

UNIVERSITY OF BIRMINGHAM

Errors in Financial Expectations, Over-Indebtedness and Unhealthy Behaviours: Evidence from Russia and South Africa

by

Wei Jin

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

Birmingham Business School

College of Social Sciences

University of Birmingham

October 2020

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

This thesis presents three empirical studies on errors in financial expectations. All studies employ micro-level survey data.

In the first study, using the Russian Longitudinal Monitoring Survey over the period 2002-2016, I examine the determinants of individuals' errors in financial expectations, focusing on the role of past income changes and job insecurity. Consistent with the over-extrapolation of their past experience, I observe that individuals become over-optimistic after their financial situation improves, and over-pessimistic after their financial situation worsens. Furthermore, I find that individuals over-extrapolate income gains more than income losses. Turning to the role of job insecurity, I observe that individuals who are confident at finding a new job if laid-off are more likely to show over-optimism. In addition, I document that the more confident respondents are at finding a new job, the larger the forecast error they make. These results are robust to using a wide range of estimators including fixed-effects estimators, fixed-effects instrumental variable estimators and correlated random effects Probit estimators. They are also robust to using different proxies for errors in financial expectations.

In the second study, using the Russia Longitudinal Monitoring Survey, which includes over 120,000 observations spanning 16 years, I investigate the relationship between errors in financial expectations and unhealthy behaviours. I employ a wide range of estimation methods including fixed effects linear probability models, conditional fixed effects Logit models, fixed effects quantile regression models, fixed effects Negative Binomial models, and the generalized method of moments (GMM). I find that extreme optimism (i.e. overestimating future household finances) is associated with higher probabilities of smoking and drinking, and higher cigarette and alcohol consumption. Consistent with the Tension Reduction Theory, I find that psychological ill-being is the channel through which extreme optimism is associated with unhealthy behaviours. Finally,I find that errors in financial expectations tend to affect drinking for males, and smoking for females, which indicates Russian males and females have different coping strategies in response to stress. The results are robust to using both static and dynamic models.

In the third study, using the 2010 and 2017 waves of the South Africa National Income Dynamics Survey, I examine the relationship between errors in financial expectations and individuals' borrowing decisions. I find robust evidence that individuals with optimistic forecast errors show a higher probability of taking financial, personal and short-term debt. They also show higher levels of debt and a higher probability of being over-indebted. Furthermore, I document positive moderating effects of low education and debt illiteracy in the relationship between optimistic forecast errors and borrowing. This suggests that optimists may take more debt due to a lack of understanding of debt contracts and mis-judgement of their debt positions. My results are robust to using a wide range of estimation methods including fixed-effects linear models and correlated random effects Probit and Tobit models. They are also robust to using different measures of forecast errors.

Declaration

I hereby declare that this thesis is my own original work except where stated otherwise by reference in the text. It has not been submitted, in whole or in part, for any other degree or qualification at this or any other university. Further, I have acknowledged and correctly referenced the work of others.

Copyright © 2020 by Wei Jin

The copyright of this thesis rests with the author. No quotations from it should be published without the author's prior written consent and information derived from it should be acknowledged.

This thesis is dedicated to my best friend Dos Zhu.

Without him,

I will not be here.

"*The very beginning mind itself is*

the most accomplished mind of true enlightenment"

勿忘初心，方得始终

Acknowledgements

First and foremost, I would like to express my greatest gratitude to my primary supervisor Professor Alessandra Guariglia (also known as Prof. Panda) for her professional guidance, valuable feedback and genuine support throughout my PhD study in Birmingham. She is not only a caring and loving proud mama panda, but also a challenging and passionate co-worker in the field. Over the past four years, she has provided me with sufficient funding and mental supports in every corner, nurtured my intuition and sensitivity towards what represents new and cutting-edge happening in the field. She has been my greatest inspiration and admiration and I hope I can share her everlasting passion and determination towards socially and economically devoted research.

Furthermore, I would like to thank my supervisor Mr Nicholas Horsewood for his genuine help and insightful comments on my work. I have learned many econometrics methods and skills from him, which contributes greatly to my work. I also appreciate that he is here for us whenever we need help.

I am extremely grateful for Professor Anindya Banerjee and the Department of Economics for the scholarship that supported my study in the department. And I would like to thank Dr Kamilya Sulemenova, Dr Rebecca McDonald and Dr Frank Strobel for giving the opportunities to work and encourage me to be a better teacher.

I would like to thank my seniors Dr Danying Li, Dr Yanan Zhang, Dr Tian Lin, Dr Fu Lin and Dr Weifeng Zhou for sharing their valuable experiences and excellent advice with me and for helping and encouraging me in this journey. I would like to express my special thanks to my wonderful colleagues Yaopeng Wang, Pei Yu, Gianluca Bortoletto, Chenyang Wang and others in our Panda Team for providing valuable feedback on my work and sharing my joy and sorrow in this journey.

At last, my deepest gratitude goes to my father and my mother and my family. You are my rock and my home. It will not be possible for me to complete anything in my life without my family's support and love. I love you.

Wei Jin

October 2020

Table of Contents

Abstract..... I

Declaration..... II

Acknowledgements..... V

Table of Contents..... VI

List of Tables..... X

List of Abbreviations..... XIII

Chapter One: Introduction..... 1

1.1 Research background..... 1

1.2 Motivation and research question..... 3

1.3 Thesis outline..... 5

1.4 Contributions to the literature..... 7

Chapter Two: Longitudinal Evidence on Errors in Financial Expectations: The role of past income changes and job insecurity..... 10

