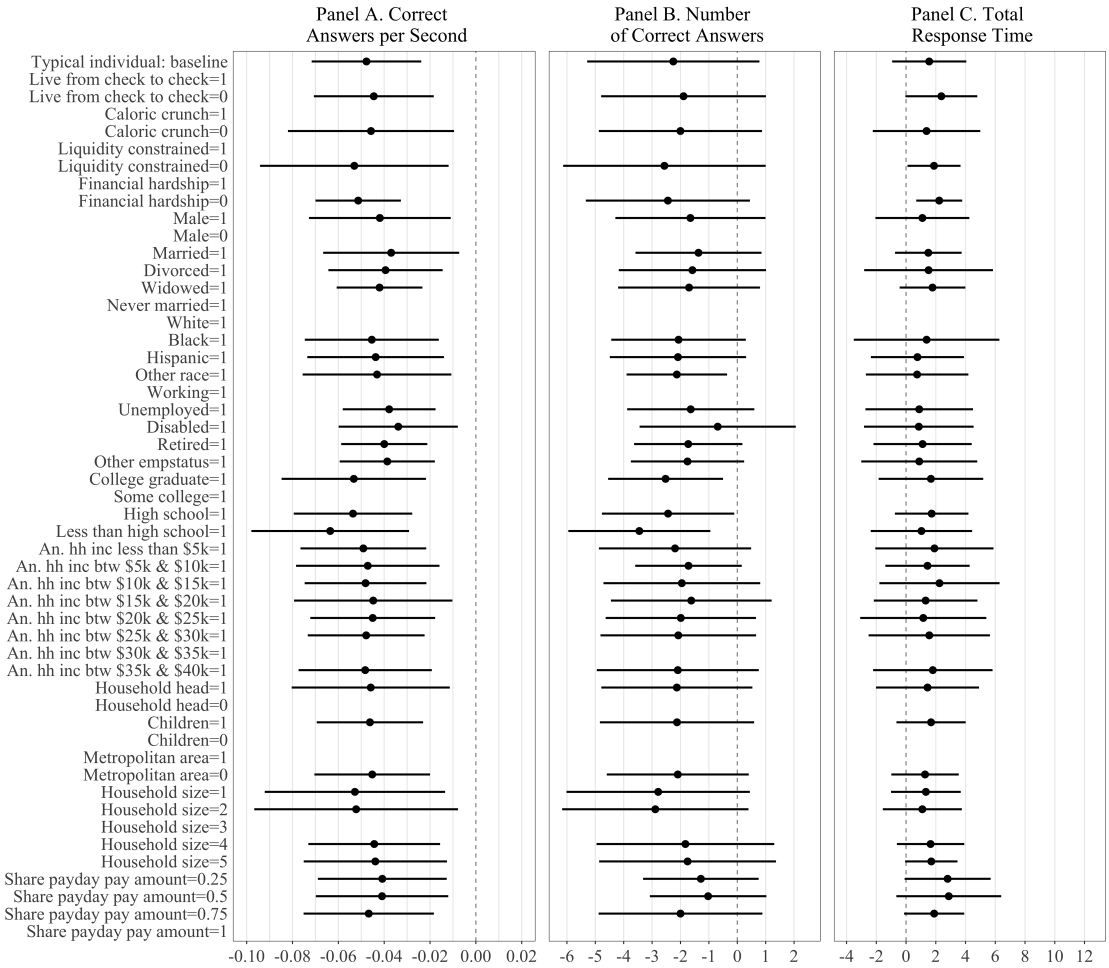
In line with the findings from the heatmaps, row one in Panels B and C in Figures 2.3 and 2.4 suggests that the detrimental effect on the main outcome results from the financial circumstances before payday having a detrimental effect on both its numerator and denominator. The estimate for the effect on the number of correct answers is negative,

[16] See Appendix B.4 for further details.

[17] Tables B.4 and B.5 in the appendix display the estimates discussed in the text.

Figure 2.3. Causal Forest Effect Estimates for the Typical Younger Individual

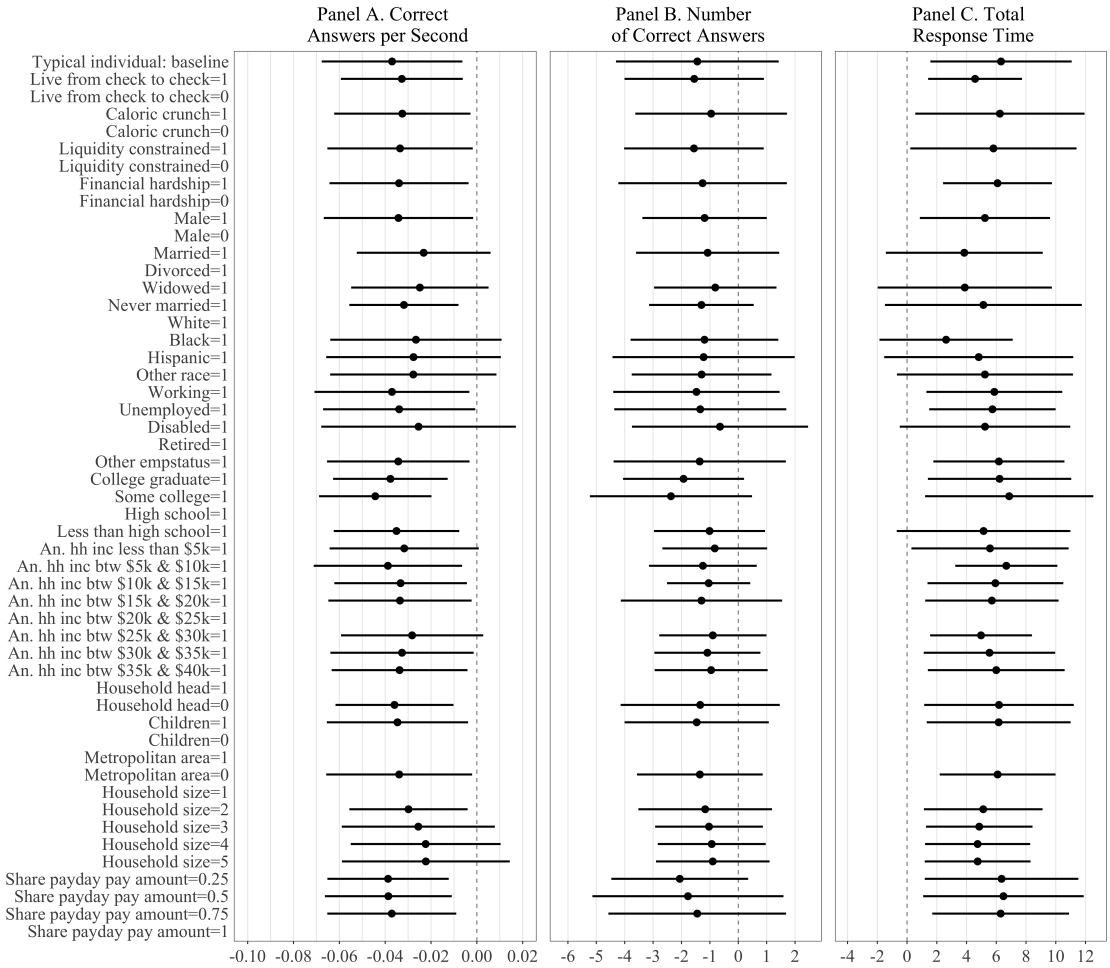
Age = 20, current income = $450



Notes: N=2,480. The data are from the KnowledgePanel experiment by Carvalho et al. (2016). The plots show conditional average treatment effect estimates obtained using the causal forest method. The horizontal bars indicate 90 percent confidence intervals. For the covariates household size and share payday pay amount, the plots give effect estimates at selected points. For the definitions of the outcome variables, see Table 2.1.

Figure 2.4. Causal Forest Effect Estimates for the Typical Older Individual

Age = 75, current income = $450



Notes: N=2,480. The data are from the KnowledgePanel experiment by Carvalho et al. (2016). The plots show conditional average treatment effect estimates obtained using the causal forest method. The horizontal bars indicate 90 percent confidence intervals. For the covariates household size and share payday pay amount, the plots give effect estimates at selected points. For the definitions of the outcome variables, see Table 2.1.

and the estimate for the effect on the total response time is positive. However, only the response-time effect estimate for the older individual is significant at conventional levels.

The rows below the first row in Figures 2.3 and 2.4 show that changing a single characteristic of the two typical individuals does not yield estimates that differ much compared with the baseline estimates. The sign of the effect estimates never changes, and the magnitude of the point estimates remains similar.[18] This behavior is in line with the conclusion from the variable importance plots that age and current income are by far the most important variables for effect heterogeneity.

2.4.3 Subgroup Analysis

Overall, our heterogeneity analysis suggests that poorer financial circumstances before payday are especially detrimental for individuals who have a current income below approximately $750 and whose age is either roughly below 30 or above 70 years. Based on this insight, we next estimate average treatment effects for this subgroup of interest in our sample. Subsequently, to verify the findings in our main analysis sample, we estimate average treatment effects for the subgroup of interest in an independent experiment that Carvalho et al. (2016) conducted in their second online panel. Based on this additional experiment, we only perform a traditional OLS subgroup analysis. We do not use the additional experiment in our heterogeneity analysis using the causal forest method.

To estimate average effects in our main analysis sample, we use the augmented inverse propensity weighted estimator (Robins and Rotnitzky 1995) implemented in the grf R package. The estimator uses the causal forest estimates for all individuals in the subgroup of interest to form the average effect estimates. Table 2.4 presents the estimation results for the subgroup analysis. Column (1) shows the estimate for the effect of the financial circumstances before payday on the number of correct answers per second. The estimation yields an effect estimate of -0.098, which corresponds to approximately 69 percent of the standard deviation of the outcome. The estimate is significant at the 1

[18] Appendix B.5 shows that the conclusions based on other typical individuals in the vicinity of the two typical individuals discussed in the text are the same.

percent level. Columns (2) and (3) display the results from the estimations that use the numerator and denominator of our main outcome as dependent variables. Both regressions also give harmful effect estimates, which are significant at least at the 5 percent level. Thus, in line with the findings from our heterogeneity analysis, the estimations yield detrimental effects on cognition of the poorer financial circumstances before payday for our subgroup of interest.

Table 2.4. Subgroup Average Effect Estimates

Outcome	Correct answers per second	Number of correct answers	Total response time (in seconds)
	(1)	(2)	(3)
Before payday	-0.098^{***}	-3.660^{**}	11.823^{***}
	(0.023)	(1.539)	(2.890)

Notes: N=117. The data are from the KnowledgePanel experiment by Carvalho et al. (2016). Standard errors are in parentheses. The sample includes all individuals who have a current income below $750 and whose age is either below 30 or above 70 years. The estimates are obtained via an augmented inverse propensity weighted estimator which is based on the causal forest estimates for the individuals in the sample. For variable definitions, see Table 2.1.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

To verify the validity of this finding, we next estimate treatment effects for our subgroup of interest in the Flanker task experiment of Carvalho et al. (2016). The authors conducted this experiment in the second online panel that they used, the RAND American Life Panel. As the Stroop task studied in our main analysis, the Flanker task measures cognitive control, and its goal is also to give correct answers to a repeated stimulus as quickly as possible. Carvalho et al. (2016) ran the experiment with 20 trials per participant.

Panel A in Table 2.5 replicates Carvalho et al.'s (2016) OLS estimates for the Flanker task. The regressions do not suggest that the poorer financial circumstances before payday have an effect on cognition in the full sample. The estimated effect on the probability of giving a correct answer in a trial, in Column (1), and on the (log) time that an individual took to respond to a trial, in Column (2), is close to zero and insignificant at the 10 percent level. Panel B displays the analogous estimates for our subgroup of interest. While the estimate in Column (1) does not suggest there to be an effect on the

probability of giving a correct answer, the estimate in Column (2) does indeed suggest a detrimental effect on the log response time per trial. The latter estimate is 0.274 and significant at the 1 percent level. This suggests that the individuals responded on average approximately 27 percent more slowly to the trials of the Flanker task due to the poorer financial circumstances before payday. Thus, in line with the results of our main analysis based on the KnowledgePanel, the analysis based on the American Life Panel also yields detrimental cognitive effects of the poorer financial circumstances before payday for younger and older individuals who have a lower income around the time of the experiment.

Table 2.5. Subgroup Average Effect Estimates in an Independent Experiment

Outcome	Correct answer	Log response time per trial
	(1)	(2)
Panel A. Full sample		
Before payday	0.007	0.016
	(0.010)	(0.028)
Constant	0.863***	8.060***
	(0.012)	(0.030)
N	20,557	20,557
Individuals	1,076	1,076
Panel B. Subgroup: Current income below $750 and age below 30 or above 70 years		
Before payday	0.045	0.274***
	(0.041)	(0.099)
Constant	0.845***	7.908***
	(0.047)	(0.107)
N	1,590	1,590
Individuals	85	85

Notes: The data are from the Flanker task experiment in the RAND American Life Panel by Carvalho et al. (2016). The table reports OLS estimates. Standard errors clustered at the individual level are in parentheses. The regressions include trial-specific dummies. The outcome correct answer is a dummy that equals one if individual i answered a trial correctly. The outcome log response time per trial measures the log time in milliseconds that individual i took to respond to a trial. Panel A replicates the results from Table 6 of Carvalho et al. (2016).

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

2.5 Conclusion

In this paper, we examine heterogeneity in the effect of financial circumstances on cognition. Our analysis is based on data from an experiment by Carvalho et al. (2016), which randomly assigned low-income individuals in the US to perform a cognitive test before or after payday. To explore heterogeneity in the effect of poorer financial circumstances before payday, we use the causal forest method by Athey et al. (2019), which is designed for studying heterogeneous treatment effects in experiments.

The results of our analysis suggest that financial circumstances have heterogeneous effects on cognition. While in our full sample the estimations do not suggest that the poorer financial circumstances before payday affect cognition, we do find detrimental effects for younger and older individuals who received a very low income around the time of the experiment. Specifically, our findings suggest that cognitive test performance was worse among those who received an income of less than $750 at the time of the experiment and whose age was below 30 or above 70 years. We also find detrimental cognitive effects for this group of individuals in an additional, independent, experiment conducted by Carvalho et al. (2016), which we do not use in our heterogeneity analysis. Among the 37 covariates included in our analysis, age and current income appear to be by far the most important for effect heterogeneity. All of the other covariates, such as marital status, household size and education, do not appear to play an important role.

We derive a number of policy recommendations from our findings: First, to address the potential negative cognitive effects of poor financial circumstances, it could be especially beneficial when designing poverty reduction measures to target these at individuals with very few current financial resources and who are either relatively young or old. Second, for this group of individuals, it may prove helpful for public policy to take into account a possible variation in cognitive capacity over payment cycles. For example, to prevent potentially poor decision making due to limited cognition, public administration could try to avoid scheduling appointments with the affected individuals at the end of their payment cycles. Because the payment cycles of welfare programs, such as the food stamp program, are generally regular and set far in advance, this appears to be a feasible option, especially in cases where individuals receive welfare payments.

A fruitful avenue for further research might be to explore why the financial circumstances before payday had detrimental effects for some, but not all, individuals in the experiment. A low current income, for example, may capture particularly poor financial circumstances before payday, and younger and older individuals may be especially worried about these. To gain a deeper understanding of the mechanisms at play, it would be helpful to obtain a larger experimental data set, which focuses on our identified subgroup of affected individuals and would allow for a more detailed analysis.

Appendix B

B.1 Derivation of the Causal Forest Estimator

The causal forest estimator $\hat{\tau}(x)$ for $\tau(x)$ in the random effects model posited in Section 2.3 is based on the two local moment equations

$$E\left(Y_i - \tau(x)D_i - c(x)|X_i = x\right) = 0 \tag{2.4}$$

$$E\left(\left(Y_i - \tau(x)D_i - c(x)\right)D_i|X_i = x\right) = 0, \tag{2.5}$$

where $c(x) = E(\epsilon_i|X_i = x)$ is an intercept term. All other quantities are defined as in the main text. The estimator $\hat{\tau}(x)$ is now obtained by minimizing an empirical version of the two local moment equations:

$$(\hat{\tau}(x), \hat{c}(x)) = argmin_{\tau(x),c(x)} \left\| \sum_{i=1}^{n} \alpha_i(x) \begin{pmatrix} Y_i - \tau(x)D_i - c(x) \\ (Y_i - \tau(x)D_i - c(x))D_i) \end{pmatrix} \right\|_2 . \tag{2.6}$$

The resulting causal forest estimator can be written as

$$\hat{\tau}(x) = \frac{\sum_{i=1}^{n} \alpha_i(x)(Y_i - \bar{Y}_\alpha)(D_i - \bar{D}_\alpha)}{\sum_{i=1}^{n} \alpha_i(x)(D_i - \bar{D}_\alpha)^2}, \tag{2.7}$$

where $\bar{Y}_\alpha = \sum_{i=1}^{n} \alpha_i(x)Y_i$, $\bar{D}_\alpha = \sum_{i=1}^{n} \alpha_i(x)D_i$, and $\alpha_i(x)$ are the similarity weights. It holds that $\sum_{i=1}^{n} \alpha_i(x) = 1$.

Equation (2.7) is the expression for the causal forest estimator in Section 6 of Athey et al. (2019). To obtain the formulation of the estimator in Equation (2.2) in the main text,

rewrite Equation (2.7) as follows. For the numerator, we have

$$
\begin{aligned}
&\sum_{i=1}^{n} \alpha_i(x)(Y_i - \bar{Y}_\alpha)(D_i - \bar{D}_\alpha) \\
&= \sum_{i=1}^{n} \alpha_i(x) Y_i D_i - \left(\sum_{i=1}^{n} \alpha_i(x) D_i\right)\left(\sum_{i=1}^{n} \alpha_i(x) Y_i\right) \\
&= \sum_{\{i:D_i=1\}} \alpha_i(x) Y_i - \left(\sum_{\{i:D_i=1\}} \alpha_i(x)\right)\left(\sum_{\{i:D_i=1\}} \alpha_i(x) Y_i + \sum_{\{i:D_i=0\}} \alpha_i(x) Y_i\right) \\
&= \left(1 - \sum_{\{i:D_i=1\}} \alpha_i(x)\right) \sum_{\{i:D_i=1\}} \alpha_i(x) Y_i - \left(\sum_{\{i:D_i=1\}} \alpha_i(x)\right) \sum_{\{i:D_i=0\}} \alpha_i(x) Y_i \\
&= \left(\sum_{\{i:D_i=0\}} \alpha_i(x)\right) \sum_{\{i:D_i=1\}} \alpha_i(x) Y_i - \left(\sum_{\{i:D_i=1\}} \alpha_i(x)\right) \sum_{\{i:D_i=0\}} \alpha_i(x) Y_i \qquad (2.8)
\end{aligned}
$$

For the denominator, we have

$$
\begin{aligned}
\sum_{i=1}^{n} \alpha_i(x)(D_i - \bar{D}_\alpha)^2 &= \sum_{i=1}^{n} \alpha_i(x) D_i - \left(\sum_{i=1}^{n} \alpha_i(x) D_i\right)^2 \\
&= \left(\sum_{\{i:D_i=1\}} \alpha_i(x)\right)\left(1 - \sum_{\{i:D_i=1\}} \alpha_i(x)\right) \\
&= \left(\sum_{\{i:D_i=1\}} \alpha_i(x)\right)\left(\sum_{\{i:D_i=0\}} \alpha_i(x)\right) \qquad (2.9)
\end{aligned}
$$

The derivations for the numerator and denominator exploit $\sum_{i=1}^{n} \alpha_i(x) = 1$ and $D_i^2 = D_i$. Plugging expression (2.8) for the numerator and expression (2.9) for the denomina-

tor into $\hat{\tau}(x)$ from (2.7) yields

$$\hat{\tau}(x) = \frac{\left(\sum_{\{i:D_i=0\}} \alpha_i(x)\right) \sum_{\{i:D_i=1\}} \alpha_i(x) Y_i - \left(\sum_{\{i:D_i=1\}} \alpha_i(x)\right) \sum_{\{i:D_i=0\}} \alpha_i(x) Y_i}{\left(\sum_{\{i:D_i=1\}} \alpha_i(x)\right) \left(\sum_{\{i:D_i=0\}} \alpha_i(x)\right)}$$

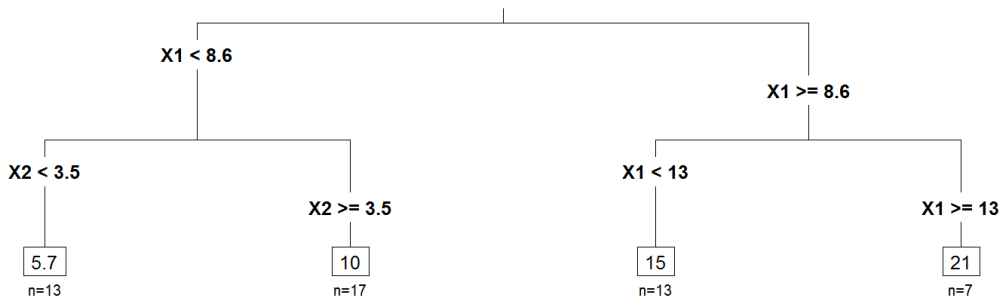
$$= \sum_{\{i:D_i=1\}} \frac{\alpha_i(x)}{\sum_{\{i:D_i=1\}} \alpha_i(x)} Y_i - \sum_{\{i:D_i=0\}} \frac{\alpha_i(x)}{\sum_{\{i:D_i=0\}} \alpha_i(x)} Y_i, \tag{2.10}$$

which is the expression for the causal forest estimator in the main text.

B.2 Tree Example

The figure below this paragraph shows an example of a single small (regression) tree. The tree is built on a sample of size $n = 50$. The data used to construct the tree includes the continuous covariates $X1$, $X2$ and the continuous outcome Y. In the first step, starting from the top of the figure, the tree splits the full sample into two partitions based on the variable $X1$. All observations with an $X1 < 8.6$ are put into the 'left' partition and all observations with an $X1 \geq 8.6$ are put into the 'right' partition. Analogously, the tree subsequently splits the resulting 'left' partition on variable $X2$ and the resulting 'right' partition on variable $X1$ again. The splitting procedure yields four leaves, which are shown at the bottom of the figure. For each leaf, the tree calculates the average outcome Y by averaging over all Y values of all observations that fall into the respective leaf. The averages are then used for predicting Y. For example, for an observation with $X1 < 8.6$ and $X2 \geq 3.5$, the tree predicts an outcome value of 10.

Regression Tree Example



Notes: The values in the boxes correspond to the average outcome Y over all observations that fall into a respective leaf. The number of observations within each leaf is denoted by n.

B.3 Calculation of the Variable Importance Measure

The variable importance measure that we use in our causal forest analysis is implemented in the function variable_importance in the R package grf. We multiply the measure by 100 for readability. The function requires to set the maximum tree depth up to which the measure considers splits, and a decay exponent that controls how the weight

that the splits receive in the overall measure changes as the tree depth increases.[19] We use the default values of the variable_importance function for the two parameters: we set the maximum tree depth to four and the decay exponent to two. For variable X_k, the measure is calculated as follows:

$$vi(X_k) = \left(\sum_{j=1}^{4} w_j \frac{n_{jk}}{n_j} \right) \times 100, \tag{2.11}$$

where n_{jk} is the number of times that all of the trees of the causal forest together split on variable X_k at tree depth $j, j = 1, \ldots, 4$. n_j is the number of times that the trees split at depth j, and $w_j = \frac{j^{-2}}{\sum_{l=1}^{4} l^{-2}}$ is a tree depth-specific weight that determines the importance of splits at a given depth.

In short, the variable importance measure $vi(X_k)$ is a weighted sum of the relative splitting frequencies for X_k over the depths $j = 1, \ldots, 4$, where the weight of the relative splitting frequencies decreases as the tree depth increases.

B.4 Procedure to Set the Covariates

For creating the heatmaps in Figure 2.2, we set the covariates household size and share of payday pay amount relative to current income to their median values in the full sample. All other covariates, which are dummies, we set according to the most frequently occurring characteristics in the full sample. To give two more examples in addition to the example in Section 2.4.2, Table 2.2 shows that the most frequent marital status category is married, with 33.5 percent. Thus, we set the dummy married equal to one and all other marital status dummies we set to zero. Furthermore, Table 2.2 shows that 80.4 percent of individuals live in a metropolitan area. Accordingly, we set the dummy metropolitan area equal to one.

To obtain the estimates for the typical individuals in Tables B.4 and B.5 and the first

[19] For a given tree, the split at depth one corresponds to the first split that a tree places, starting from the entire subsample, and splitting it into two partitions. The splits at depth two then correspond to the splits that the tree performs starting from the two partitions created at depth 1. The next depths follow analogously.

row of Figures 2.3 and 2.4, we proceed analogously to the covariate setting procedure for the heatmaps. However, rather than setting the variables according to the full sample characteristics, we determine the covariate values according to the characteristics in a five-year age and $250 current-income window which is centered at the age–current income combination for which we want to estimate an effect. For example, for the typical older individual in Figure 2.4, the relevant window for setting the covariates ranges from 73 to 77 years of age and from $325.5 to $574.5 of current income. If there are tied categories in categorical variables or dummies that relate to an ordinal characteristic, such as annual household income or education, we select the lowest tied category.[20] For example, if there are equally many individuals in a respective age–current income window with a high school degree and some college, we set the dummy high school to one and all other education dummies to zero. If there are tied categories in non-ordinal characteristics, such as martial status or being liquidity constrained, we set the covariates by extending the age–current income window by one year and $100 in each direction, i.e., we use a seven-year age and $450 current-income window.[21]

B.5 Estimates in the Vicinity of the Two Typical Individuals

In our main analysis, we estimate effects for two typical individuals who have a current income of $450 and whose age is 20 and 75 years, respectively. To assert that the insights based on the two typical individuals are not sensitive to the specific choice of the age–current income combination, we additionally estimate effects for other typical individuals that are in the vicinity of our two typical individuals from the main analysis, where the heatmaps also indicate pronounced detrimental effects. Specifically, we increase and decrease, respectively, age by one and two years and current income by $25 and $50 relative to the typical individuals from the heterogeneity analysis. We estimate

[20] Similarly, if the median household size, as calculated by R, is a non-integer value, we set the household size to the largest integer below the respective median household size. For example, a median household size of 3.5, we set to 3.

[21] For the typical individual with age 20 and current income equal to $400 in Table B.4, extending the age–current income window does not break the tie in the variable household head. In this case, we set household head equal to zero. Setting household head equal to one instead yields the same conclusions.

the effects analogously to the typical individual baseline estimates in Figure 2.3 and Figure 2.4. Tables B.4 and B.5 present the estimates for the other typical younger and older individuals. In the interest of space, we do not display the effect estimates when varying the other 35 covariates. However, very similar to the findings in our main analysis, varying the other covariates one by one does also not change the estimates much relative to the baseline estimates.

The comparison between the estimates for the two typical individuals from our main analysis, which are displayed in the gray shaded areas of Table B.4 and Table B.5, and the other typical individuals shows that overall the vicinity estimates are quite similar to the estimates from the main analysis. For our main outcome, correct answers per second, Panel A in both tables shows that the estimates for the other typical individuals are also always negative and of a similar magnitude as for the respective younger or older typical individual from the main analysis. For the typical younger individuals, all estimates, except for one, are significant at conventional levels. For the typical older individuals, the estimates sometimes lose significance at the 10 percent level.

Panel B in Table B.4 and Table B.5 shows that the estimates for the outcome number of correct answers are also always negative and the point estimates appear quite similar to the respective estimate for the main analysis typical individual, considering the magnitude of the standard errors. As in the main analysis, the estimates are insignificant at the 10 percent level in most regressions. Similar to the findings in Panel B, the estimates for the other typical individuals using the outcome total response time in Panel C are also not substantially different from the respective estimate for the typical individual in the main analysis. In all regressions, the estimations yield positive effect estimates that are insignificant at the 10 percent level for the younger individuals, and mostly significant at conventional levels for the older individuals.

B.6 Additional Tables

Table B.1. Variation in Financial Resources at Payday

Outcome	Cash	Checking and savings	Total expenditures
	(1)	(2)	(3)
Panel A. OLS regressions			
Before payday	−33.39	−6032.75	−542.88
	(73.72)	(5083.40)	(378.69)
Constant	273.18***	15520.66***	1279.50***
	(55.52)	(5000.96)	(371.45)
Panel B. Median regressions			
Before payday	−5.00	−500.00***	−200.00***
	(4.26)	(122.05)	(33.03)
Constant	50.00***	1500.00***	600.00***
	(2.19)	(109.85)	(26.08)
Panel C. p-values for Wilcoxon tests of equality of distributions			
	0.01	0.00	0.00
N	2,295	2,127	2,296

Notes: The data are from the KnowledgePanel experiment by Carvalho et al. (2016). For the OLS regressions, heteroscedasticity-robust standard errors are in parentheses. For the median regressions, bootstrap standard errors based on 1,000 replications are in parentheses. Compared with the analogous results in Carvalho et al.'s (2016) Table 1, only the before-payday estimate in the OLS regression using the outcome total expenditures and the before-payday estimate in the median regression using the outcome cash loses significance in our sample, which is smaller. The two estimates are significant at the 10 percent level in Carvalho et al.'s (2016) analysis.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table B.2. Balance Checks

	Mean		
	After payday	Before payday	p-value
	(1)	(2)	(3)
Age	56.062	55.836	0.747
Male	0.328	0.340	0.515
Household size	1.935	1.953	0.705
Household head	0.843	0.849	0.706
Children in household	0.162	0.173	0.463
Metropolitan area	0.810	0.799	0.499
Current income	1735.856	1740.043	0.937
Share of payday pay amount relative to current income	0.758	0.765	0.534
Financial strain			
Live from paycheck to paycheck	0.480	0.498	0.388
Caloric crunch	0.473	0.467	0.758
Liquidity constrained	0.500	0.506	0.752
Financial hardship	0.404	0.423	0.332
Marital status			
Married	0.323	0.346	0.213
Divorced	0.277	0.275	0.924
Widowed	0.138	0.140	0.867
Never married	0.263	0.239	0.164
Race			
White	0.756	0.766	0.556
Black	0.110	0.090	0.089*
Hispanic	0.084	0.080	0.736
Other race	0.050	0.064	0.130

Notes: The table continues on the next page.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table B.2. Continued

	Mean		
	After payday	Before payday	p-value
	(1)	(2)	(3)
Employment status			
Working	0.284	0.290	0.744
Unemployed	0.067	0.060	0.521
Disabled	0.191	0.207	0.333
Retired	0.391	0.385	0.771
Other employment status	0.067	0.058	0.330
Education			
Less than high school	0.062	0.064	0.855
High school	0.247	0.260	0.465
Some college	0.419	0.415	0.860
College	0.272	0.261	0.534
Annual household income			
Less than $5,000	0.048	0.048	0.990
Between $5,000 and $10,000	0.094	0.105	0.362
Between $10,000 and $15,000	0.134	0.152	0.193
Between $15,000 and $20,000	0.131	0.109	0.081*
Between $20,000 and $25,000	0.147	0.151	0.802
Between $25,000 and $30,000	0.144	0.143	0.941
Between $30,000 and $35,000	0.143	0.138	0.721
Between $35,000 and $40,000	0.158	0.155	0.787

Notes: $N = 2,480$. The data are from the KnowledgePanel experiment by Carvalho et al. (2016). Columns (1) and (2) show the covariate means for the individuals who are randomly assigned to be surveyed after payday or before payday. Column (3) gives the p-values from pairwise t-tests which test whether the difference in means between the before-payday group and after-payday group for a given covariate is different from zero. The difference in means for the covariate black is also significant at the 10 percent level in Carvalho et al.'s (2016) full sample. The p-value of an F-test which tests whether all of the variables jointly predict assignment into the before-payday group is 0.879.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table B.3. OLS Average Effect Estimates for the Subgroups Analyzed by Carvalho et al. (2016)

Outcome	Number of correct answers per second	Number of correct answers	Total response time (in seconds)
	(1)	(2)	(3)
Panel A. Subgroup: One payment			
Before payday	0.003	−0.014	−1.500
	(0.008)	(0.663)	(1.332)
Constant	0.419***	41.799***	104.461***
	(0.006)	(0.478)	(0.986)
N	1,265	1,265	1,265
Panel B. Subgroup: Financial hardship			
Before payday	0.007	0.066	−0.949
	(0.009)	(0.670)	(1.474)
Constant	0.447***	42.638***	99.784***
	(0.007)	(0.490)	(1.050)
N	1,026	1,026	1,026
Panel C. Subgroup: Live paycheck to paycheck			
Before payday	0.012	0.435	−1.737
	(0.008)	(0.602)	(1.284)
Constant	0.441***	42.629***	100.863***
	(0.006)	(0.450)	(0.933)
N	1,213	1,213	1,213
Panel D. Subgroup: Annual household income less than $20,000			
Before payday	0.000	−0.376	−0.321
	(0.009)	(0.756)	(1.522)
Constant	0.424***	41.686***	102.278***
	(0.007)	(0.534)	(1.045)
N	1,020	1,020	1,020
Panel E. Subgroup: Caloric crunch			
Before payday	0.011	0.666	−1.190
	(0.009)	(0.645)	(1.353)
Constant	0.433***	42.040***	101.913***
	(0.006)	(0.482)	(0.982)
N	1,165	1,165	1,165
Panel F. Subgroup: Liquidity constrained			
Before payday	0.013	0.257	−1.753
	(0.008)	(0.619)	(1.388)
Constant	0.437***	42.332***	101.796***
	(0.006)	(0.449)	(0.959)
N	1,248	1,248	1,248