2.1 Introduction..... 10

2.2 Literature Review..... 13

2.2.1 The determinants of financial expectations..... 13

2.2.2 The determinants of errors in financial expectations..... 17

2.2.3 Contribution..... 23

2.3 Hypotheses..... 25

2.3.1 Over-extrapolation hypotheses..... 25

2.3.2 Job insecurity hypotheses..... 26

2.4 Data and summary statistics..... 28

2.4.1 Data description..... 28

2.4.2 Measurements of key variables..... 29

2.4.3 Summary statistics..... 33

2.5 Model specifications and estimation strategy....37
2.5.1 The role of past income changes....37
2.5.2 The role of job insecurity....40
2.5.3 The roles of past income changes and job insecurity....40
2.5.4 Estimation strategy....42
2.6 Empirical results....47
2.6.1 Empirical results - the role of past income changes....47
2.6.2 Empirical results - the role of job insecurity....53
2.6.3 Empirical results - the roles of past income changes and job insecurity....56
2.7 Robustness checks....59
2.7.1 Correlated random effects - ordered Probit model....59
2.7.2 Attrition in the sample....63
2.8 Conclusion....66
Chapter Three: Errors in financial expectations and unhealthy behaviours - Evidence from Russia....115
3.1 Introduction....115
3.2 Literature Review....120
3.2.1 Definitions of optimism and pessimism....120
3.2.2 Optimism/pessimism and unhealthy behaviours....122
3.2.3 Mechanisms underlying the link between optimism/pessimism and unhealthy behaviours....126
3.2.4 Contribution....129
3.3 Hypotheses....131
3.3.1 General hypotheses....131
3.3.2 Psychological ill-being channel hypotheses....133
3.4 Data and descriptive statistics....134
3.4.1 Data....134

3.4.2 Measures of errors in financial expectations and unhealthy behaviours..... 135
3.4.3 Descriptive statistics..... 138
3.5. Baseline specification and estimation strategy..... 142
3.5.1 Main empirical models..... 142
3.5.2 Interaction models..... 145
3.5.3 Estimation strategy..... 147
3.6 Empirical Results..... 152
3.6.1 Main results..... 152
3.6.2 Interaction results..... 166
3.6.3 Sample attrition..... 174
3.7 Conclusion..... 176
Chapter Four: How does errors in financial expectations affect individuals' borrowing decisions? - Evidence from South Africa..... 232
4.1 Introduction..... 232
4.2 Literature Review..... 236
4.2.1 Theoretical background..... 236
4.2.2 Optimism/pessimism and borrowing..... 239
4.2.3 Contribution..... 246
4.3 Hypotheses..... 248
4.3.1 General hypotheses..... 248
4.3.2 Low education hypotheses..... 249
4.3.3 Debt illiteracy hypotheses..... 250
4.4 Data and summary statistics..... 252
4.4.1 Data..... 252
4.4.2 Measurements of dependent and independent variables..... 254
4.4.3 Summary statistics..... 259
4.5 Baseline specification and methodology 262

4.5.1 Main specification..... 262
4.5.2 Moderating effects of low education and debt illiteracy..... 265
4.5.3 Estimation methodology.....266
4.6 Empirical results.....268
4.6.1 Main results..... 268
4.6.2 Moderating results.....274
4.7 Robustness checks.....279
4.7.1 Alternative optimism biases indicators.....279
4.7.2 Correlated random effects.....283
4.8 Conclusion.....288
Chapter Five: Conclusion..... 331
5.1 Summary of findings and policy implication..... 331
5.2 Future research.....333
Reference..... 335

List of Tables

Chapter Two:

Table 2.1 Structure of the unbalanced panel..... 69

Table 2.2 Distribution and classification of expectation errors..... 70

Table 2.3 Distribution of past income changes..... 71

Table 2.4 Expectation errors conditional on past income changes..... 72

Table 2.5 Distribution of job insecurity..... 73

Table 2.6 Summary statistics..... 74

Table 2.7 Sample means of errors in financial expectations by past income changes and job insecurity..... 77

Table 2.8 Correlation Matrix..... 78

Table 2.9 FE and FEIV results of past income changes..... 80

Table 2.10 FE and FEIV results of job insecurity..... 83

Table 2.11 FE and FEIV results of past income changes and job insecurity..... 86

Table 2.12 CRE ordered Probit results..... 89

Appendix 2.1 First stage results of past income changes..... 91

Appendix 2.2 First stage results of job insecurity..... 95

Appendix 2.3 First stage results of past income changes and job insecurity..... 99

Appendix 2.4 CRE ordered Probit results (other controls)..... 104

Appendix 2.5 CRE ordered Probit results (other controls)..... 107

Appendix 2.6 Variable addition tests 1 (num_wave)..... 109

Appendix 2.7 Variable addition tests 2 (next_wave)..... 112

Chapter Three

Table 3.1 Structure of the panel....180

Table 3.2 Summary of dependent variables.... 181

Table 3.3 Distribution of errors in financial expectations.... 182

Table 3.4 Descriptive statistics and differences between male and female samples....183

Table 3.5 Descriptive statistics by errors in financial expectations in the full sample and sub-samples.... 185

Table 3.6 Correlation Matrix between errors in financial expectations indicators....186

Table 3.7 LPM and Logit models for binge drinking....187

Table 3.8 LPM and Logit models for problem drinking.... 190

Table 3.9 GMM models for drinking.... 193

Table 3.10 Quantile regressions for alcohol consumption in the male sample.... 196

Table 3.11 LR and GMM models for alcohol consumption....198

Table 3.12 LPM and Logit models for smoking.... 201

Table 3.13 GMM models by smoking....203

Table 3.14 LR and Negative Binomial models by cigarette consumption....206

Table 3.15 GMM models by cigarette consumption.... 208

Table 3.16 Interaction results for binge and problem drinkings....211

Table 3.17 Interaction results for alcohol consumption.... 213

Table 3.18 Interaction results for smoking....215

Table 3.19 Interaction results for cigarette consumption.... 217

Table 3.20 Variable additional tests in the full sample (*num_wave* and *next_wave*)....219

Table 3.21 Variable additional tests in the male sample (*num_wave* and *next_wave*)....221

Table 3.22 Variable additional tests in the female sample (*num_wave* and *next_wave*)....223

Appendix 3.1 Definitions of Variables....225

Appendix 3.2 Quantile regressions for alcohol consumption in the full sample....228

Appendix 3.3 Quantile regressions for alcohol consumption in the female sample.... 230

Chapter Four:

Table 4.1 Structure of the unbalanced panel..... 291
Table 4.2 Errors in financial expectations categories.....291
Table 4.3 Description of variables.....292
Table 4.4 Summary statistics.....297
Table 4.5 Comparison between Optimism and Pessimism groups..... 299
Table 4.6 Correlation matrix.....301
Table 4.7 FE results for the probability and amount of borrowing..... 303
Table 4.8 FE results for the probability of over-indebtedness.....306
Table 4.9 Moderating results for low education.....308
Table 4.10 Moderating results for debt illiteracy..... 311
Table 4.11 Summary statistics and distribution for optimistic forecast errors.....313
Table 4.12 Results for alternative optimism measures.....315
Table 4.13 Results for correlated random effects.....318
Appendix 4.1 Borrowing decisions by 3-step optimists at time t-2..... 322
Appendix 4.2 Quantile regressions for debt-to-income ratio..... 323
Appendix 4.2a FE Quantile Regressions results for the debt-to-income ratio..... 324
Appendix 4.3 Coefficients of other control variables - low education models..... 327
Appendix 4.4 Coefficients of other control variables - debt illiteracy models.....329

List of Abbreviations

BCEA	Basic Conditions of Employment Act
BHPS	British Household Panel Survey
CDF	Cumulative Distribution Function
CPI	Consumer Price Index
CRE	Correlated Random Effects
DAT	Decision Affect Theory
FE	Fixed-effects
FEI	Financial Expectations Index
FEIV	Fixed-effects Instrumental Variable
GMM	Generalized Method of Moments
IDS	Income Distribution Statistics
IQR	Interquartile Range
IRR	Incidence-Rate Ratio
LAD	Least-Absolute-Deviations
LCH	Life-Cycle Hypothesis
LDS	Life Dissatisfaction
LOT-R	Revised Version of the Life Orientation Test
LPM	Linear Probability Model
LR	Linear Regression
LSMS	Living Standard Measurement Study
MED	Median
NB	Negative Binomial
NIDS	National Income Dynamics Survey
NLSY97	National Longitudinal Study of Youth 1997
NPM	Netspar Pension Monitor

OR	Odds-Ratio
PSU	Primary Sampling Unit
RLMS	Russian Longitudinal Monitoring Survey
Rosstat	Federal State Statistics Service
RUB	Rouble
SA	South Africa
SCF	Survey of Consumer Finances
SEE	Survey of Economic Expectations
SEP	Socio-Economic Panel
SSA	Statistics South Africa
TRT	Tension Reduction Theory
US	United States
WHO	World Health Organization

Chapter One: Introduction

1.1 Research background

Individuals' financial expectations play a central role in determining economic behaviours (Puri and Robinson, 2007). Theoretical evidence like the Life-Cycle Theory (LCH) (Modigliani and Brumberg, 1954) embody the assumption that individuals are forward-looking decision makers. The LCH suggests that individuals' decisions on consumption and savings are not only determined by current variables (present income and assets), but also on subjective expectations of future income, prices and so on. Subject to their current income and discounted future income and resources, individuals smooth lifetime consumption and borrow when facing income volatility during the life-cycle. Empirical evidence has generally corroborated these predictions. Individuals who anticipate higher income over the next year are less likely to have precautionary savings (Harris *et al.*, 2002), and more likely to consume (Brown and Taylor, 2006) and borrow (Lee, 2008).

The role of financial expectations themselves has drawn most of the attention in the economics literature, and less attention has been paid to the investigation of the consequences of making inaccurate financial expectations. The adverse effects of inaccurate financial expectations on individuals' economic behaviours have been documented in some papers (Hyytinen and Putkuri, 2018). For instance, Grohmann *et al.* (2019) suggest that financially optimistic respondents are more likely to smooth their current consumption by saving less and borrowing more at time *t-1*. However, when the expectations are not realised at time *t*, over-optimistic

respondents face lower income than anticipated and have to adjust their consumption to correct the erroneous financial expectations. In some cases, overconfident individuals have no precautionary savings and are unable to pay off their debts completely and keep on borrowing.

A variety of psychology and economics papers have also investigated the negative consequences of errors in financial expectations on individuals' mental health (anxiety, depression) and physical health (abuse of unhealthy behaviours; Shepperd *et al.*, 2017). For example, Armor and Taylor (1998) show that when individuals' expectations fall short, those who display over-optimism may experience emotional let-down or even anxiety and depressive syndromes. Furthermore, according to the Tension-Reduction Theory, individuals may use alcohol (Conger, 1956) and cigarettes (Ikard *et al.*, 1969) as self-medication to regulate negative emotions and psychological distress. Thus, over-optimistic respondents may regulate their negative emotions and moods by consuming more alcohol and cigarettes.

The relationships between errors in financial expectations and individuals' economic behaviours (Hyytinen and Putkuri, 2018; Massenot and Pettinicchi, 2019) and health behaviours (Floyd *et al.*, 2000; Dillard *et al.*, 2009) have been well-documented in the developed countries, while few have examined these relationships in the developing context (Duval and Wolff, 2012).

1.2 Motivation and research question

This thesis is inspired by the following phenomena[1]: First, individuals who make optimistic forecast errors have been found to display imprudent financial behaviours, such as saving less (Cocco *et al.*, 2019), borrowing more and having a higher debt-to-income ratios (Hyytinen and Putkuri, 2018). Given that people living in developing/transitional countries have lower standards of living and greater financial stress, inaccurate forecast errors may have devastating consequences (e.g., concerns over getting basic necessities). It is therefore central and of enormous interest for economists and policy makers to understand what drives individuals to make forecast errors in developing economies. An investigation of the determinants of expectation errors can in fact provide important information that can help prevent individuals from making further forecast errors.

Second, one of the most astonishing developments in world health in the early 21th century has been the mortality crisis in Russia (Cockerham *et al.*, 2006). Stuckler and Basu (2013) suggest that Russia had a long history of drinking, and the high prevalence of smoking and drinking have been major factors contributing to the Russian mortality crisis in the past decades. For instance, Zaridze *et al.* (2014) conduct a retrospective study in three typical Russian cities and find that 47% of the males and 11% of the females were reported to have drunk at least a bottle of vodka a week and many of them were reported to have drunk on average a bottle of

1 The background and motivation of this thesis are only briefly introduced in this chapter. More details on the background and motivation can be found in each of the empirical chapters that follow. This is to avoid repetition.

vodka per day. Thus, policy makers in Russia have launched various anti-drinking (from the mid 2000s) and anti-smoking campaigns (since 2013) to address these unhealthy behaviours.