Notes: The data are from the KnowledgePanel experiment by Carvalho et al. (2016). Heteroscedasticity-robust standard errors are in parentheses. For the definitions of the outcome variables and the regressor before payday, see Table 2.1.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table B.4. Causal Forest Estimates for Typical Individuals in the Vicinity of the Typical Younger Individual

Age	18	19	20	21	22
Current income	(1)	(2)	(3)	(4)	(5)
Panel A. Outcome: Correct answers per second					
\$400	−0.0520*	−0.0388**	−0.0385**	−0.0434**	−0.0420**
	(0.0272)	(0.0172)	(0.0187)	(0.0198)	(0.0166)
\$425	−0.0545**	−0.0451**	−0.0499***	−0.0418**	−0.0386*
	(0.0224)	(0.0183)	(0.0167)	(0.0198)	(0.0211)
\$450	−0.0545**	−0.0525**	−0.0477***	−0.0415*	−0.0387*
	(0.0226)	(0.0227)	(0.0145)	(0.0225)	(0.0207)
\$475	−0.0620**	−0.0369**	−0.0405**	−0.0380**	−0.0377**
	(0.0252)	(0.0160)	(0.0184)	(0.0188)	(0.0182)
\$500	−0.0396*	−0.0296**	−0.0424*	−0.0396	−0.0350*
	(0.0236)	(0.0140)	(0.0220)	(0.0284)	(0.0212)
Panel B. Outcome: Number of correct answers					
\$400	−2.281	−1.569	−1.572	−1.566	−1.376
	(1.848)	(1.276)	(1.282)	(1.467)	(1.430)
\$425	−2.692*	−2.014	−2.221*	−1.398	−1.263
	(1.488)	(1.697)	(1.233)	(1.581)	(1.103)
\$450	−2.693*	−2.419	−2.253	−1.256	−1.268
	(1.475)	(1.597)	(1.844)	(1.356)	(1.073)
\$475	−3.384*	−1.379	−1.817	−1.143	−1.260
	(2.009)	(1.566)	(1.503)	(0.999)	(1.182)
\$500	−1.241	−0.739	−2.138*	−1.546	−1.453
	(1.273)	(1.445)	(1.252)	(1.814)	(1.282)
Panel C. Outcome: Total response time					
\$400	1.626	1.445	1.424	2.705	2.587*
	(2.758)	(2.925)	(2.883)	(2.179)	(1.466)
\$425	2.286	1.068	1.984	3.255*	2.224
	(1.962)	(2.249)	(1.730)	(1.834)	(1.874)
\$450	2.293	2.062	1.559	2.342	2.232
	(2.022)	(2.327)	(1.515)	(2.366)	(1.894)
\$475	1.442	1.042	0.901	2.480	2.146
	(2.881)	(2.886)	(1.628)	(2.072)	(1.833)
\$500	1.572	1.118	0.759	0.657	1.197
	(1.991)	(2.483)	(2.209)	(1.810)	(1.608)

Notes: The data are from the KnowledgePanel experiment by Carvalho et al. (2016). Standard errors are in parentheses. For the definitions of the outcome variables, see Table 2.1. The table shows conditional average treatment effect estimates obtained using the causal forest method. For more information, see Appendix B.5.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table B.5. Causal Forest Estimates for Typical Individuals in the Vicinity of the Typical Older Individual

Age	73	74	75	76	77
Current income	(1)	(2)	(3)	(4)	(5)
Panel A. Outcome: Correct answers per second					
\$400	−0.0340	−0.0356*	−0.0361**	−0.0359*	−0.0247
	(0.0226)	(0.0193)	(0.0184)	(0.0199)	(0.0205)
\$425	−0.0359**	−0.0364*	−0.0370**	−0.0367*	−0.0253
	(0.0160)	(0.0199)	(0.0188)	(0.0200)	(0.0186)
\$450	−0.0359**	−0.0365*	−0.0370**	−0.0367*	−0.0253
	(0.0156)	(0.0197)	(0.0187)	(0.0200)	(0.0184)
\$475	−0.0366**	−0.0373**	−0.0378**	−0.0376**	−0.0256
	(0.0142)	(0.0176)	(0.0167)	(0.0183)	(0.0175)
\$500	−0.0369**	−0.0376**	−0.0381**	−0.0379**	−0.0369*
	(0.0156)	(0.0178)	(0.0166)	(0.0182)	(0.0197)
Panel B. Outcome: Number of correct answers					
\$400	−0.850	−1.204	−1.436	−1.529	−0.985
	(1.258)	(1.616)	(1.791)	(1.642)	(1.585)
\$425	−0.840	−1.209	−1.445	−1.543	−1.004
	(1.693)	(1.553)	(1.753)	(1.582)	(1.550)
\$450	−0.831	−1.206	−1.442	−1.540	−0.998
	(1.671)	(1.546)	(1.742)	(1.573)	(1.546)
\$475	−0.786	−1.212	−1.452	−1.551	−0.993
	(1.410)	(1.316)	(1.555)	(1.369)	(1.299)
\$500	−0.778	−1.203	−1.440	−1.537	−1.540
	(1.420)	(1.360)	(1.601)	(1.411)	(1.537)
Panel C. Outcome: Total response time					
\$400	4.971	6.238**	6.195**	6.221**	3.822
	(4.102)	(2.828)	(2.887)	(2.895)	(3.701)
\$425	5.834*	6.375**	6.318**	6.350**	3.932
	(3.266)	(2.804)	(2.874)	(2.954)	(3.893)
\$450	5.856*	6.376**	6.319**	6.351**	3.928
	(3.344)	(2.817)	(2.888)	(2.981)	(3.916)
\$475	6.020*	6.517**	6.461**	6.488**	4.022
	(3.538)	(2.535)	(2.695)	(2.767)	(3.704)
\$500	6.200*	6.701**	6.646**	6.668**	6.372**
	(3.741)	(3.209)	(3.321)	(3.234)	(2.715)

Notes: The data are from the KnowledgePanel experiment by Carvalho et al. (2016). Standard errors are in parentheses. For the definitions of the outcome variables, see Table 2.1. The table shows conditional average treatment effect estimates obtained using the causal forest method. For more information, see Appendix B.5.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Chapter 3

A Natural Experiment on the Role of Response Uncertainty in Household Surveys

3.1 Introduction

For many years, large-scale household surveys have been one of the most important sources of data for empirical research in economics and the social sciences more generally. For example, the questions of how consumption levels and patterns change after retirement, whether such changes are expected or not, and how they are related to financial planning in pre-retirement years are of major importance in current debates about public policy. The empirical analysis of these issues requires reliable data on households' income, consumption expenditure, and many other variables. Typically, such data are taken from household surveys.[1]

Even though large-scale household surveys have been a major success story for more than three decades, concerns about the quality of data collected in social surveys are growing, for instance because of increasing rates of nonresponse and potentially severe

[1] Important examples are the US Health and Retirement Study (HRS), see Juster and Suzman (1995), or its European counterpart, the Survey of Health, Aging and Retirement in Europe (SHARE), see Börsch-Supan et al. (2013b).

measurement error (see, e.g., Meyer et al. 2015). At the same time, the research possibilities afforded by administrative data, 'big data' and other sources of naturally occurring data receive considerable attention (see Einav and Levin 2014; Varian 2014). Data from such sources are increasingly viewed as superior alternatives to survey data, and there is a debate about whether such data will replace those from households surveys. However, as Groves (2011) emphasizes, the key feature shared by these data is that their collection is not designed with research in mind, but they exist for other reasons and are then 'harvested' for research purposes. To us, it therefore seems that despite their merits, naturally occurring data cannot replace survey data. We rather believe that for the foreseeable future, household surveys will continue to be an important source of data for the social sciences. For this reason, we think that further improving response quality in household surveys will be crucial, in particular as competing data sources become available.

In this paper, we focus on a determinant of response quality in household surveys that has received relatively little attention in the literature: the respondents' uncertainty about the quantities they are asked to report. We use data from a natural experiment that allows us to test hypotheses about the relationship between survey respondents' subjective uncertainty, their memory capacity, and data quality. The natural experiment is provided by the fact that in the United States, Social Security checks used to be delivered on the third of each month.[2] As interview dates in large household surveys span over field periods of several weeks (and are as good as random in an econometric sense), we argue that this leads to exogenous variation in the time passed since the last Social Security check delivery, which can be considered a key determinant of response uncertainty in Social Security income.

Our study is related to a small literature in economics that builds on insights from social and cognitive psychology as well as survey research to construct models of survey response error. These models recognize that survey respondents are often uncertain about the quantities a researcher would like them to report. It is easy for respondents to answer

[2]Stephens (2003) uses the same natural experiment to investigate consumption smoothing. He finds that some spending on some expenditure categories rises sharply on the days after the delivery of the Social Security check, which indicates that people are unable to smooth consumption over short horizons. Mastrobuoni and Weinberg (2009) also use the timing of Social Security check delivery in a study of consumption patterns.

questions about their age, marital status, and family relations, but other quantities of interest to economists, including consumption and income, are not easily recalled from memory. It might also be impractical for respondents to look them up, for instance in bank records, during a survey interview. Consequently, respondents use heuristics when they construct their answers on the spot, and these heuristics might bias their responses.[3] Several studies, such as the ones by Battistin et al. (2003), Battistin and Padula (2016), and Angel et al. (2019) as well as those reviewed by Bound et al. (2001), Browning et al. (2003), and Browning et al. (2014), document how recall biases lead to measurement error in variables such as consumption and income constructed from survey data. Hoderlein and Winter (2010) study how recall errors affect the estimation of econometric models in a general, nonparametric framework.

An important conclusion from this literature is that recall errors should best be avoided or at least mitigated, as they are statistically more complex than classical measurement error and difficult to correct in econometric models. Existing insights from neighboring disciplines on the determinants of survey response behavior and of recall errors more specifically have not yet been fully explored in the specific contexts of surveys on economic quantities. One issue that has not received much attention, at least in the literature on responses to household surveys that focus on economic variables, is the *direct* role of subjective uncertainty about the quantities in question – this is the topic of the present paper. The natural experiment described above overcomes a key challenge, namely obtaining reliable measures of survey respondents' subjective uncertainty about the exact values of quantities such as income or expenditure items that are otherwise hard to obtain.

As the natural experiment induces variation in respondent uncertainty, we can test the following hypotheses. Uncertainty about the amount of the Social Security payment should be the higher the longer ago the check was delivered. This effect should be more pronounced for persons with limited memory capacity. In addition, we expect the effect to be even more pronounced for those persons with a limited memory capacity who are unaware of their poor mental ability, because their distorted perception of their mental ability may lead them to putting less effort into recalling the answer to the question. We

[3] Tourangeau et al. (2000) provide an overview of the literature on survey response behavior and question design in social psychology.

operationalize these hypotheses by using the fraction of rounded responses as the main dependent variable. Rounding has been shown to be related to subjective uncertainty in a few studies, including Ruud et al. (2014). The fact that rounding is associated with subjective uncertainty has also been exploited to construct measures of aggregate macroeconomic uncertainty (see Binder 2017; Rossmann 2019).[4]

We use data from the second wave of the Health and Retirement Study, a national survey of persons aged 50 and older that was fielded in 1994. This survey is special in that it contains the exact interview date, which allows us to construct a measure of the time passed since the social security check has been received. Moreover, this survey contains measures of cognitive ability and memory capacity which allow us to test the additional hypotheses stated above. Overall, our hypotheses that postulate a relationship between the time span since the last income receipt, subjective uncertainty and response quality, moderated by memory capacity, are supported by these data.

The remainder of this paper is structured as follows. Section 3.2 describes the HRS data and the natural experiment on response uncertainty. Section 3.3 reports the main results. Section 3.4 presents results from robustness checks. Section 3.5 provides some concluding remarks.

3.2 Natural Experiment and Data

3.2.1 Natural Experiment on Response Uncertainty in the HRS

To test the hypotheses outlined in the introduction, we use data from the Health and Retirement Study (HRS). The HRS is a bi-annual national panel study which surveys Americans over the age of 50 (and their partner). The survey started with an initial sample of about 12,700 individuals in 1992 and collects information about a wide range of topics, such as household finances, cognition, and retirement decisions. For our analysis, we use data from the HRS wave 1994. We use wave 1994 data because, in

[4]There is also a related literature that studies the properties of measurement error induced by rounding and the implications for the estimation of econometric models (see Manski and Molinari 2010; Hoderlein et al. 2015).

addition to last month's Social Security income and memory measures, the wave also contains the exact interview date in the public use file, which is crucial for our analysis. This makes wave 1994 well-suited for the investigation.[5] Before explaining the idea of the natural experiment in more detail, we describe the central survey questions for our analysis and how we define the key variables.

The income question our analysis focuses on relates to the amount of the Social Security check that an individual received last month. The question is asked of all individuals who report receiving income from Social Security at the time of the interview and is worded as follows: 'How much did you receive from Social Security last month?'. To investigate uncertainty in the answer to this question, we look at two types of uncertainty measures. Our first response uncertainty measure is a dummy which equals one if the Social Security amount for an individual is missing in the data.[6] Reasons for missing check amount information include the answer 'don't know', values that the survey termed inappropriate, and refusal of the individual to provide the amount. The motivation for this measure is that if an individual is uncertain about his or her Social Security check amount, one may expect that he or she becomes more likely to provide a response to the check amount question that results in a missing. As our second type of uncertainty measure, we construct dummies that indicate if an individual stated a rounded check amount. Specifically, in our analysis we include three dummies that equal one if an individual's reported Social Security check amount is a multiple of 10, 50, and 100, respectively. The rounding measures relate to the observation that individuals tend to report a rounded number when they are uncertain about an underlying quantity of interest, as discussed in the introduction above.

For constructing our memory capacity measure, we use information from a memory test called delayed word recall task. In wave 1994, the test proceeds as follows. The

[5]HRS wave 1992 and AHEAD wave 1993 also contain the exact interview date. However, HRS wave 1992 does not include information about the Social Security check amount last month and, unlike HRS wave 1994, AHEAD wave 1993 does not include the date of the end of the interview; if a scheduled interview was postponed, the end-of-the-interview date is usually the date on which the interview took place and thus would be needed for our analysis. Additionally, the AHEAD 1993 wave's word recall task, which we use to construct our memory capacity variable, asks respondents to remember a different number of words than the word recall task in HRS wave 1994.

[6]The HRS imputes Social Security check amounts for these individuals. We set their check amounts to missing in our analysis, as this is a relevant outcome.

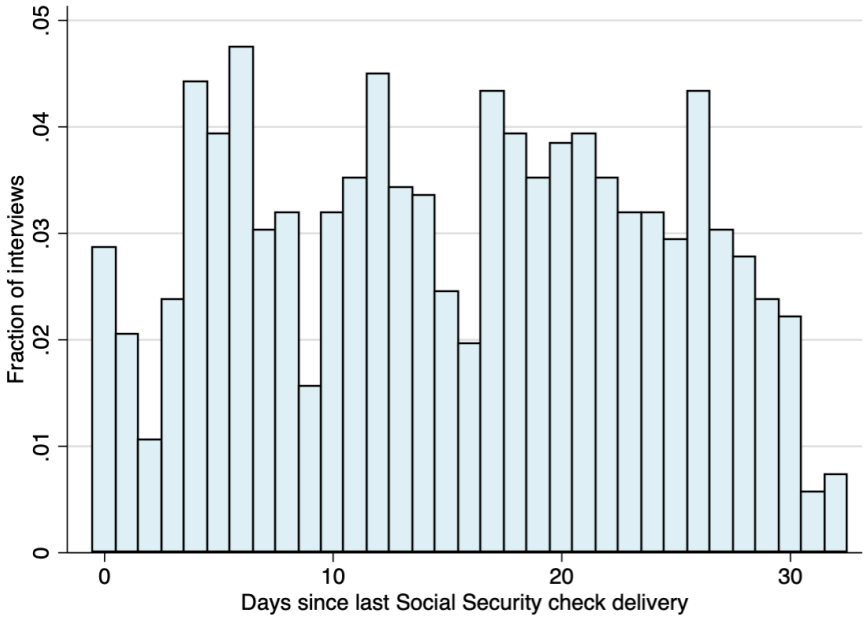
interviewer reads out 20 nouns to the respondent (e.g., mountain, coffee, door). After approximately five minutes during which other survey questions are asked, the respondent then has to repeat as many of the 20 read out words as possible. Based on this task, we consider an individual to have a bad memory if he or she remembers at most three out of the 20 words, and to have a good memory if he or she recalls more than three words.

Our measure indicating if an individual with a bad memory is unaware of his or her bad mental ability we construct based on the question 'First, how would you rate your ability to think quickly at the present time? Would you say it is excellent, very good, good, fair, or poor?'. Individuals answering the question with good, very good, or excellent we define to be unaware of their bad mental ability, and individuals answering the question with fair or poor, we consider to be aware of their bad mental ability.

During our sample period, Social Security checks are delivered on the third day of each month.[7] Combining this knowledge with an individual's interview date, we construct a variable which gives the number of days since the last Social Security check receipt by calculating the difference between an individual's interview date and the last Social Security payday. Based on this variable, we define our regressor of interest below. Figure 3.1 displays the distribution of the days since the last Social Security check delivery variable for our analysis sample. The figure shows that there is some cyclical variation in the number of days since the last check arrival, because Social Security checks are not delivered on weekends, and the longest time spans of 31 and 32 days occur less frequently, because in most months the payment cycle is shorter. Nevertheless, the time spans are overall quite evenly distributed over the entire range of possible values from 0 to 32. This supports the notion that the survey's interview dates are quasi-randomly distributed, leading also to a quasi-random assignment of the time since the last Social Security check receipt. The latter observation gives rise to the following natural experiment exploited in this paper.

[7] If the third of the month is a weekend day or holiday, Social Security checks are delivered on the first day before the respective day that is neither a weekend day nor holiday. Our analysis takes this into account. In our analysis sample, the interviews take place between 7 May 1994 and 21 December 1994. We use the end-of-the-interview date.

Figure 3.1. Distribution of the Days since the Last Social Security Check Delivery



Notes: $N = 1,223$. Authors' calculations based on 1994 HRS data.

3.2.2 Econometric Approach

To test if survey respondents' uncertainty in their reported Social Security check amount increases in the time since their last Social Security check receipt, we compare Social Security income recipients who are interviewed at a late stage of the Social Security payment cycle with Social Security recipients who are interviewed at an early stage of the Social Security payment cycle. The quasi-random assignment of the days since the last Social Security check arrival in our sample implies that the individuals are quasi-randomly interviewed at either the early or late stage of the payment cycle. To investigate how response uncertainty varies by memory capacity and unawareness of bad mental ability, we additionally conduct this comparison for different sub-samples of our analysis sample. Specifically, for individual i, we estimate versions of the following OLS regression equation:

$$y_i = \alpha + \beta late_i + x_i'\gamma + \epsilon_i, \tag{3.1}$$

where y_i is one of our described uncertainty measures, x_i is a vector of covariates described in the next section, $late_i$ is our regressor of interest, which equals one if an individual is interviewed 15 to 32 days after his or her last Social Security check receipt, and zero otherwise, and ϵ_i is a zero-mean error term. Due to the quasi-random assignment of the number of days since the last check arrival, $late_i$ is uncorrelated with the error term.

3.2.3 Sample and Descriptive Statistics

For our analysis sample, we select all individuals in the HRS wave 1994 who report receiving currently Social Security income and who do not have missing values in our analysis variables. This selection procedure yields 1,223 individuals out of the 6,979 individuals reporting about Social Security income receipt.[8]

Table 3.1 presents descriptive statistics for the reported Social Security check amount last month and the covariates we include in our regressions. The average age in our sample is approximately 62 years, 58.8 percent of the individuals are male and almost 80 percent are white/Caucasian. Additionally, the majority of respondents is retired, 76.2 percent, and approximately two thirds of the individuals have at most a high school degree. On average, the individuals with a non-missing Social Security check amount report having received approximately $658 last month.

Table 3.2 displays descriptive statistics and summarizes the definitions for our uncertainty measures, regressor of interest, and the two grouping variables for the sub-sample analyses. The table shows that 6.9 percent of individuals in our sample have a missing check amount. This number appears relatively low and may suggest that the individuals report a check amount even if they are uncertain about the exact figure. Conversely, a relatively high fraction of individuals provide focal responses to the Social Security check amount question: 34.8 percent of respondents who have a non-missing check

[8] Additionally, we drop four individuals with reported Social Security check amounts of $0 and two individuals who report an implausibly large amount of $10,000 and above. In every HRS household, only one individual called financial respondent answers the questions of the survey's income section, which contains our required Social Security income receipt and corresponding check amount question. Thus, all individuals in our sample are financial respondents. We use the RAND HRS data files for our analysis. The data files are carefully pre-processed versions of the raw HRS data, which facilitate the use of the HRS data.

Table 3.1. Descriptive Statistics for the Covariates and Social Security Check Amount Variable

Variable	Mean	Standard deviation	N
Social Security amount (in Dollars)	657.549	262.706	1,139
Male	0.588	0.492	1,223
Age (in years)	62.048	5.066	1,223
White/Caucasian	0.789	0.408	1,223
Married	0.644	0.479	1,223
Labor force status			
Working	0.103	0.304	1,223
Unemployed	0.029	0.167	1,223
Retired	0.762	0.426	1,223
Disabled	0.065	0.246	1,223
Not in labor force	0.042	0.200	1,223
Education			
Less than high school	0.343	0.475	1,223
GED	0.050	0.218	1,223
High school	0.311	0.463	1,223
Some College	0.161	0.368	1,223
College and above	0.135	0.342	1,223

Notes: Authors' calculations based on 1994 HRS data.

amount state a figure which is a multiple of 10, 19.1 percent report multiples of 50, and 14.5 percent report multiples of 100. The shares suggest that indeed a non-negligible number of respondents may be uncertain about their Social Security check amount last month. Considering that the average Social Security check amount is about $658, rounding to multiples of 10 does not appear to indicate substantial uncertainty in the reported figure and does likely not create substantial measurement error. Conversely, rounding to multiples of 50 or even 100 seems to indicate significant uncertainty in the reported amount and may lead to substantial measurement error for analyses using last month's Social Security income as a variable.

In addition, Table 3.2 gives that 52.8 percent of individuals are interviewed late in the Social Security payment cycle. This share is intuitively plausible given the relatively uniform distribution of the number of days since the last Social Security check arrival in our sample and considering that the late variable splits the payment cycle approximately into two equal parts. For our grouping variables, the table indicates that about a third of

the individuals have a bad memory, and out of the 343 bad memory individuals who do not have a missing Social Security check amount, 63.3 percent are unaware of their bad mental ability.

Table 3.2. Variable Definitions and Descriptive Statistics for the Uncertainty Measures, Regressor of Interest, and Grouping Variables

Variable	Definition	Mean	Standard deviation	N
Uncertainty Measures				
Missing check amount	= 1 if the Social Security amount is missing	0.069	0.253	1,223
10-focal response	= 1 if the Social Security amount is a multiple of 10	0.348	0.476	1,139
50-focal response	= 1 if the Social Security amount is a multiple of 50	0.191	0.394	1,139
100-focal response	= 1 if the Social Security amount is a multiple of 100	0.145	0.352	1,139
Regressor of interest				
Late	= 1 if the interview takes place 15–32 days after the last Social Security check delivery	0.528	0.499	1,223
Grouping variables				
Memory bad	= 1 if the individual remembered $\leq$ 3 out of the 20 words in the delayed word recall task	0.303	0.460	1,223
Unaware of bad mental ability	Among all individuals with a non-missing Social Security amount and bad memory: = 1 if the individual rates his or her ability to think quickly at the present time as good, very good, or excellent (as opposed to fair or poor).	0.633	0.483	343

Notes: Authors' calculations based on 1994 HRS data.

3.3 Main Results

In this section, we present our main results, based on a series of regressions with the missing check amount indicator and the three focal response indicators as the dependent variables. We complement the regression tables with figures that highlight the key contrasts graphically.

Table 3.3. Change in the Fraction of Missing Check Amounts between the Early and Late Stage of the Social Security Payment Cycle

Sample	All		Memory good		Memory bad	
	(1)	(2)	(3)	(4)	(5)	(6)
Late	0.018	0.018	0.013	0.014	0.030	0.028
	(0.014)	(0.014)	(0.017)	(0.017)	(0.027)	(0.027)
Constant	0.059***	−0.624**	0.060***	−0.554	0.057***	−0.684
	(0.010)	(0.249)	(0.012)	(0.370)	(0.018)	(0.538)
Controls	–	✓	–	✓	–	✓
N	1,223		853		370	

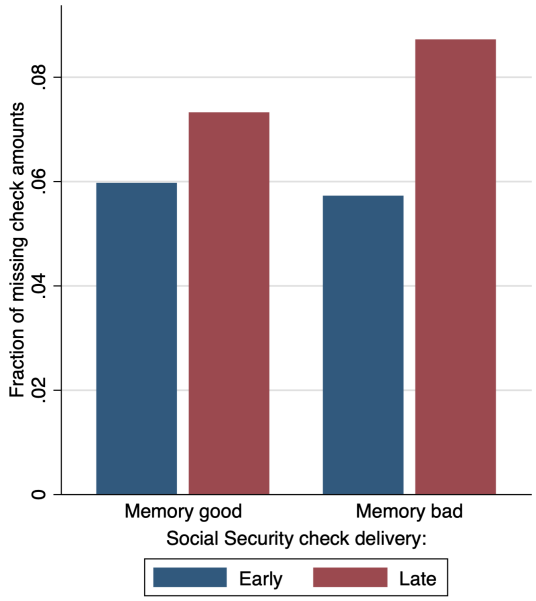
Notes: Heteroscedasticity-robust standard errors are in parentheses. The estimations are based on 1994 HRS data. All models are estimated via OLS. The outcome variable is the missing check amount measure defined in Table 3.2. The set of control variables consists of the dummy variables male, white, married, working, unemployed, retired, disabled, GED, high school, some college, college and above, and the continuous variables age, age^2. The labor force status not in labor force and the education category less than high school are omitted because of multicolinearity. All individuals for whom the dummy variable memory bad equals one belong to the 'memory bad sample'. All other individuals belong to the 'memory good sample'. For the definition of the regressor late, and memory bad variable, see Table 3.2.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table 3.3 displays the estimation results for the missing check amount measure. The corresponding Figure 3.2 shows the coefficient estimates for the models without controls, stratified by memory capacity, as shown in Columns (3) and (5) of Table 3.3. Columns (1) and (2) show for the full sample that the fraction of missing check amounts is 1.8 percentage points higher for the individuals interviewed late in the Social Security payment cycle compared with the individuals interviewed early in the cycle. The estimate is insignificant at the 10 percent level, with and without controls, however. Looking at the estimates by memory capacity, Columns (3)–(6) yield that the increase in the fraction of missing values between the early and late payment cycle stage is larger for the individuals with a bad memory (approximately 2.8 percentage points) than for the individuals

with a good memory (about 1.4 percentage points). However, these estimates are also insignificant at the 10 percent level. Thus, the missing check amount measure does not appear to capture an increase in response uncertainty between the interviews conducted early and late in the payment cycle.

Figure 3.2. Fraction of Missing Check Amounts



Notes: The estimates are based on HRS 1994 data. Memory good: $N = 853$. Memory bad: $N = 370$. The figure illustrates the estimates by memory capacity, without controls, from Table 3.3.

Conversely, Table 3.4 indicates that uncertainty does increase based on the focal response measures. For the full sample, the estimations in Columns (1) and (2) yield that the fraction of check amounts rounded to multiples of 10, 50, and 100, respectively, increases by about 18 to 26 percent in the late payment cycle stage relative to the corresponding rounding shares in the early payment cycle stage (which are given by the estimates for the constants in the first column). After adding controls in Column (2), all of the estimates are significant at conventional levels.

Columns (3)–(6) present the estimation results by memory capacity. In Figures 3.3–3.5, Panel A illustrates the estimates without controls, as shown in Columns (3) and (5) of Table 3.4. For the memory good group, the estimates are smaller compared with the full sample results, and insignificant at the 10 percent level. While the point estimates

for the 10- and 50-focal response measures may still suggest a small increase in the fraction of rounded responses for the memory good individuals interviewed late in the payment cycle, the point estimates in Columns (3) and (4) of Panel C do not suggest that rounding to multiples of 100 increases late in the payment cycle for respondents with a good memory. Conversely, for the respondents who have a bad memory, the estimates in Columns (5) and (6) give that the increase in the fraction of rounded Social Security check amounts between the early and late payment cycle stage is even larger than for the full sample. In all regressions but one, these increases are significant at conventional levels. For the 100-focal response measure in Panel C, the estimations indicate a particularly large increase of 10.6 percentage points after adding controls. This corresponds to an approximately 110 percent increase relative to the bad memory group's fraction of 100-focal responses early in the payment cycle, which is 9.7 percent.

Columns (7)–(10) report the estimates from regressions where the individuals with a bad memory are grouped by their unawareness of their bad mental ability. The estimates without controls are illustrated in Figures 3.3–3.5, Panel B; they correspond to Columns (7) and (9) of Table 3.4. Similar to the previous grouping based on memory capacity, splitting by unawareness of bad mental ability again leads to increased estimates for one group and decreased estimates for the other group: for individuals who are unaware of their bad mental ability, the change in the fraction of rounded responses between the early and late payment cycle stage goes up across all three rounding measures, and is significant at conventional levels in all regressions, whereas for the individuals who are aware of their bad mental ability, the estimates become smaller and insignificant at the 10 percent level.

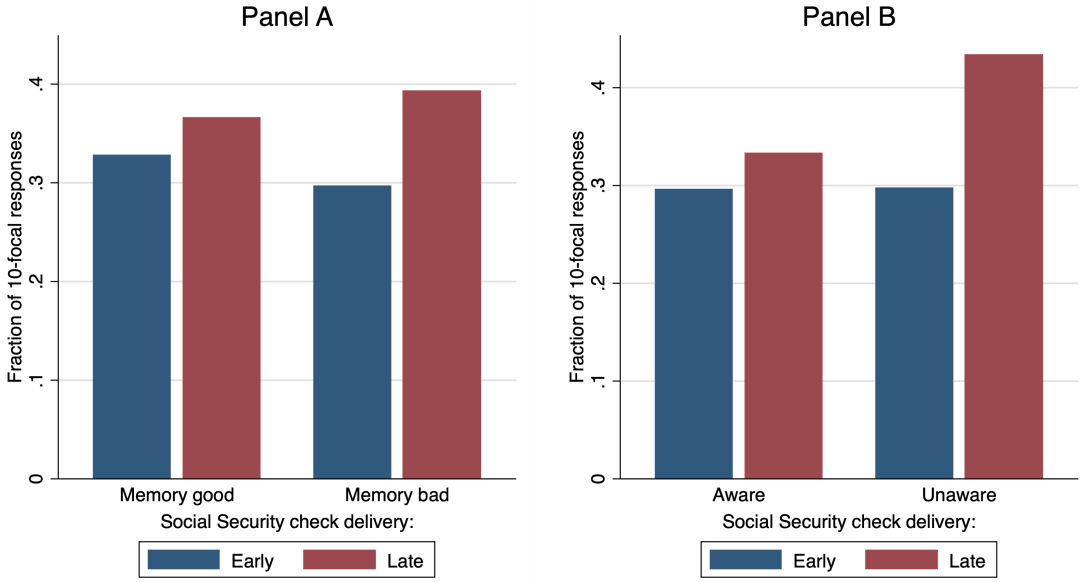
In sum, our analysis findings based on the focal response measures suggest that the uncertainty in the reported Social Security check amount increases in the time since the last Social Security check receipt. The increase appears to be especially driven by respondents who have a bad memory, and among these individuals, respondents who are unaware of their bad mental ability seem to exhibit a particularly large increase in their response uncertainty over the payment cycle. These findings are in line with our hypotheses.