Third, in the past decades, South Africa, as an emerging market, has witnessed a rapid growth of consumer debt and credit markets. A wide range of new credit products have been developed and marketed to specific consumers from banks and other financial retailers. Consumers nowadays are getting used to employ debt as a way to redistribute future finances and smooth their current consumption. Statistics shows that the ratio of household debt to disposable income stands at 73% and has increased by over 40% since 1990 (South African Reserve Bank, 2017). Moreover, the interest rates has dropped significantly from 14% to 8% in the early 2000s. The decrease in interest rates reduces debt repayments and further fuels the growth of consumer debt in South Africa. With escalating debt, consumers may run the risk of accumulating an overwhelming level of debt, which could lead to a drop in their future consumption. Thus, the over-indebtedness of individuals is gradually becoming a concern for policy makers in South Africa (Meniago *et al.*, 2013).

Informed by the aforementioned issues, this thesis presents three related empirical studies on errors in financial expectations in developing/transitional countries (Russia and South Africa[2]). More specifically, the first study investigates the role of past income changes and job insecurity in determining the errors in financial expectations in Russia. The second study focuses on two research questions: (1) the relationship between errors in financial expectations and health behaviours (smoking and drinking) in Russia; (2) whether psychological ill-being is the channel

2 According to the United Nation (2019), Russia and South Africa are both classified as developing/transitional economies, not developed economies.

through which errors in financial expectations may affect unhealthy behaviours. The third study examines: (1) whether errors in financial expectations affect individuals' borrowing behaviour in South Africa; (2) the moderating effects of low education and debt illiteracy on the relationship between forecast errors and borrowing. Micro-level survey data are employed in all three studies.

1.3 Thesis outline

Using the 2002-2016 waves of the Russian Longitudinal Monitoring Survey (RLMS), in Chapter two, I empirically examine the determinants of errors in financial expectations in Russia, with particular focus on the role of past income changes and job insecurity. In order to obtain robust results, I use a fixed-effects instrumental variable (FEIV) estimator to control for unobserved individual heterogeneity and endogeneity, as well as different proxies of expectation errors. Focusing on the role of past income changes, I find that individuals whose financial situation has improved over the last year are more likely to overestimate their future income growth. This could be explained by the over-extrapolation of past experience, which suggests that individuals tend to build their expectations based on past experience (Massenot and Pettinicchi, 2019). Focusing on the role of job insecurity, I find that the more secure individuals are at finding a new job, the larger the forecast errors they make.

In Chapter Three, to provide a new perspective for Russia's intensive anti-drinking and anti-smoking campaigns, I empirically examine whether and how individuals' forecast errors affect smoking and drinking decisions. A novel approach is adopted by constructing four indicators of forecast errors: extreme optimism, moderate optimism, extreme pessimism and moderate pessimism (Puri and Robinson, 2007). Given that Russian men are infamous for their binge

drinking (Yakovlev, 2018), I first estimate my models over the full sample, and then estimate the models for male and female samples separately. Using the 2001-2016 waves of RLMS with over 120,000 observations, I find that extreme optimists show a higher probability of binge drinking in the male sample, while extreme optimists show a higher probability of smoking in the female sample.

Furthermore, I use life dissatisfaction as an indicator for psychological ill-being to examine whether psychological ill-being is the channel through which forecast errors affect smoking and drinking. The results show that only extreme optimists with psychological ill-being display a higher probability of binge drinking in the male sample and of smoking in the female sample. In line with the Decision Affect Theory (Shepperd and McNulty, 2002) and the Tension Reduction Theory (Conger, 1956; Ikard *et al.*, 1969), this finding suggests that extreme optimists who suffer from psychological ill-being consequently consume higher levels of alcohol and cigarettes as an effort to relieve and regulate negative emotions. As for the differential effects of extreme optimism on smoking and drinking between the male and female samples, Cockerham *et al.* (2006) explain that Russian females translate their psychological stress into smoking instead of drinking due to social norms and family obligations. Thus, male optimists are found to consume more alcohol, while female optimists show higher cigarette consumption. The findings of this Chapter suggest that providing extreme optimists with financial education programmes aimed at avoiding making inaccurate forecast errors may have important implication on reducing stress-related smoking and drinking in Russia.

In Chapter Four, using the 2010 and 2017 waves of the National Income Dynamics Survey (NIDS), I empirically test whether errors in financial expectations affect individuals' borrowing

decisions in South Africa (SA). Three proxies for the borrowing decisions are used: the probability of borrowing, the amount of borrowing and the probability of being over-indebted. I find evidence that individuals who make optimistic forecast errors display a higher probability of borrowing and being over-indebted, while those who make pessimistic forecast errors do not show differences in their borrowing decisions. To better understand why optimists display higher levels of borrowing, I examine the moderating role of low education and debt illiteracy in the relationship between optimistic forecast errors and borrowing. I find that low educated or debt illiterate optimists show a higher probability of borrowing and being over-indebted. This suggests that optimists are more likely to engage in borrowing if they have a lower education and/or are debt illiterate. Thus, introducing financial education workshops focused on increasing individuals' debt literacy could help policy makers in SA reduce indebtedness (Meniago *et al.*, 2013).

1.4 Contributions to the literature

This thesis contributes to the existing literature in the following ways. In Chapter Two, the contributions are threefold. Firstly, to the best of my knowledge, no other paper has investigated the determinants of errors in financial expectations in Russia. Next, this study takes into account, for the first time, the role of both past income changes and job insecurity on the determinants of expectation errors. Third, most studies on this area, such as Das and van Soest (1997, 1999) and Massenot and Pettinicchi (2019), have only used cross-sectional or short panel data and failed to account for endogeneity. This study adds to the literature by employing a FEIV estimator to obtain unbiased estimates.

Chapter Three provides the following four contributions to the literature. First, I examine for the first time the relationship between errors in financial expectations and health behaviours in Russia. Second, to the best of my knowledge, no study has examined this relationship using a large panel dataset. Third, this study is the first to investigate possible channel through which errors in financial expectations affect unhealthy behaviours in Russia. Fourth, this study contributes to the literature by applying a dynamic model to control for the habit persistence of smoking and drinking.