Table 3.4. Change in the Fraction of Focal Responses between the Early and Late Stage of the Social Security Payment Cycle

Sample	All		Memory good		Memory bad		Memory bad: Unaware of bad mental ability		Memory bad: Aware of bad mental ability	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Outcome: 10-focal response										
Late	0.056**	0.062**	0.038	0.049	0.096*	0.106**	0.137**	0.147**	0.037	0.019
	(0.028)	(0.028)	(0.034)	(0.034)	(0.051)	(0.050)	(0.065)	(0.064)	(0.084)	(0.089)
Constant	0.319***	1.099	0.328***	0.445	0.297***	1.314	0.297***	1.222	0.296***	2.523
	(0.020)	(1.074)	(0.024)	(1.341)	(0.036)	(1.514)	(0.044)	(1.663)	(0.063)	(5.047)
Panel B. Outcome: 50-focal response										
Late	0.031	0.040*	0.019	0.028	0.061	0.072*	0.093*	0.097*	0.019	0.044
	(0.023)	(0.023)	(0.028)	(0.028)	(0.043)	(0.042)	(0.056)	(0.055)	(0.066)	(0.069)
Constant	0.175***	0.151	0.180***	0.421	0.164***	−0.917	0.171***	−0.786	0.148***	−0.925
	(0.016)	(0.614)	(0.020)	(0.977)	(0.029)	(0.916)	(0.036)	(1.011)	(0.049)	(3.020)
Panel C. Outcome: 100-focal response										
Late	0.027	0.034*	−0.002	0.005	0.094**	0.106***	0.119**	0.125**	0.069	0.092
	(0.021)	(0.021)	(0.025)	(0.025)	(0.038)	(0.037)	(0.052)	(0.051)	(0.050)	(0.056)
Constant	0.131***	0.566	0.146***	0.899	0.097***	−0.365	0.117***	−0.570	0.056*	−1.615
	(0.014)	(0.606)	(0.018)	(0.960)	(0.023)	(0.757)	(0.031)	(0.913)	(0.031)	(2.076)
N	1,139		796		343		217		126	
Controls	–	✓	–	✓	–	✓	–	✓	–	✓

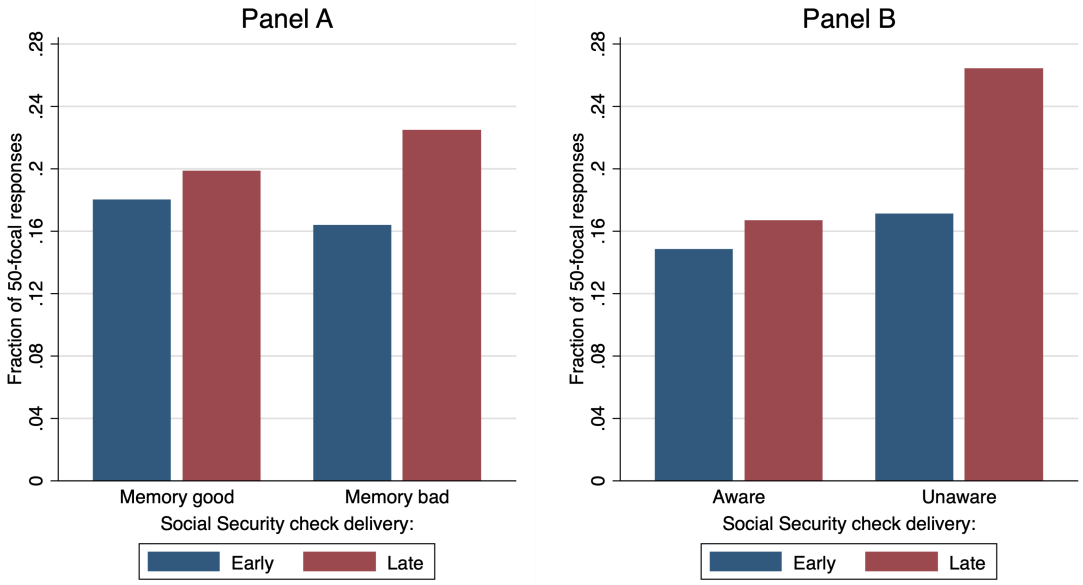
Notes: Heteroscedasticity-robust standard errors are in parentheses. The estimations are based on 1994 HRS data. All models are estimated via OLS. The set of control variables consists of the dummy variables male, white, married, working, unemployed, retired, disabled, GED, high school, some college, college and above, and the continuous variables age, age^2. The labor force status not in labor force and the education category less than high school are omitted because of multicolinearity. All individuals for whom the dummy variable memory bad equals one belong to the 'memory bad sample'. All other individuals belong to the 'memory good sample'. All individuals with a bad memory for whom the dummy variable unaware of bad mental ability equals one belong to the 'unaware of bad mental ability sample'. All other individuals with a bad memory belong to the 'aware of bad mental ability sample'. For the definitions of the outcome variables, regressor late, and grouping variables see Table 3.2. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Figure 3.3. Fraction of 10-Focal Responses



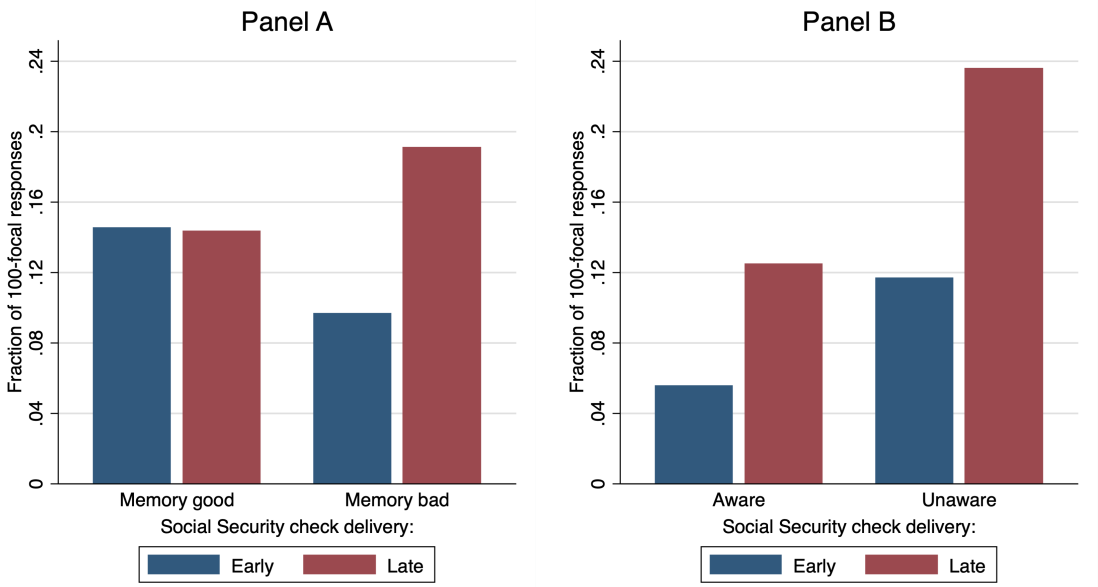
Notes: Memory good: $N = 796$. Memory bad: $N = 343$. Aware: $N = 126$. Unaware: $N = 217$. The estimates are based on HRS 1994 data. The figure illustrates the estimates by memory capacity and unawareness of bad mental ability, without controls, from Panel A in Table 3.4.

Figure 3.4. Fraction of 50-Focal Responses



Notes: Memory good: $N = 796$. Memory bad: $N = 343$. Aware: $N = 126$. Unaware: $N = 217$. The estimates are based on HRS 1994 data. The figure illustrates the estimates by memory capacity and unawareness of bad mental ability, without controls, from Panel B in Table 3.4.

Figure 3.5. Fraction of 100-Focal Responses



Notes: Memory good: $N = 796$. Memory bad: $N = 343$. Aware: $N = 126$. Unaware: $N = 217$. The estimates are based on HRS 1994 data. The figure illustrates the estimates by memory capacity and unawareness of bad mental ability, without controls, from Panel C in Table 3.4.

3.4 Robustness Checks

The previous section showed that the uncertainty in the Social Security check amount increases for individuals who are interviewed late rather than early in the monthly Social Security payment cycle, based on our focal response measures. To assert that this finding is not sensitive to the definition of the late Social Security payment cycle stage, we re-estimate our rounding analysis, using two alternative late dummy definitions. Tables 3.5 and 3.6 present results where we decrease the late payment cycle stage window by three and five days, respectively.[9]

Overall, the alternative estimations yield quite similar results as our main analysis. In

[9]Thus, the late dummy equals now one if an individual is interviewed 18–32 days and 20–32 days, respectively, after the last Social Security check receipt.

the full sample, the fraction of rounded check amounts increases between the early and late interviews for all three focal response measures, even though some estimates lose significance at the 10 percent level compared with the respective main estimates. The increases for the memory bad group are again larger than for the full sample, and in most regressions even slightly larger than the corresponding main analysis results. A reason for the slightly larger increases could be that decreasing the late payment cycle stage window may especially increase the check amount uncertainty in the late group.[10] For the individuals with a good memory, the point estimates in Panel A of both tables may also suggest a small increase in 10-focal responses between the early and late stage of the payment cycle. The point estimates in Panels B and C, however, do not support the notion that there is an increase in 50- or 100-focal responses. Additionally, all of the memory good group estimates are insignificant at the 10 percent level, just as in the main analysis. Dividing the individuals with a bad memory by unawareness of their bad mental ability, the regressions for the unaware group again yield estimates that are larger than for the whole memory bad sample. All of these estimates, except for one, are also again significant at conventional levels. For the respondents who are aware of their bad mental ability, the point estimates generally decrease compared with the estimates for all bad memory individuals, and all of these estimates, apart from two, are insignificant at the 10 percent level. The findings for the respondents who are aware of their bad mental ability are thus also similar to the corresponding main estimates.

[10]Decreasing the late payment cycle stage window implies that the late group individuals are on average further away from their last Social Security check arrival, which likely increases the group's overall response uncertainty. However, decreasing the window also shifts individuals, who are likely more uncertain, from the late to the early group. For this reason, it is unclear how changing the late dummy definition affects the estimates in general, and depends on how exactly the uncertainty increases over the Social Security payment cycle.

Table 3.5. Alternative Late Dummy Definition 1: Change in the Fraction of Focal Responses between the Early and Late Stage of the Social Security Payment Cycle

Sample	All		Memory good		Memory bad		Memory bad: Unaware of bad mental ability		Memory bad: Aware of bad mental ability	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Outcome: 10-focal response										
Late(18–32)	0.055*	0.061**	0.023	0.036	0.130**	0.129**	0.183***	0.167**	0.065	0.060
	(0.029)	(0.028)	(0.034)	(0.034)	(0.052)	(0.051)	(0.068)	(0.067)	(0.083)	(0.088)
Constant	0.323***	1.141	0.338***	0.507	0.290***	1.239	0.293***	1.172	0.283***	2.222
	(0.019)	(1.064)	(0.022)	(1.332)	(0.033)	(1.450)	(0.040)	(1.578)	(0.059)	(4.972)
Panel B. Outcome: 50-focal response										
Late(18–32)	0.016	0.022	−0.011	−0.001	0.079*	0.075*	0.113*	0.089	0.048	0.097
	(0.024)	(0.023)	(0.028)	(0.028)	(0.044)	(0.042)	(0.059)	(0.058)	(0.065)	(0.069)
Constant	0.184***	0.204	0.195***	0.501	0.161***	−0.943	0.173***	−0.796	0.133***	−1.292
	(0.015)	(0.611)	(0.019)	(0.960)	(0.027)	(0.901)	(0.033)	(0.995)	(0.044)	(3.005)
Panel C. Outcome: 100-focal response										
Late(18–32)	0.024	0.029	−0.012	−0.005	0.108***	0.106***	0.142**	0.123**	0.086*	0.108*
	(0.021)	(0.021)	(0.025)	(0.025)	(0.040)	(0.038)	(0.056)	(0.055)	(0.051)	(0.058)
Constant	0.134***	0.596	0.150***	0.922	0.098***	−0.395	0.120***	−0.593	0.050*	−1.648
	(0.013)	(0.608)	(0.017)	(0.952)	(0.022)	(0.755)	(0.028)	(0.898)	(0.028)	(2.071)
N	1,139		796		343		217		126	
Controls	–	✓	–	✓	–	✓	–	✓	–	✓

Notes: Heteroscedasticity-robust standard errors are in parentheses. The estimations are based on 1994 HRS data. All models are estimated via OLS. The dummy late(18–32) equals one if an individual is interviewed 18 to 32 days after the last Social Security check arrival and zero otherwise. The controls are the same as the ones given in the notes for Table 3.4. All individuals for whom the dummy variable memory bad equals one belong to the 'memory bad sample'. All other individuals belong to the 'memory good sample'. All individuals with a bad memory for whom the dummy variable unaware of bad mental ability equals one belong to the 'unaware of bad mental ability sample'. All other individuals with a bad memory belong to the 'aware of bad mental ability sample'. For the definitions of the outcome variables and grouping variables, see Table 3.2. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 3.6. Alternative Late Dummy Definition 2: Change in the Fraction of Focal Responses between the Early and Late Stage of the Social Security Payment Cycle

Sample	All		Memory good		Memory bad		Memory bad: Unaware of bad mental ability		Memory bad: Aware of bad mental ability	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Outcome: 10-focal response										
Late(20–32)	0.047	0.055*	0.016	0.032	0.116**	0.111**	0.195***	0.164**	0.006	0.024
	(0.029)	(0.029)	(0.035)	(0.035)	(0.054)	(0.053)	(0.070)	(0.071)	(0.085)	(0.090)
Constant	0.331***	1.151	0.342***	0.530	0.304***	1.225	0.299***	1.089	0.315***	2.501
	(0.018)	(1.064)	(0.021)	(1.333)	(0.031)	(1.478)	(0.038)	(1.615)	(0.055)	(4.956)
Panel B. Outcome: 50-focal response										
Late(20–32)	0.029	0.037	0.002	0.015	0.093**	0.088*	0.148**	0.115*	0.019	0.087
	(0.025)	(0.024)	(0.029)	(0.029)	(0.046)	(0.045)	(0.063)	(0.062)	(0.067)	(0.074)
Constant	0.181***	0.181	0.189***	0.474	0.161***	−1.008	0.167***	−0.885	0.151***	−1.217
	(0.014)	(0.612)	(0.017)	(0.965)	(0.025)	(0.907)	(0.031)	(0.989)	(0.042)	(3.047)
Panel C. Outcome: 100-focal response										
Late(20–32)	0.031	0.038*	−0.002	0.007	0.108**	0.106***	0.170***	0.144**	0.031	0.070
	(0.022)	(0.022)	(0.026)	(0.026)	(0.042)	(0.041)	(0.060)	(0.059)	(0.055)	(0.064)
Constant	0.133***	0.583	0.145***	0.900	0.106***	−0.444	0.118***	−0.691	0.082**	−1.356
	(0.013)	(0.608)	(0.016)	(0.955)	(0.021)	(0.764)	(0.027)	(0.899)	(0.032)	(2.201)
N	1,139		796		343		217		126	
Controls	–	✓	–	✓	–	✓	–	✓	–	✓

Notes: Heteroscedasticity-robust standard errors are in parentheses. The estimations are based on 1994 HRS data. All models are estimated via OLS. The dummy late(20–32) equals one if an individual is interviewed 20 to 32 days after the last Social Security check arrival and zero otherwise. The controls are the same as the ones given in the notes for Table 3.4. All individuals for whom the dummy variable memory bad equals one belong to the 'memory bad sample'. All other individuals belong to the 'memory good sample'. All individuals with a bad memory for whom the dummy variable unaware of bad mental ability equals one belong to the 'unaware of bad mental ability sample'. All other individuals with a bad memory belong to the 'aware of bad mental ability sample'. For the definitions of the outcome variables and grouping variables, see Table 3.2. *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

3.5 Conclusions

Using data from a natural experiment, this paper has demonstrated that subjective uncertainty about a quantity – specifically, Social Security income – influences responses to an open-ended survey question. In the natural experiment, variations in uncertainty were induced by the fact that Social Security checks are always delivered on the same day of the month while survey interview dates are distributed over the whole month. Hence, time elapsed between check delivery and interview varies randomly across respondents. The analysis in this paper has also shown that these response effects interact with the respondents' memory capacity in non-trivial ways.

The size of the effects we found is small, which raises the question of whether the findings of this study are relevant for practical work. However, reporting one's Social Security income is, arguably, a relatively simple task so one would probably expect to find no such effects at all. Other tasks such as reporting consumption expenditure are more difficult and known to suffer from recall error. More research on the role of respondents' uncertainty about such quantities is important but also difficult because uncertainty is unobserved – this is why the natural experiment we studied here is so valuable. An important implication is that survey agencies should make survey metadata, such as interview dates, routinely available with the survey data itself. Survey metadata can provide information that can be exploited in the estimation of econometric models of survey response error, as in the framework laid out by Hoderlein and Winter (2010).

Chapter 4

Testing under a Special Form of Heteroscedasticity[1]

4.1 Introduction

In the June 2011 issue of the *American Economic Review*, Vikesh Amin commented on an article by Dorothe Bonjour et al. published in December 2003 also in the *American Economic Review*. Bonjour et al. (2003) estimated the private return to education using a data set containing 428 female monozygotic twins. One of their main findings was an estimated return to one additional year of education of 7.7 percent, which is statistically significant at the 5 percent level. Amin (2011) replicated their results and performed similar estimations where he excluded outliers. He found that many of Bonjour et al.'s within-twin pair estimates became smaller in magnitude and significant only at lower levels or insignificant when removing these extreme values.

In this study, we show that the inference in Amin (2011) is mostly incorrect due to the presence of a special form of heteroscedasticity. The correct standard errors turn out to be around 15 percent lower, leading to different policy conclusions. In contrast to Amin (2011), we find a significant positive return to education for most of the within-twin pair regressions.

[1]This chapter is an extended version of the paper by Farbmacher and Kögel (2017), published in *Applied Economics Letters*.

In Section 4.2, we provide a theoretical background for the situation when an upward bias in conventional standard errors occurs. There, we also discuss the difficulties in using standard tests for heteroscedasticity in such settings. We then propose a heteroscedasticity test which has better power properties. Section 4.3 presents the results of a series of Monte Carlo simulations based on data exhibiting this special form of heteroscedasticity. In Section 4.4, we use three test procedures to test for heteroscedasticity in Bonjour et al.'s (2003) data set. The Koenker variant of the Breusch-Pagan test and the White test do not reject the hypothesis of homoscedasticity, which is as expected, due to the special form of heteroscedasticity present. However, our proposed test rejects the null hypothesis, in favor of the special form of heteroscedasticity. Also in Section 4.4, we present the within-twin pair regressions using the appropriate standard errors. Section 4.5 concludes.

4.2 Inference Issues and Test Procedure

In the presence of heteroscedasticity, conventional standard errors (which assume homoscedasticity) can be biased up or down. The most common form of heteroscedasticity, where the residual variance rises in increasing regressor values, usually leads to conventional standard errors that are too small. When Wald tests based on these standard errors are insignificant, heteroscedasticity-robust standard errors do not change inference. On the other hand, inference is conservative in a setting with upward-biased conventional standard errors. Using heteroscedasticity-robust standard errors may change inference in this case.

Angrist and Pischke (2010) derive the condition for such an upward bias in the classical bivariate linear regression model[2]

$$y_i = \alpha + \beta x_i + e_i, \tag{4.1}$$

[2]A similar insight can be derived in the multivariate regression model by partialling out all other covariates.

where the true sampling variance for the OLS estimator $\hat{\beta}$ can be written as

$$\sigma^2_{\hat{\beta}} = \frac{1}{n}\frac{Var[e_i(x_i - \bar{x})]}{Var[x_i]^2}. \tag{4.2}$$

Under the assumption of homoscedasticity, $Var(e_i|x_i) = \sigma^2_e$, the equation simplifies to the conventional standard error

$$[\sigma^2_{\hat{\beta}}]_{conv} = \frac{1}{n}\frac{\sigma^2_e}{Var[x_i]}. \tag{4.3}$$

Thus,

$$[\sigma^2_{\hat{\beta}}]_{conv} > \sigma^2_{\hat{\beta}} \iff \sigma^2_e > \frac{Var[e_i(x_i - \bar{x})]}{Var[x_i]}. \tag{4.4}$$

Since

$$Var[e_i(x_i - \bar{x})] = E[e_i^2(x_i - \bar{x})^2] \tag{4.5}$$

$$= E[e_i^2]E[(x_i - \bar{x})^2] + Cov[e_i^2, (x_i - \bar{x})^2] \tag{4.6}$$

$$= \sigma^2_e Var[x_i] + Cov[e_i^2, (x_i - \bar{x})^2], \tag{4.7}$$

the relationship in (4.4) can further be rewritten as

$$[\sigma^2_{\hat{\beta}}]_{conv} > \sigma^2_{\hat{\beta}} \iff Cov[e_i^2, (x_i - \bar{x})^2] < 0. \tag{4.8}$$

An upward bias in conventional standard errors occurs if there is a negative covariance between the squared residual e_i^2 and the squared deviation of x_i from its mean $\bar{x}$. The further away x_i is from $\bar{x}$, the smaller becomes $Var[e_i|x_i] = E[e_i^2|x_i]$, the conditional

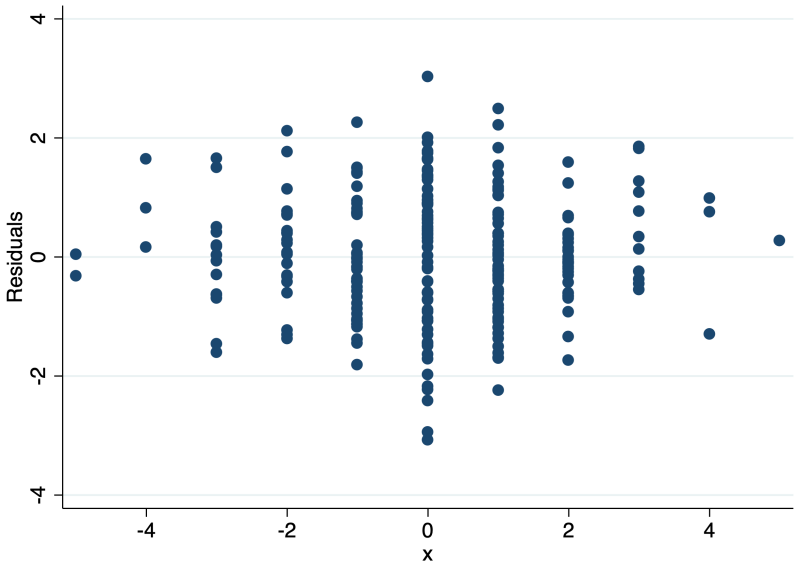
variance of residual e_i. When $Cov[e_i^2, (x_i - \bar{x})^2] < 0$, the corresponding scatter plot of e_i on the regressor x_i often resembles an ellipse. That is why we refer to this form of heteroscedasticity as *elliptical heteroscedasticity*. Panel A in Figure 4.1 illustrates the elliptical shape of the residuals based on simulated data, exhibiting elliptical heteroscedasticity.

If the data exhibit elliptical heteroscedasticity, the usual Wald tests for hypotheses about β in the bivariate regression model using conventional standard errors give an actual size smaller than the nominal size. Policy conclusions based on estimates with conventional standard errors are thus conservative. Conversely, Wald tests using heteroscedasticity-robust standard errors are size-correct and yield therefore valid policy conclusions. Furthermore, heteroscedasticity-robust Wald tests lead to power gains compared to tests using conventional standard errors in this case.

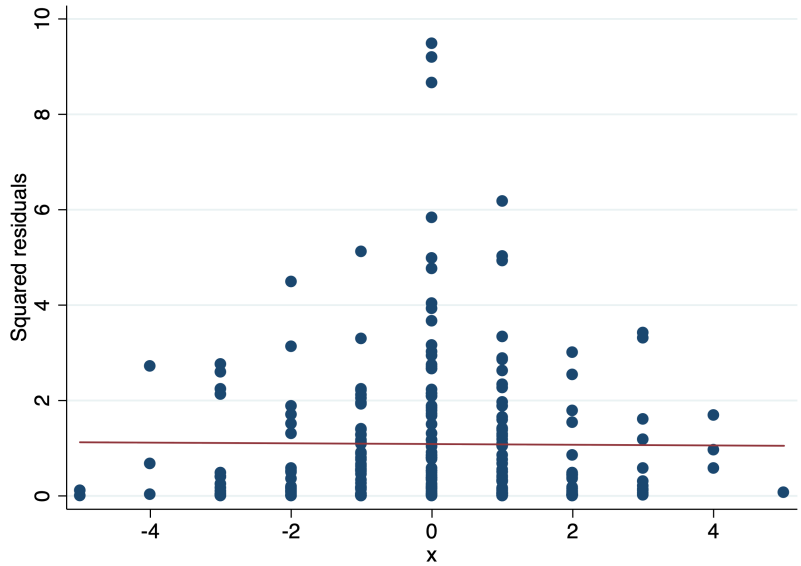
When elliptical heteroscedasticity is present, a reverse 'U'-shaped relation between the squared residual e_i^2 and the regressor x_i often occurs. Hence, statistical procedures testing for linear forms of heteroscedasticity, based on e_i^2 as the dependent variable, usually fail to detect elliptical heteroscedasticity. Panel B in Figure 4.1 illustrates how the linear regression line from the regression of e_i^2 on x_i is close to zero, as the squared residuals first rise and then fall in an increasing x_i. Therefore, tests such as the Breusch-Pagan (1979) test with x_i as the only independent variable included usually do not reject the hypothesis of homoscedasticity. In addition, more general tests, for example the White (1980) test, to detect also non-linear heteroscedasticity, do not give information about the form of heteroscedasticity that is present. This is because such test procedures test the null hypothesis of homoscedasticity against the unspecific alternative of no homoscedasticity. Moreover, due to their open formulation of null and alternative hypothesis, more general tests can possess a lower power in detecting elliptical heteroscedasticity.

Figure 4.1. Elliptical Heteroscedasticity Example

Panel A. Scatter Plot Illustrating the Elliptical Shape of the Residuals



Panel B. Regression of the Squared Residuals on x



Notes: The data are simulated data based on the data generating process in Section 4.3. $N = 250$; $a = 0.2$.

By exploiting the relationship in (4.8), we can test specifically for elliptical heteroscedasticity in the classical bivariate regression model. To derive our elliptical heteroscedasticity test, consider the regression

$$e_i^2 = \delta_0 + \delta_1(x_i - \bar{x})^2 + \xi_i, \tag{4.9}$$

where the squared residuals e_i^2 are obtained from the regression of y_i on x_i. Under elliptical heteroscedasticity, we know that $Cov[e_i^2, (x_i - \bar{x})^2] < 0$ and therefore

$$\delta_1 = \frac{Cov[e_i^2, (x_i - \bar{x})^2]}{Var[(x_i - \bar{x})^2]} < 0\,. \tag{4.10}$$

Thus, by using this knowledge, we can test specifically for elliptical heteroscedasticity. Our elliptical heteroscedasticity test conducts a one-sided Wald test for $H_0 : \delta_1 \geq 0$ against $H_a : \delta_1 < 0$ in the regression $e_i^2 = \delta_0 + \delta_1(x_i - \bar{x})^2 + \xi_i$. The hypotheses are H_0: *no elliptical heteroscedasticity* and H_a: *elliptical heteroscedasticity*.

4.3 Monte Carlo Simulations

To illustrate the testing issues arising from elliptical heteroscedasticity, we run a series of Monte Carlo simulations. The design of our Monte Carlo simulations is based on the following data generating process.

$$y_i = 0.04x_i + e_i$$

$$e_i = \sqrt{\frac{1}{\{(x_i - \bar{x})^2 + 0.1\}^a}}\,\epsilon_i$$

$$x_i = [x_i^*],\ x_i^* \sim N(0.04, 1.8^2),\ \epsilon_i \sim N(0,1),$$

$$a = 0,\ 0.01,\ 0.02,\ 0.03,\ \ldots,\ 0.5$$

We choose the model so that the shape of the resulting y-x scatter plot resembles Panel A of Figure 1 by Amin (2011), which is replicated in Panel A of Figure 4.4 in the next section. For values of a between 0.15 and 0.3, the y-x scatter plot is most similar to Panel A. The operator $[.]$ rounds x_i^* to the nearest integer. Hence, x_i is an integer, just as the within-twin difference in years of schooling in Bonjour et al. (2003). Furthermore, also in accordance with the within-twin difference in years of schooling, the values of x_i are centered around the mean $\bar{x}$. The structure of the error term e_i implies that $Cov[e_i^2, (x_i - \bar{x})^2] < 0$ if $a > 0$. The larger is the parameter a, the more negative is the covariance between e_i^2 and $(x_i - \bar{x})^2$, and therefore the stronger is the upward bias caused by elliptical heteroscedasticity. For $a = 0$, the error term is homoscedastic. The number of observations is set to $N = 214$, as in Bonjour et al.'s (2003) data set, and additionally to $N = 2,140$. The number of replications is 10,000.

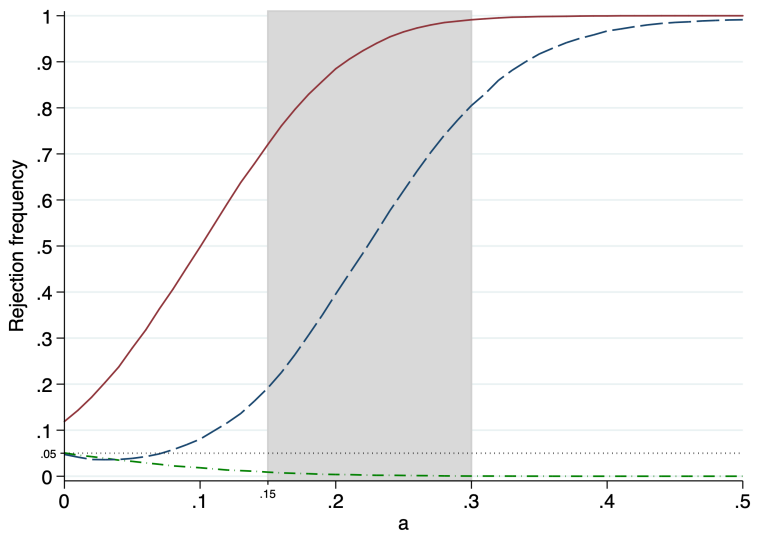
In each simulation, we evaluate the size and power of three different tests for heteroscedasticity: the Koenker (1981) variant of the Breusch-Pagan (1979) test, which drops the assumption of normality of the error term, with x as the independent variable, the White test, and our elliptical heteroscedasticity test introduced in Section 4.2. In addition, we compare the size and power for the parameter of interest in the causal model, using Wald tests for the hypothesis $H_0 : \beta = k$ against $H_a : \beta \neq k$, for $k = 0, 0.01, 0.02, \ldots, 0.12$, in the regression of y_i on x_i using robust and conventional standard errors.

Figure 4.2 shows the power plots for the heteroscedasticity tests. The simulation with $a = 0$ gives the actual size of each test. While the rejection frequency of the Breusch-Pagan and White test is close to the given significance level of $\alpha = 5\%$ for $N = 214$, the actual size of the elliptical heteroscedasticity test is above this value, with 11.9%. However, the actual test size for the latter test approaches the theoretically given significance level for larger numbers of observations. The simulation with $N = 2,140$ yields an actual size of 7.7% for the elliptical heteroscedasticity test.

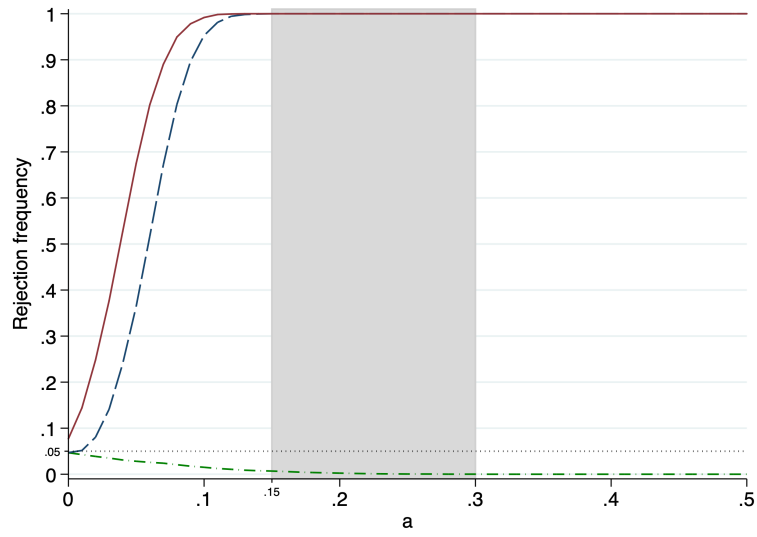
For $a > 0$, Figure 4.2 displays the power of each test. The rejection frequency of the White test and our elliptical heteroscedasticity test increases with stronger elliptical heteroscedasticity, i.e., with increasing values of a. Compared to the elliptical heteroscedasticity test, the White test performs worse in detecting heteroscedasticity,

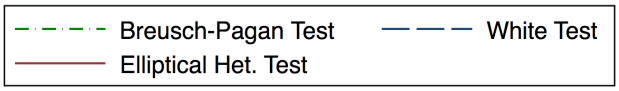
Figure 4.2. Power Plots for the Heteroscedasticity Tests

Panel A. $N = 214$



Panel B. $N = 2,140$





Notes: The plots show the heteroscedasticity tests' rejection frequencies of H_0 at $\alpha = 0.05$ from the Monte Carlo Simulations. For the Breusch-Pagan and White test: H_0: Homoscedasticity; H_a: Heteroscedasticity. For the elliptical heteroscedasticity test: H_0: No elliptical heteroscedasticity; H_a: Elliptical heteroscedasticity. For $a = 0$, the data exhibit homoscedasticity. For $a > 0$, the data exhibit elliptical heteroscedasticity. For values of a in the gray shaded area, the resulting y-x scatter plot of the generated data is most similar to Amin's (2011) Panel A in Figure 1, which is replicated in Figure 4.4.