Chapter Four mainly fills the following three gaps in the literature. First, to the best of my knowledge, no other study has investigated the effect of errors in financial expectations on individuals' borrowing decisions in South Africa. Next, the majority of the literature on the relationship between expectation errors and borrowing has used either cross-sectional data (Del Rio and Young, 2006) or random-effects models (Brown *et al.*, 2008). This study makes use of a fixed-effects estimator to eliminate the unobserved individual fixed traits that are prevalent among optimism literature (Hyytinen and Putkuri, 2018). Third, to the best of my knowledge, no existing study has analysed whether low education and debt illiteracy moderate the relationship between expectation errors and borrowing in SA.

The remainder of this thesis is organised as follows. Chapter Two, Chapter Three and Chapter Four present the three above-mentioned empirical studies. Chapter Five concludes the thesis.

Chapter Two: Longitudinal Evidence on Errors in Financial Expectations: The role of past income changes and job insecurity

2.1 Introduction

In recent years, errors in individuals' financial expectations have gradually attracted more attention from researchers. Errors in financial expectations are typically calculated by taking the difference between individuals' financial expectations at time *t-1* and financial realisations at time *t*. Individuals are classified as over-optimistic if the actual realisation fails to match the expectation, and as over-pessimistic if the realisation exceeds the expectation[3]. The relationship between expectation errors and individuals' economic behaviours has been examined in the developed countries. Over-optimistic individuals have been found to have shorter planning horizons and display imprudent financial habits (Puri and Robinson, 2007), such as saving less (Cocco *et al.*, 2019), consuming more (Massenot and Pettinicchi, 2019) and borrowing (Hyytinen and Putkuri, 2018).

Considering that errors in financial expectations play an important role in predicting individuals' economic behaviours, it is essential for economists and policy makers to understand what drives individuals to make optimistic or pessimistic forecast errors. Empirical studies on the determinants of expectation errors can in fact provide crucial information that helps understand individuals' economic habits. Such information also enables policy makers to assess the impacts of their economic policies on consumption and savings. For example, respondents who make

3 In this chapter, I define individuals with inaccurate financial expectations as being either over optimistic or pessimistic. I use 'errors in financial expectations', 'optimistic and pessimistic forecast errors' and 'optimistic and pessimistic expectation errors' interchangeably.

optimistic forecast errors may be less responsive to saving stimulation policies (e.g., rises in the interest rate). Hence, policy makers may be interested in identifying which particular groups of individuals are more prone to make optimistic forecast errors.

There are a handful of studies focused on the determinants of expectation errors in developed countries. For instance, using cross-sectional Dutch data, Das and van Soest (1997, 1999) suggest that past income changes are the main determinant of expectation errors and find that individuals who experience a positive income change over the year are more likely to make optimistic forecast errors. Massenot and Pettinicchi (2019) note that individuals will over-extrapolate their past experience and expect any positive (negative) streak to be continued. As a result, they are more likely to overestimate their future finances after experiencing a positive income change and to underestimate them after a negative experience.

Surprisingly, empirical evidence on the determinants of errors in financial expectations is scarce in developing countries due to limited data availability on individuals' financial expectations (Duval and Wolff, 2012). As individuals living in developing countries face lower standards of living and higher financial pressure, inaccurate forecast errors may have devastating consequences (e.g., no precautionary savings, higher debt burdens). Thus, empirical studies on the determinants of errors in financial expectations in developing countries are much warranted.

The objective of this chapter is to provide an in-depth analysis of the determinants of errors in financial expectations in Russia, with particular focus on the role of past income changes and job insecurity. The contributions of this chapter to the empirical literature are three-fold. First, to the best of my knowledge, no other study has examined the determinants of expectation errors in

developing countries. Second, whilst previous studies only take into account the role of past income changes (Cocco *et al.* 2019), this study adds to the literature by accounting for the potential role of job insecurity in determining expectation errors. Third, most of the literature on the determinants of expectation errors has only used cross-sectional (Das and van Soest, 1997, 1999) or short panel data (Massenot and Pettinicchi, 2019). Yet, empirical results derived from cross-sectional data may be largely driven by unobserved heterogeneity and endogeneity, This study improves on this literature by using a fixed-effects instrumental variable (FEIV) estimator to control for potential bias.

The remainder of this paper is organised as follows. Section 2.2 provides a review of the empirical literature on the determinants of errors in financial expectations. Section 2.3 proposes the main hypotheses. Section 2.4 presents the data and summary statistics. Section 2.5 illustrates the baseline specifications and estimation methodology. Section 2.6 reports and discuss the empirical results. Section 2.7 presents some robustness checks and Section 2.8 concludes.

2.2 Literature Review

A growing literature has been exploring the determinants of errors in financial expectations in developed countries (Massenot and Pettinicchi, 2019; Cocco *et al.* 2019). However, it draws less attention in developing countries due to the scarcity of individual data measuring subjective financial expectations and realizations (Duval and Wolff, 2012). Errors in financial expectations are measured as the differences between financial expectation and financial realization. Thus, the determinants of financial expectations and errors in financial expectations are closely related and examined. In fact, it is common practice to apply the same set of variables to examine the determinants of financial expectations as well as that of the errors in financial expectations (Das and van Soest, 1997, 1999).

This section proceeds as follows. Section 2.2.1 examines the literature on the determinants of financial expectations. Section 2.2.2 presents the literature on the determinants of errors in financial expectations. Section 2.2.3 summarises the contributions of this chapter.

2.2.1 The determinants of financial expectations

A vast number of predictors determining financial expectations have been heavily researched and documented. Among the numerous predictors investigated, there are generally two key determinants established in the literature: past income changes (Das and Donker, 1999; Duval and Wolff, 2012) and job insecurity (Dominitz, 1998; Bissonnette and Soest, 2015).

The first stream of literature focuses on the role of past income changes in predicting financial expectations. Das and Donker (1999) were one of the pioneers to use subjective

questions to explore the determinants of income expectations. Their study uses the 1993 Dutch CentER Panel Survey with a sample of 1,122 household heads. Household' income expectations are elicited from their subjective expected distribution of future income. Specifically, household heads are first asked to provide the expected highest and lowest possible weekly income one year from now. They are then asked to evaluate the probability with which their weekly income will fall into the specific categories. Each individual will be assigned within four future income categories (in increasing order), which are calculated by a computer-algorithm using the lowest and highest future income provided by the respondents. Finally, the author derives the subjective cumulative distribution function (CDF) of future income for each household, and uses the median (MED) and interquartile range (IQR) of the CDF as proxies for income expectations.