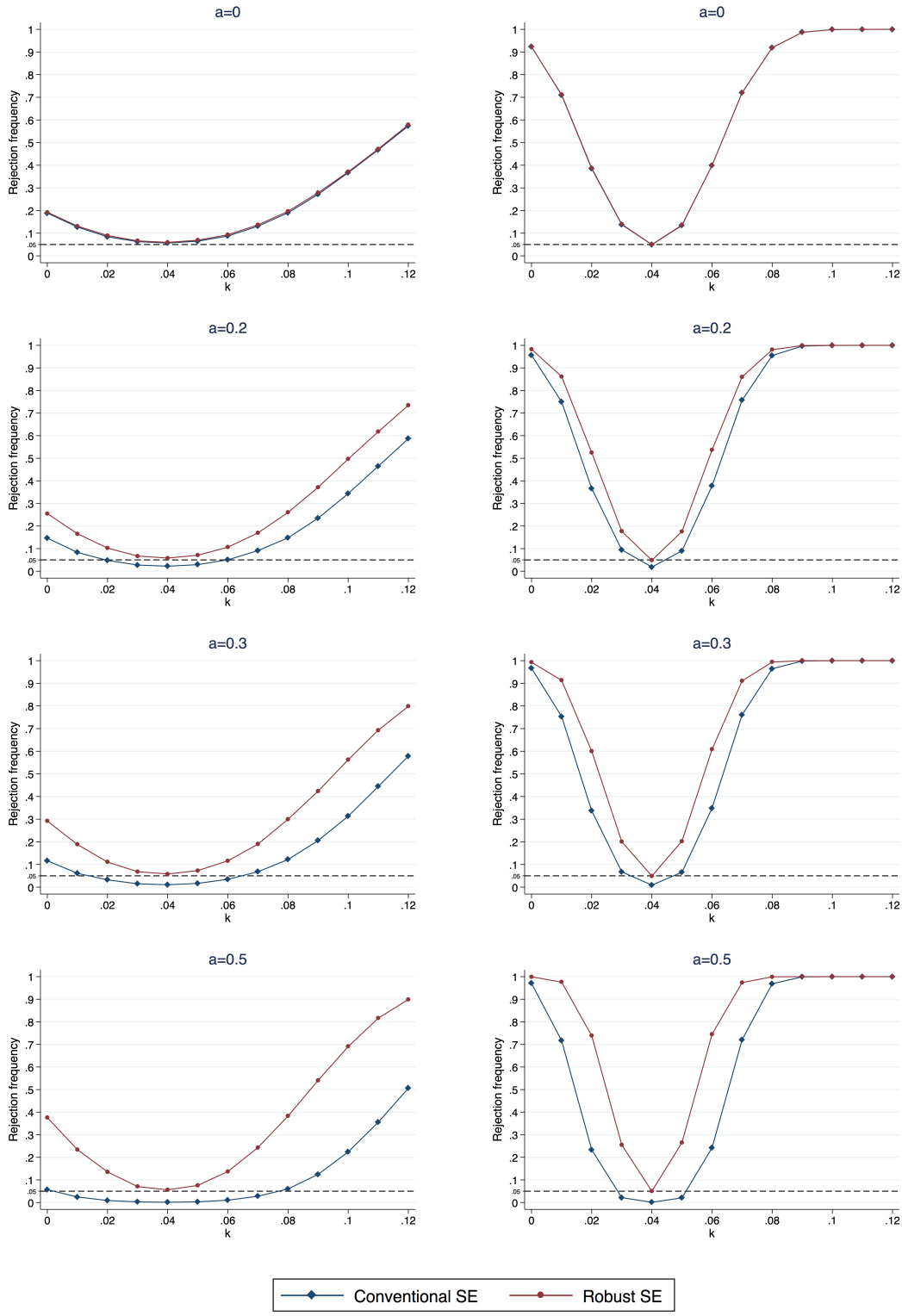
although the difference in power gets smaller for larger values of a. In contrast to the elliptical heteroscedasticity test, the White test does not have elliptical heteroscedasticity as the alternative hypothesis, but rather heteroscedasticity in general. The less specific formulation of H_a may explain the White test's worse performance. The Breusch-Pagan test has considerably smaller rejection frequencies than the two other tests throughout the whole range of $a > 0$. For $N = 214$ and $N = 2,140$, it does not reach a power of 5% for any given positive value of a. This result may be related to the fact that the basic specification of the Breusch-Pagan test is for detecting linear forms of heteroscedasticity, whereas elliptical heteroscedasticity implies a non-linear form of heteroscedasticity.

Figure 4.3 displays the power and size of the Wald tests. The actual size of the tests is given at $H_0 : \beta = 0.04$. Under homoscedasticity, $a = 0$, both test versions' actual sizes are close to the given significance level of 5%, for $N = 214$ and $N = 2,140$. In the presence of heteroscedasticity, $a > 0$, the Wald tests using robust standard errors yield also an actual size around 5%. The size of the Wald tests using conventional standard errors, however, decreases with increasing a, such that at $a = 0.5$, the actual size for both sample sizes is only around 0.1%. Hence, t-tests with conventional standard errors do not reject the correct null hypothesis often enough for $a > 0$. This is due to the upward bias in conventional standard errors in this case.

For $H_0 : \beta \neq 0.04$, Figure 4.3 shows the power of the Wald tests. At $a = 0$, the power curves of both tests are almost the same. However, an ever increasing gap between them generally arises as a gets larger. The Wald test using robust standard errors becomes more powerful whereas the test using conventional standard errors loses power. The loss in power can be attributed to the increasing upward bias in conventional standard errors for rising values of $a > 0$. As expected, the tests' power gets larger the further away the null hypothesis is from the true parameter $\beta = 0.04$, and the tests have a higher power for $N = 2,140$ than for $N = 214$.

Figure 4.3. Power Plots for Wald Tests Using Conventional and Robust Standard Errors

$N = 214$ $N = 2140$

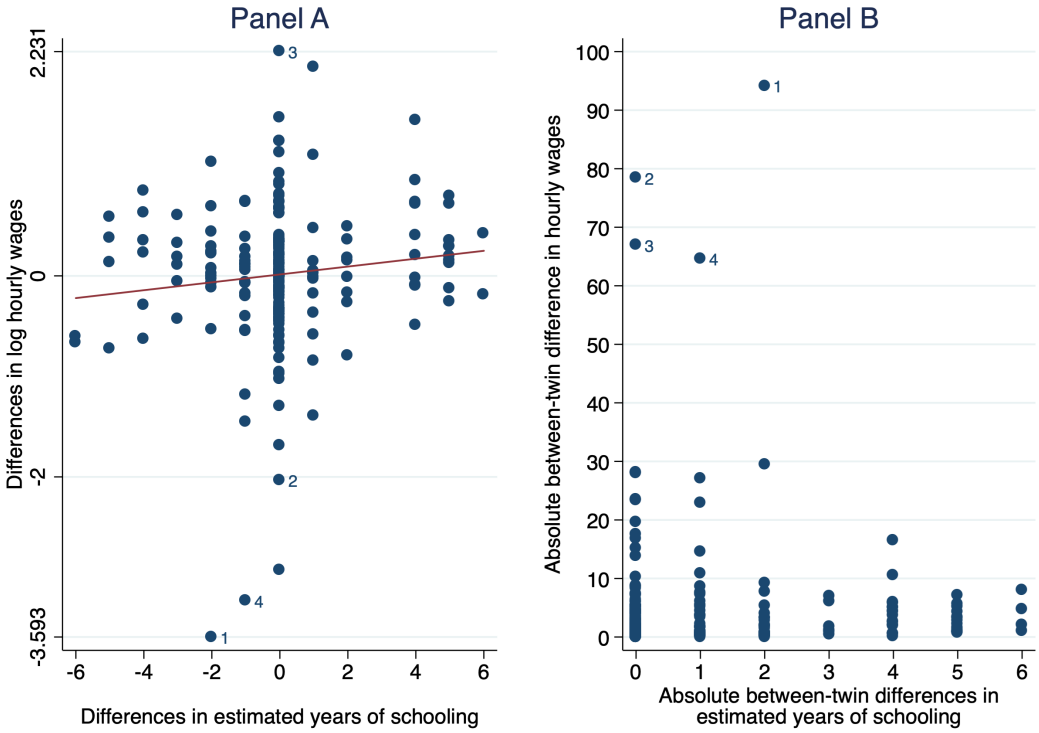


Notes: The plots show the Wald tests' rejection frequencies for $H_0 : \beta = k$ against $H_a : \beta \neq k$, $k = 0, 0.01, 0.02, \ldots, 0.12$, at $\alpha = 0.05$ from the Monte Carlo Simulations. The true parameter β equals 0.04. For $a = 0$, the data exhibit homoscedasticity. For $a > 0$, the data exhibit elliptical heteroscedasticity.

4.4 Empirical Illustration: Returns to Education

As discussed in the introduction, Amin (2011) replicated Bonjour et al.'s (2003) estimates of the return to education and performed similar regressions where he excluded outliers from their sample of monozygotic twins. Specifically, he excluded up to four twin pair outliers on the basis of the absolute between-twin difference in hourly wages. Figure 4.4, which replicates Figure 1 by Amin (2011), illustrates which data points he removed. Panel A already suggests that the data exhibit the elliptical heteroscedasticity discussed in Section 4.2, which leads to an upward bias in conventional standard errors.[3]

Figure 4.4. Replication of Figure 1, Amin (2011)



Notes: $N = 214$. The data are from Bonjour et al. (2003). The twin pairs in the data set are from the St. Thomas' UK Adult Twin Registry. The labelled observations correspond to the outliers which Amin (2011) excluded.

[3]We noticed that outlier number 2 in Amin's (2011) Panel B does not correspond to the data point labelled 2 in his Panel A. As shown in our Figure 4.4, observation number 2 in Panel B is actually the data point with a difference in log hourly wages of approximately -2 instead of the point at approximately -3. Despite this graphical error, Amin (2011) excluded the correct observations in his analysis.

To test for the presence of heteroscedasticity, we perform the three tests outlined in Section 4.2 for all within-twin pair OLS and IV regressions in columns (3), (4), and (7), (8) of Table 1 in Amin (2011). In all regressions, the dependent variable is the within-twin difference in log hourly wages. The regressor of interest is the within-twin difference in self-reported education. In the IV regressions, this variable is instrumented by the within-twin difference in the co-twin's report of the other twin's education. The regressions in columns (7) and (8) include the covariates within-twin difference in marital status, current job tenure, part-time status, and whether a person lives in London or the south-east of the UK.

Table 4.1 provides the p-values for the Koenker variant of the Breusch-Pagan test with within-twin difference in years of schooling as the only independent variable, the White test, and our proposed elliptical heteroscedasticity test. In the regressions including covariates, we partialled them out before testing. The elliptical heteroscedasticity test rejects for all regression specifications at least at the 10 percent level. In contrast, the Breusch-Pagan and White test do not reject the hypothesis of homoscedasticity in any regression. This may be attributed to the difficulties and lower power in detecting elliptical heteroscedasticity when using more general tests as discussed in Section 4.2. Based on our proposed elliptical heteroscedasticity test, there is thus evidence for the presence of elliptical heteroscedasticity in the data. This suggests that conventional standard errors are incorrect and may lead to false policy conclusions.

Table 4.2 shows the return to education estimates. Our replication results are very similar to the estimates by Amin (2011). The regressions based on the full sample are the ones which Bonjour et al. (2003) also performed. Amin (2011) and Bonjour et al. (2003) both used conventional standard errors. In addition to the replications using conventional standard errors, Table 4.2 reports robust standard errors and the corresponding significance levels. In all but two regressions, the robust standard error is smaller than the conventional one. This result is in line with the suspicion that elliptical heteroscedasticity is present in the data, which causes an upward bias in conventional standard errors. It also supports the conclusions from our elliptical heteroscedasticity test.

In many regressions where the estimate for the parameter of interest is insignificant using conventional standard errors, it becomes significant at the 5 percent or 10 percent

Table 4.1. Heteroscedasticity Test Results for the Within-Twin Pair Regressions in Table 1, Amin (2011)

		Within-twin pair without covariates		Within-twin pair with covariates	
		OLS	IV	OLS	IV
Sample		(3)	(4)	(7)	(8)
Full Bonjour et al. (2003) data set	Breusch-Pagan Test	0.3645	0.3124	0.6090	0.5435
	White Test	0.4805	0.4581	0.5300	0.5603
	Elliptical Het. Test	0.0096	0.0234	0.0113	0.0295
	Observations	214	214	187	187
Drop if abs. wage difference $> £90$	Breusch-Pagan Test	0.7221	0.6719	0.8713	0.8276
	White Test	0.4982	0.4906	0.5451	0.5491
	Elliptical Het. Test	0.0109	0.0107	0.0194	0.0211
	Observations	213	213	186	186
Drop if abs. wage difference $> £75$	Breusch-Pagan Test	0.7207	0.6737	0.7799	0.7437
	White Test	0.5488	0.5421	0.6034	0.6075
	Elliptical Het. Test	0.0176	0.0173	0.0341	0.0383
	Observations	212	212	185	185
Drop if abs. wage difference $> £65$	Breusch-Pagan Test	0.7143	0.7065	0.8136	0.8126
	White Test	0.6147	0.6101	0.6562	0.6559
	Elliptical Het. Test	0.0297	0.0286	0.0464	0.0461
	Observations	211	211	184	184
Drop if abs. wage difference $> £60$	Breusch-Pagan Test	0.9861	0.9466	0.8177	0.8014
	White Test	0.6310	0.6237	0.7247	0.7194
	Elliptical Het. Test	0.0586	0.0534	0.0824	0.0813
	Observations	210	210	183	183

Notes: The data are from Bonjour et al. (2003). The table reports p-values from heteroscedasticity tests for each sample and regression specification. The column numbers indicate which column in Amin's (2011) Table 1 the results refer to. See the notes for Table 4.2 for further information on the regression specifications.

Table 4.2. Replication and Re-Estimation of the Within-Twin Pair Regressions in Table 1, Amin (2011)

		Within-twin pair without covariates		Within-twin pair with covariates	
		OLS	IV	OLS	IV
Sample		(3)	(4)	(7)	(8)
	$\hat{\beta}_{education}$	0.039	0.077	0.039	0.082
Full Bonjour et al.	Conventional SE	(0.023)*	(0.033)**	(0.024)	(0.036)**
(2003) data set	Robust SE	(0.018)**	(0.039)**	(0.018)**	(0.043)*
	Observations	214	214	187	187
	$\hat{\beta}_{education}$	0.032	0.050	0.034	0.053
Drop if abs. wage	Conventional SE	(0.021)	(0.031)	(0.023)	(0.033)
difference $>$ £90	Robust SE	(0.016)**	(0.027)*	(0.017)**	(0.030)*
	Observations	213	213	186	186
	$\hat{\beta}_{education}$	0.032	0.050	0.036	0.055
Drop if abs. wage	Conventional SE	(0.021)	(0.030)*	(0.022)	(0.032)*
difference $>$ £75	Robust SE	(0.016)**	(0.027)*	(0.017)**	(0.030)*
	Observations	212	212	185	185
	$\hat{\beta}_{education}$	0.032	0.036	0.036	0.039
Drop if abs. wage	Conventional SE	(0.020)	(0.029)	(0.021)*	(0.031)
difference $>$ £65	Robust SE	(0.016)**	(0.022)	(0.017)**	(0.024)
	Observations	211	211	184	184
	$\hat{\beta}_{education}$	0.028	0.036	0.036	0.041
Drop if abs. wage	Conventional SE	(0.019)	(0.027)	(0.019)*	(0.028)
difference $>$ £60	Robust SE	(0.016)*	(0.022)	(0.016)**	(0.023)*
	Observations	210	210	183	183

Notes: The data are from Bonjour et al. (2003). The table reports estimates of the return to one additional year of education based on a sample of monozygotic twins from the UK. The columns are numbered according to the corresponding columns in Amin's (2011) Table 1. The dependent variable is the within-twin difference in log hourly wages. The regressor of interest is the within-twin difference in years of schooling. The covariates are the within-twin differences in the following variables: marital status, current job tenure, part-time status, and whether a person lives in London or the south-east of the UK. In the IV regressions, the within-twin difference in self-reported education is instrumented by the within-twin difference in the co-twin's report of the other twin's education.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

level when using robust standard errors. With conventional standard errors, 13 out of the 20 regressions yield an insignificant parameter estimate. In contrast, only in three out of the 20 regressions do we fail to find a return to education estimate that is significantly different from zero when using robust standard errors. In particular, all point estimates based on the full sample as well as the sample excluding observations with an absolute wage difference of more than £90 and £75, respectively, are significant at the usual levels. Regarding the regressions based on samples with three or four outliers removed, three more estimates turn significant at least at the 10 percent level with robust standard errors compared to the results which use conventional standard errors. Thus, when using robust rather than conventional standard errors, the vast majority of regressions in Table 4.2 suggests that there is a positive return to education.

4.5 Conclusion

In this study, we discuss the conditions under which conventional, homoscedasticity-assuming, standard errors are upward biased. In such settings, standard tests of heteroscedasticity may fail and leave the heteroscedasticity undetected. When Wald tests based on downward-biased conventional standard errors are insignificant, heteroscedasticity-robust standard errors do not change inference. On the other hand, inference is conservative in a setting with upward-biased conventional standard errors. We discuss the power gains when using robust standard errors in this case and also potential problems of heteroscedasticity tests. In Monte Carlo simulations we show that our proposed heteroscedasticity test has a higher power in detecting this special form of heteroscedasticity. In our application only this test detects the heteroscedasticity, and using then the appropriate standard errors leads to different test decisions.

Bibliography

Abadie, A., and Cattaneo, M. D. (2018). Econometric Methods for Program Evaluation. *Annual Review of Economics*, *10*, 465–503.

Åkerstedt, T. (2006). Psychosocial Stress and Impaired Sleep. *Scandinavian Journal of Work, Environment & Health*, *32*(6), 493–501.

Amin, V. (2011). Returns to Education: Evidence from UK Twins: Comment. *American Economic Review*, *101*(4), 1629–1635.

Angel, S., Disslbacher, F., Humer, S., and Schnetzer, M. (2019). What Did You Really Earn Last Year?: Explaning Measurement Error in Survey Income Data. *Journal of the Royal Statistical Society, Series A*, *forthcoming*.

Angrist, J., and Pischke, J.-S. (2010). *A Note on Bias in Conventional Standard Errors under Heteroskedasticity*. Mathematical Note. Retrieved from http://econ.lse.ac.uk/staff/spischke/mhe/josh/Notes%20on%20conv%20std%20error.pdf

Athey, S., and Imbens, G. (2016). Recursive Partitioning for Heterogeneous Causal Effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360.

Athey, S., and Imbens, G. (2019). Machine Learning Methods That Economists Should Know About. *Annual Review of Economics*, *11*, 685–725.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized Random Forests. *Annals of Statistics*, *47*(2), 1148–1178.

Barr, M. S. (2012). *No Slack: The Financial Lives of Low-Income Americans*. Washington, DC: Brookings Institution Press.

BIBLIOGRAPHY

Battistin, E., Miniaci, R., and Weber, G. (2003). What Do We Learn from Recall Consumption Data? *Journal of Human Resources*, *38*(2), 354–385.

Battistin, E., and Padula, M. (2016). Survey Instruments and the Reports of Consumption Expenditures: Evidence from the Consumer Expenditure Surveys. *Journal of the Royal Statistical Society, Series A*, *179*(2), 559–581.

Benson, H., and Proctor, W. (2010). *Relaxation Revolution: Enhancing Your Personal Health Through the Science and Genetics of Mind Body Healing*. New York: Scribner.

Bertrand, M., Mullainathan, S., and Shafir, E. (2004). A Behavioral-Economics View of Poverty. *American Economic Review*, *94*(2), 419–423.

Bertrand, M., Mullainathan, S., and Shafir, E. (2006). Behavioral Economics and Marketing in Aid of Decision Making among the Poor. *Journal of Public Policy and Marketing*, *25*(1), 8–23.

Binder, C. C. (2017). Measuring Uncertainty Based on Rounding: New Method and Application to Inflation Expectations. *Journal of Monetary Economics*, *90*, 1–12.

Bonjour, D., Cherkas, L. F., Haskel, J. E., Hawkes, D. D., and Spector, T. D. (2003). Returns to Education: Evidence from U.K. Twins. *American Economic Review*, *93*(5), 1799–1812.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., and Zuber, S. (2013b). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology*, *42*(4), 992–1001.

Börsch-Supan, A., Brandt, M., Litwin, H., and Weber, G. (Eds.). (2013a). *Active Ageing and Solidarity between Generations in Europe: First Results from SHARE after the Economic Crisis*. Berlin: De Gruyter.

Börsch-Supan, A., Bristle, J., Andersen-Ranberg, K., Brugiavini, A., Jusot, F., Litwin, H., and Weber, G. (Eds.). (2019). *Health and Socio-Economic Status over the Life Course. First Results from SHARE Waves 6 and 7*. Berlin: De Gruyter.

BIBLIOGRAPHY

Bound, J., Brown, C., and Mathiowetz, N. (2001). Measurement Error in Survey Data. In J. J. Heckman, and E. Leamer (Eds.), *Handbook of Econometrics* (Vol. 5, pp. 3705–3843). Amsterdam: Elsevier.

Bourke, N., Horowitz, A., and Roche, T. (2012). *Payday Lending in America: Who Borrows, Where They Borrow, and Why*. Payday Lending in America Series, Pew Charitable Trusts.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Breusch, T. S., and Pagan, A. R. (1979). A Simple Test for Heteroskedasticity and Random Coefficient Variation. *Econometrica*, *47*(5), 1287–1294.

Browning, M., Crossley, T. F., and Weber, G. (2003). Asking Consumption Questions in General Purpose Surveys. *Economic Journal*, *113*(491), F540–F567.

Browning, M., Crossley, T. F., and Winter, J. (2014). The Measurement of Household Consumption Expenditures. *Annual Review of Economics*, *6*, 475–501.

Butterworth, P., Rodgers, B., and Windsor, T. D. (2009). Financial Hardship, Socio-Economic Position and Depression: Results from the PATH Through Life Survey. *Social Science & Medicine*, *69*(2), 229–237.

Cameron, A. C., and Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, *50*(2), 317–372.

Carrington, W., Dahl, M., and Falk, J. (2013). *Growth in Means-Tested Programs and Tax Credits for Low-Income Households*. Congressional Budget Office, Congress of the United States.

Carvalho, L. S., Meier, S., and Wang, S. W. (2016). Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday. *American Economic Review*, *106*(2), 260–284.

Castellari, E., Cotti, C., Gordanier, J., and Ozturk, O. (2017). Does the Timing of Food Stamp Distribution Matter? A Panel-Data Analysis of Monthly Purchasing Patterns of US Households. *Health Economics*, *26*(11), 1380–1393.

Clotfelter, C. T., Cook, P. J., Edell, J. A., and Moore, M. (1999). *State Lotteries at the Turn of the Century: Report to the National Gambling Impact Study Commission*. Research Report, Duke University.

Cole, N., and Lee, E. (2005). *Analysis of EBT Redemption Patterns: Methods and Detailed Tables*. Abt Associates. Cambridge, MA.

Couper, M. P. (2013). Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys. *Survey Research Methods*, *7*(3), 145–156.

DeNavas-Walt, C., and Proctor, B. D. (2015). *Income and Poverty in the United States: 2014*. Current Population Report No. P60-252, U.S. Census Bureau.

Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., Aptowicz, C., and Pack, A. I. (1997). Cumulative Sleepiness, Mood Disturbance, and Psychomotor Vigiliance Performance Decrements During a Week of Sleep Restricted to 4–5 Hours per Night. *Sleep*, *20*(4), 267–277.

Dube, S. R., Asman, K., Malarcher, A., and Carabollo, R. (2009). Cigarette Smoking among Adults and Trends in Smoking Cessation – United States, 2008. *Morbidity and Mortality Weekly Report*, *58*(44), 1227–1232.

Edin, K., Boyd, M., Mabli, J., Ohls, J., Worthington, J., Greene, S., Redel, N., and Sridharan, S. (2013). *SNAP Food Security In-Depth Interview Study: Final Report*. Food and Nutrition Service, U.S. Department of Agriculture.

Edin, K., and Lein, L. (1997). *Making Ends Meet: How Single Mothers Survive Welfare and Low-Wage Work*. New York: Russel Sage Foundation.

Einav, L., and Levin, J. (2014). The Data Revolution and Economic Analysis. In J. Lerner, and S. Stern (Eds.), *Innovation Policy and the Economy* (Vol. 14, pp. 1–24). Chicago: University of Chicago Press.

Farbmacher, H., and Kögel, H. (2017). Testing under a Special Form of Heteroscedasticity. *Applied Economics Letters*, *24*(4), 264–268.

Ferrie, J. E., Martikainen, P., Shipley, M. J., and Marmot, M. G. (2005). Self-Reported Economic Difficulties and Coronary Events in Men: Evidence from the Whitehall II Study. *International Journal of Epidemiology*, *34*(3), 640–648.

Foley, C. F. (2011). Welfare Payments and Crime. *Review of Economics and Statistics*, *93*(1), 97–112.

Fontenot, K., Semega, J., and Kollar, M. (2018). *Income and Poverty in the United States: 2017*. U.S. Census Bureau. Washington, DC.

Goldin, J., Homonoff, T., and Meckel, K. (2016). *Is there an Nth of the Month Effect? The Timing of SNAP Issuance, Food Expenditures, and Grocery Prices*. Working Paper.

Gray, K. F., and Kochhar, S. (2015). *Characteristics of Supplemental Nutrition Assistance Program Households: Fiscal Year 2014*. Nutrition Assistance Program Report Series, No. SNAP-15-CHAR, Food and Nutrition Service, U.S. Department of Agriculture.

Groves, R. M. (2011). Three Eras of Survey Research. *Public Opinion Quarterly*, *75*(5), 861–871.

Gunasekara, F. I., Carter, K. N., Crampton, P., and Blakely, T. (2013). Income and Individual Deprivation as Predictors of Health over Time. *International Journal of Public Health*, *58*(4), 501–511.

Hamrick, K. S., and Andrews, M. (2016). SNAP Participants' Eating Patterns over the Benefit Month: A Time Use Perspective. *PLoS ONE*, *11*(7), e0158422.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

Hastings, J., and Washington, E. (2010). The First of the Month Effect: Consumer Behavior and Store Responses. *American Economic Journal: Economic Policy*, *2*(2), 142–162.

Hoderlein, S., Siflinger, B., and Winter, J. (2015). *Identification of Structural Models in the Presence of Measurement Error Due to Rounding in Survey Responses*. Working Paper No. 869, Department of Economics, Boston College.

Hoderlein, S., and Winter, J. (2010). Structural Measurement Errors in Nonseparable Models. *Journal of Econometrics*, *157*(2), 432–440.

Hokayem, C., and Heggeness, M. L. (2014). *Living in Near Poverty in the United States: 1966–2012*. Current Population Report No. P60-248, U.S. Census Bureau.

Hoynes, H. W., and Schanzenbach, D. W. (2009). Consumption Reponses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program. *American Economic Journal: Applied Economics*, *1*(4), 109–139.

Hoynes, H. W., and Schanzenbach, D. W. (2016). US Food and Nutrition Programs. In R. Moffitt (Ed.), *Economics of Means-Tested Transfer Programs in the United States, Volume 1* (pp. 219–301). Chicago: University of Chicago Press.

Johnson, D. S., Parker, J. A., and Souleles, N. S. (2006). Household Expenditure and the Income Tax Rebates of 2001. *American Economic Review*, *96*(5), 1589–1610.

Juster, F. T., and Suzman, R. (1995). An Overview of the Health and Retirement Study. *Journal of Human Resources*, *30*, S7–S56.

Kahn, J. R., and Pearlin, L. I. (2006). Financial Strain over the Life Course and Health among Older Adults. *Journal of Health and Social Behavior*, *47*(1), 17–31.

Kahn, M., Sheppes, G., and Sadeh, A. (2013). Sleep and Emotions: Bidirectional Links and Underlying Mechanisms. *International Journal of Psychophysiology*, *89*(2), 218–228.

Koenker, R. (1981). A Note on Studentizing a Test for Heteroscedasticity. *Journal of Econometrics*, *17*(1), 107–112.

Mani, A., Mullainathan, S., Shafir, E., and Zhao, J. (2013). Poverty Impedes Cognitive Function. *Science*, *341*(6149), 976–980.

Manski, C. F., and Molinari, F. (2010). Rounding Probabilistic Expectations in Surveys. *Journal of Business and Economic Statistics*, *28*(2), 219–231.

Mastrobuoni, G., and Weinberg, M. (2009). Heterogeneity in Intra-Monthly Consumption Patterns, Self-Control, and Savings at Retirement. *American Economic Journal: Economic Policy*, *1*(2), 163–189.

Meyer, B. D., Mok, W. K. C., and Sullivan, J. X. (2015). Household Surveys in Crisis. *Journal of Economic Perspectives*, *29*(4), 199–226.

Ouellette, T., Burstein, N., Long, D., and Beecroft, E. (2004). *Measures of Material Hardship, Final Report*. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services.

Paiva, T., Gaspar, T., and Matos, M. G. (2015). Sleep Deprivation in Adolescents: Correlations with Health Complaints and Health-Related Quality of Life. *Sleep Medicine*, *16*(4), 521–527.

Robins, J. M., and Rotnitzky, A. (1995). Semiparametric Efficiency in Multivariate Regression Models with Missing Data. *Journal of the American Statistical Association*, *90*(429), 122–129.

Ross, J. S., Bernheim, S. M., Bradley, E. H., Teng, H.-M., and Gallo, W. T. (2007). Use of Preventive Care by the Working Poor in the United States. *Preventive Medicine*, *44*(3), 254–259.

Rossmann, T. (2019). *Economic Uncertainty and Subjective Inflation Expectations*. CRC TRR 190 Discussion Paper No. 160.

Ruud, P., Schunk, D., and Winter, J. (2014). Uncertainty Causes Rounding: An Experimental Study. *Experimental Economics*, *17*(3), 391–413.

Saslow, E. (2013, March 16). Food Stamps Put Rhode Island Town on Monthly Boom-and-Bust Cycle. *Washington Post*. Retrieved from https://www.washingtonpost.com/national/food-stamps-put-rhode-island-town-on-monthly-boom-and-bust-cycle/2013/03/16/08ace07c-8ce1-11e2-b63f-f53fb9f2fcb4_story.html?utm_term=.6e47d4d87e47

Schilbach, F., Schofield, H., and Mullainathan, S. (2016). The Psychological Lives of the Poor. *American Economic Review: Papers & Proceedings*, *106*(5), 435–440.

Schulz, A. J., Israel, B. A., Zenk, S. N., Parker, E. A., Lichtenstein, R., Shellman-Weir, S., and Klem, A. B. L. (2006). Psychosocial Stress and Social Support as Mediators of Relationships between Income, Length of Residence and Depressive Symptoms among African American Women on Detroit's Eastside. *Social Science & Medicine*, *62*(2), 510–522.

Shapiro, J. M. (2005). Is there A Daily Discount Rate? Evidence from the Food Stamp Nutrition Cycle. *Journal of Public Economics*, *89*(2-3), 303–325.

Shapiro, M. D., and Slemrod, J. (1995). Consumer Response to the Timing of Income: Evidence from a Change in Tax Withholding. *American Economic Review*, *85*(1), 274–283.

Stephens, M. Jr. (2003). "3rd of tha Month": Do Social Security Recipients Smooth Consumption between Checks? *American Economic Review*, *93*(1), 406–422.

Stephens, M. Jr. (2006). Paycheque Receipt and the Timing of Consumption. *Economic Journal*, *116*(513), 680–701.

Stronks, K., van de Mheen, H. D., and Mackenbach, J. P. (1998). A Higher Prevalence of Health Problems in Low Income Groups: Does It Reflect Relative Deprivation? *Journal of Epidemiology and Community Health*, *52*(9), 548–557.

Szanton, S. L., Thorpe, R. J., and Whitfield, K. (2010). Life-Course Financial Strain and Health in African-Americans. *Social Science & Medicine*, *71*(2), 259–265.

Tibshirani, J., Athey, S., Wager, S., Friedberg, R., Miner, L., and Wright, M. (2018). *grf: Generalized Random Forests (Beta). R package version 0.9.6.*

Tourangeau, R., Rips, L. J., and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Truffer, C. J., Klemm, J. D., Wolfe, C. J., Rennie, K. E., and Shuff, J. F. (2013). *2013 Actuarial Report on the Financial Outlook for Medicaid*. Centers for Medicare & Medicaid Services, U.S. Department of Health and Human Services.

Tucker-Seeley, R. D., Li, Y., Subramanian, S. V., and Sorensen, G. (2009). Financial Hardship and Mortality among Older Adults Using the 1996–2004 Health and Retirement Study. *Annals of Epidemiology*, *19*(12), 850–857.

U.S. Department of Agriculture. (2006). *An Analysis of Food Stamp Benefit Redemption Patterns*. Food and Nutrition Service, U.S. Department of Agriculture.

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28.

Wager, S., and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, *48*(4), 817–838.

Wickrama, K. A. S., Lorenz, F. O., Conger, R. D., Elder, G. H. Jr., Abraham, W. T., and Fang, S.-A. (2006). Changes in Family Financial Circumstances and the Physical Health of Married and Recently Divorced Mothers. *Social Science & Medicine*, *63*(1), 123–136.

Wilde, P. E., and Ranney, C. K. (2000). The Monthly Food Stamp Cycle: Shopping Frequency and Food Intake Decisions in an Endogenous Switching Regression Framework. *American Journal of Agricultural Economics*, *82*(1), 200–213.

Wolkwitz, K., and Leftin, J. (2008). *Characteristics of Food Stamp Households: Fiscal Year 2007*. Nutrition Assistance Program Report Series, No. FSP-08-CHAR, Food and Nutrition Service, U.S. Department of Agriculture.

World Health Organization. (2000). *Obesity: Preventing and Managing the Global Epidemic*. WHO Technical Report Series, No. 894, World Health Organization.

Eidesstattliche Versicherung

Ich versichere hiermit eidesstattlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht. Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht. Sofern ein Teil der Arbeit aus bereits veröffentlichten Papers besteht, habe ich dies ausdrücklich angegeben.

Datum: 18. September 2019

Unterschrift: Heinrich Kögel