The authors use the least-absolute-deviations (LAD) estimates and find that household income in the past 12 months is a significant and dominant factor in predicting the expected household income in the next 12 months. For instance, they find that every one dollar increase in past income is associated with a 8.13 dollar increase in expected future income. However, this study is limited by its cross-sectional natural failing to control for unobserved heterogeneity.

Duval and Wolff (2012) utilise a sample of 5,618 observations on the 2002-2004 Living Standard Measurement Study (LSMS) to examine the determinants of individuals' financial expectations in Albania. A subjective financial expectations indicator is formulated, equaling to 1 if respondents expect their financial situation to 'deteriorate a lot', 2 if they expect it to 'deteriorate somewhat', 3 if they expect it 'remaining the same' and 4 if they expect it to 'somewhat or greatly improve'. The authors employ random effects and quasi fixed effect ordered probit models and present robust finding that realised income changes in the past 12

months play a significant role in forming individuals' expectations. This study shows that those who experienced an improvement (deterioration) in their financial situation in the past year are more likely to expect a continuous improvement (deterioration) in their financial situation over the next year, suggesting a strong inertia between individuals' realised income changes and financial expectations.

Brown and Taylor (2006) extend the analysis by taking into account the persistence of individuals' financial expectations. They use the 1991-2003 British Household Panel Survey (BHPS) which contains 118,664 observations to examine whether past expectations errors affect individuals' financial expectations. The authors first construct a financial expectations index (FEI), which is coded 0 if respondents expect their financial situation to be worse off in the next 12 months, 1 if they expect no changes, 2 if they expect to be financially better off. Next, they compare individuals' expectations at time *t-1* regarding whether they will be financially worse-off, the same or better-off in the next 12 months to individuals' response at time *t* regarding whether their financial situation actually deteriorated, remained the same or improved in the past 12 months.

Four binary variables are formulated denoting incorrect/correct optimistic/pessimistic expectations errors. The authors employ a random effects ordered probit model and find that individuals making incorrect or correct optimistic expectations errors both show positive financial expectations. Meanwhile, an incorrect or correct pessimistic financial prediction is associated with a negative financial expectation. The study also shows that the effects of optimistic expectation errors on individuals' expectations outweigh those of pessimistic expectation errors. Their findings are robust to controlling for past income changes and suggest

that there may be habit persistence in the process of individuals' expectation formation. Thus, the use of a dynamic model is of particular importance.

The second stream of literature has drawn attention to the role of job insecurity in determining financial expectations. Dominitz (1998) examines the role of job loss in determining individuals' financial expectations. He use the 1993 US Survey of Economic Expectations (SEE), which is a national telephone survey with a representative sample of 622 respondents. His study uses financial expectation data, as well as the same approach as in Das and Donker (1999), to fit an individual-specific subjective distribution for future income. He employs the least-absolute-deviations (LAD) estimates and finds that respondents with higher probabilities of job loss show lower income expectations. The study also shows that those with higher probabilities of job loss face higher income uncertainty.

In a more recent study, Bissonnette and Soest (2015) use a sample of 2,519 observations from the 2009-2012 Dutch Netspar Pension Monitor (NPM) to investigate the determinants of financial expectations, with particular focus on the role of job insecurity. Financial expectations are measured by two subjective questions on the probability that the purchasing power of respondents' future income will be larger/lower than it is now in the next 12 months. The authors employ an ordinal variable (ranging from 1 to 10) to measure individuals' job insecurity. *Crisis_job* equals to 1 (10) if an individual is not at all (completely) confident at keeping the current job in the next 12 months. The authors then use tobit models and find that despite the effect of the crisis, individuals who are (not) confident at keeping their jobs show positive (negative) income expectations. Thus, this study documents a positive correlation between individuals' job insecurity and financial expectations.

In addition to the two key determinants surveyed above, a number of other significant determinants of individuals' financial expectations have also been documented in the empirical literature. Consistent with the life-cycle theory (Modigliani and Brumberg, 1954), an U-shaped association between age and financial expectations has been established, suggesting that younger (older) individuals are more likely to be financially optimistic (pessimistic) (Das and Donker, 1999). Socioeconomic variables have also been found to be significant determinants. Being male, married, with a larger household and a higher education are all associated with optimistic financial expectations (Bissonnette and Soest, 2015). In terms of income and wealth, higher income, more savings and wealth are associated with having optimistic financial expectations (Dominitz, 1998; Duval and Wolff, 2012). Turning to employment status, individuals in labour market (employed or self-employed) are more prone to be financially optimistic (Brown and Taylor, 2006).

2.2.2 The determinants of errors in financial expectations

The aforementioned literature explicitly investigates the determinants of financial expectations, and only a handful of studies have examined whether these financial expectations are realised over the next year. Empirical studies on the determinants of errors in financial expectations primarily focus on the over-extrapolation of past income changes (Das and van Soest, 1997, 1999; Massenot and Pettinicchi, 2019).

Das and van Soest (1997) conduct a cross-sectional analysis to explore the determinants of errors in financial expectations, using data on 2,179 households from the 1984-85 waves of the Dutch Socio-Economic Panel (SEP). An income expectation variable (INCEXP) is constructed

based on individuals' response to the question "what will happen to your household's income in the next 12 months". INCEXP is an ordinal variable measured on a 5-point scale, running from "strong decrease" to "strong increase". With respect to the individuals' income realisations, respondents are asked to answer the following question "did your household's income change during the previous 12 months". An ordinal variable (PREV) is then formulated, equaling to 1 if household income strong decreased, 2 if household income decreased, 3 if household income remained the same, 4 if household income increased and 5 if household income strong increased.

By utilising data on two waves, the authors subsequently compare individuals' expectation (INCEXP84) in 1984 with the realisation (PREV85) in 1985 and derive a variable DEV as the difference between PREV85 and INCEXP84. DEV, ranges from -3 to 4, is larger than 0 if a household underestimated their future income growth, is less than 0 if the household overestimated it, and 0 if the household correctly predicted it. Interestingly, they find that over a third (34.9%) of the households underestimated their future income growth, while only 15.5% of the households overestimated their income growth. This suggests that people, on average, significantly underestimate their future income growth. To examine the determinants of errors in financial expectations, the authors employ an ordered probit model and find that the lagged income relisation (PREV84) is the most important predictor of DEV. The study shows that compared to households who experienced no income change over the past year, those whose income has decreased or strongly decreased are more likely to underestimate their future income growth, while those whose income has increased are more likely to overestimate future income growth. However, this study fails to take into account the unobserved household heterogeneity.

For instance, households could be simply too pessimistic in general and the effect of decreases in past income on expectation errors could be exaggerated.

Using an unbalanced panel of 6,845 households from the 1984-1989 SEP, Das and van Soest (1999) aim to answer two questions: whether the determinants of errors in financial expectations are robust over time and of the same expected signs as in Das and van Soest (1997). The authors construct a DEV as in Das and van Soest (1997), and convert the ordinal variable into two binary variables denoting optimistic (overestimating) and pessimistic (underestimating) expectations errors respectively. Controlling for unobserved household heterogeneity and macro-economic shocks, they employ a conditional fixed-effects logit model and find that household whose income has decreased (increased) in the past show a higher probability of underestimating (overestimating) future income growth. This result suggests that the findings of the role of past income changes in determining expectation errors are robust using a fixed-effects estimator as well as controlling for year fixed effects (macro-economic shocks).

Massenot and Pettinicchi (2019) use the 2009-2016 Tilburg Consumer Outlook Monitor in the Netherlands with over 31,442 observations to examine the effect of over-extrapolation of past income changes on expectation errors. They make use of the following two questions of interests: (1) whether the financial situation of your household has changed in the last 12 months; (2) do you think the financial situation of your household will change in the next 12 months. There are five responses to the above questions: clearly gotten better (2), gotten a bit better (1), stayed the same (0), gotten a bot worse (-1), clearly gotten worse (-2). An expectation error is then computed by taking the difference between the expected financial situation last year and the realised financial situation this year. Households are classified as overoptimistic if the error is

positive, and overpessimistic if it is negative. The study uses a fixed-effects estimator and finds that households who experienced improved (worsen) financial situations in the previous year are more likely to be overoptimistic (overpessimistic). The authors suggest that households who experienced positive (negative) income changes in the previous year may exaggerate the persistence of the positive (negative) shocks, which leads them to make overoptimistic (overpessimistic) errors. This is called over-extrapolation of past income changes (past experience).

The authors further extend their analysis by examining the persistence and asymmetry of over-extrapolation. They add a second lag of past income change into the regressions and find that both the first and second lags of past income change are positively associated with expectation errors. This implies two things: First, there is a persistent effect of the past income change on expectation errors. Households who experienced positive (negative) income changes in the previous two years are more likely to make overoptimistic (overpessimistic) errors. Secondly, since the first and second lags of past income change are both positive and significant determinants of expectation errors, this suggests that household extrapolate more strongly if they experienced a streak of similar experience. For instance, households who experienced two positive income shocks on a streak are more likely to make overoptimistic errors than household who experienced one positive and one negative income shocks in the last 2 years.

To examine the asymmetry of past income changes, the authors construct four binary variables that represents the four possible directions of past income change: a lot better, a bit better, a bit worse, a lot worse. These binary variables equal to 0 if a household answered “stay the same”. They find that households with negative income changes (a bit/lot worse) are more

likely to make overpessimistic errors, while households with positive income changes are more likely to make overoptimistic errors. Albeit the sign differences, the effects of negative shocks (a bit/lot worse) on expectation errors are significantly larger than the effect of positive shocks in absolute value. This implies that households react asymmetrically differently to positive and negative income shocks, and tend to make larger expectation errors after experienced a negative shock than a positive shock.

The prior-mentioned literature focuses on the role of past income changes in determining errors in financial expectations at the household level. Instead, Cocco *et al.* (2019) use the British Household Panel Survey (BHPS) over the period 1991-2008 with a sample of 79,204 observations to explore the relationship between changes in financial situation and errors in financial determinants at the individual level. In the questionnaires, each individual is asked: 1) whether he/she is financially better, worse-off, or about the same as last year; 2) whether his/her financial situation would be better, worse, or the same in the next year. Two binary indicators for expectation errors are computed: overoptimism and overpessimism. Individuals are classified as being overoptimistic (overpessimistic) if they expect to be better (worse) off at time t than the actual realised financial changes at time $t+1$. The authors use conditional fixed-effects models and find that individuals who have experienced an improvement in their financial situation are indeed more likely to over-extrapolate the positive income changes, and more (less) likely to be overoptimistic (overpessimistic). This is consistent with the over-extrapolation of positive income changes found in Massenot and Pettinicchi (2019).

As for other determinants of errors in financial expectations, findings in the empirical literature are fairly consistent. First, a U-shaped relationship has been detected between age and

expectation errors. In line with the life-cycle theory, young individuals at the beginning of their career and more likely to experience large income increases (Das and van Soest, 1997). Compared to their older counterparts, young people are more impatient and thus more prone to anticipate high income and make optimistic forecast errors. On the contrary, older individuals are more likely to lower their expectations and employ prudent financial behaviours. As a result, they are more likely to show pessimistic forecast errors (Puri and Robinson, 2007).

Next, Massenot and Pettinicchi (2019) find that men are generally more prone to make optimistic forecast errors than women. This could be explained by the traditional findings that women are more risk-averse than men (Jacobsen *et al.*, 2014). Based on their risk-averse nature, women are less likely to over-estimate their future financial situation, and more prone to make pessimistic forecast errors, while men are the opposite. Third, Das and van Soest (1997) find that being married or living with a partner is positively associated with over-optimism. According to marriage theories (Khodarahimi, 2015), married households tend to have the same desired goals as a couple, and they are more optimistic fighting towards the desired future due to the characteristics of a marriage, such as shared responsibility and commitment, supportive resources and so on. As a result, they are more likely to be over-optimistic regarding future finance than singles and widowed households.

Fourth, as for education, Duval and Wolff (2012) suggest that education is closely related to individuals' permanent income and wealth. Compared to the less-educated, the well-educated are more likely to have a higher income and get more job opportunities in the future. They are therefore more likely to over-estimate their future financial situation. Fifth, turning to wealth controls (income, consumption and saving), wealthier individuals are better shielded from future

downfalls than the poor. They are more likely to use higher income and saving as a buffer against economic volatility. Given that they have something to fall back on, they are more likely to show imprudent financial expectations, such as optimistic forecast errors (Carroll, 1997).

Sixth, Individuals living in urban area face more job and investment opportunities, and thus more likely to overestimate their future income growth (Duval and Wolff, 2012). Seventh, in terms of occupation, compared to those who are not in the labor force, those who are in the labor market (employed, self-employed and unemployed) face less earnings volatility and are more likely to be overoptimistic (Das and van Soest, 1999). Eighth, as for retirees, most of them rely on pensions and have a stable and predictable income source, which suggests that they are less likely to over-estimate their future finance growth (Das and van Soest, 1997). Last, compared to those with poor health condition, individuals with a better health condition are more likely to have a proper financial planning horizon and thus to overestimate their future financial situation (Bissonnette and Soest, 2015).

2.2.3 Contribution

This chapter fills the following three gaps in the literature. First, the determinants of errors in financial expectations have been well-examined in developed countries. However, to the best of my knowledge, no study has examined the determinants of expectation errors in developing countries. Panel data from the RLMS for the years 2002 to 2016 is utilised for the first time to explore the determinants of expectation errors in Russia. Considering that people living in developing/transitional countries face greater financial pressure, individuals with inaccurate forecast errors may suffer from devastating economic consequences (e.g., concerns over getting basic necessities, more debt). Thus, understanding what drives individuals to make forecast errors

can help policy makers target audience more efficiently and improve the precision of policy targeting (Duval and Wolff, 2012).

Next, previous literature on the determinants of expectation errors has mainly focused on the role of past income changes as a determinant of expectation errors (Massenot and Pettinicchi, 2019). However, some studies also suggest that employment insecurity can contribute to the determination of expectation errors (Dominitz, 1998; Bissonnette and Soest, 2015). For the first time, I take into account the role of both past income changes and job insecurity in my empirical analysis.

Third, the majority of the studies in this field have only used cross-sectional or short panel data to examine the determinants of expectation errors (Das and van Soest, 1997, 1999). Massenot and Pettinicchi (2019) suggest that the established empirical findings in this area are subject to concerns over unobserved heterogeneity and endogeneity. Following Xu (2013) and Staudigel (2016), this chapter employs a fixed-effects instrumental variable estimator to address these issues.

2.3 Hypotheses

This chapter examines the determinants of errors in financial expectations with particular focus on the role of past income changes and job insecurity. Specifically, errors in financial expectations are measured by the difference between individuals' financial expectations for next year and the actual realised financial changes in the following year. Four categories of expectation errors are formulated: extreme optimism, moderate optimism, extreme pessimism and moderate pessimism. Past income changes are based on whether individuals' financial situation improved or deteriorated over the past year. The responses are divided into four categories: greatly improved, improved, deteriorated and greatly deteriorated. As for job insecurity, I construct four dummy variables regarding individuals' confidence to find a job if laid off: absolutely/fairly certain, absolutely/fairly uncertain. This section will present in turn the two main hypotheses namely: the over-extrapolation hypotheses, and the job insecurity hypotheses.

2.3.1 Over-extrapolation hypotheses

Literature in the area of expectation errors first examines data at the household level. Das and van Soest (1997, 1999) employ Dutch data from the SEP and find that households' past income changes are the most important predictor of errors in financial expectations. Massenot and Pettinicchi (2019) classify households who overestimate their future income growth as overoptimistic and those underestimating it as overpessimistic. They find that households who experienced a great/slight improvement (deterioration) in their financial situation in the past year are more likely to be overoptimistic (overpessimistic). This suggests that following a great/slight

improvement (deterioration) in their financial situation, households over-extrapolate their latest experience and expect their situation to keep improving (even at the same rate), which leads them to make overoptimistic (overpessimistic) forecast errors. At the individual level, Cocco *et al.* (2019) use data from the BHPS and find similar evidence that individuals are more likely to over-extrapolate their recent improvement (deterioration) in their financial situation and make optimistic (pessimistic) expectation errors.

To the best of my knowledge, no study has examined the relationship between past income changes and errors in financial expectations in developing countries. I believe that a pattern similar to the one observed in developed countries will also apply to Russia. This leads to the following two hypotheses:

H1a: Individuals whose financial situation has greatly/slightly improved are more (less) likely to show extreme/moderate optimism (pessimism).

H1b: Individuals whose financial situation has greatly/slightly deteriorated are more (less) likely to show extreme/moderate pessimism (optimism).

2.3.2 Job insecurity hypotheses

No study has examined the relationship between job insecurity and errors in financial expectations. Existing studies (Dominitz, 1998; Bissonnette and Soest, 2015) mainly focus on the role of job insecurity in determining financial expectations instead of the expectation errors. Dominitz (1998) uses U.S data from the SEE and finds that individuals who have a higher chance of job loss show lower financial expectations. Bissonnette and Soest (2015) use Dutch data from the NPM to examine whether economic crises may alter the negative relationship found between

job insecurity and financial expectation. Controlling for the crisis dummy, they find that individuals who are confident at keeping their jobs are more likely to show optimistic expectations, while individuals who are unconfident at keeping the jobs have a higher chance of showing pessimistic expectations. Thus, a negative relationship between job insecurity and financial expectations is documented.

Furthermore, Das and van Soest (1999) utilise panel data from the SEP to examine whether individuals' expectations in this year are realised over the next year. They regress financial realisations on expectation and find that the coefficient on expectation is significantly positive and lower than 1, which suggests that individuals have the tendency to overestimate their future income growth and make optimistic forecast errors. To sum up, individuals who are confident at keeping their jobs are likely to show optimistic expectations. At the same time, those with optimistic expectations are more likely to make optimistic forecast errors. On the basis of these studies, the job insecurity hypotheses read:

H2a: Individuals who are highly/relatively confident at finding a new job, are more (less) likely to display extreme/moderate optimism (pessimism).

H2b: Individuals who are highly/relatively unconfident at finding a new job, are more (less) likely to display extreme/moderate pessimism (optimism).

2.4 Data and summary statistics

This section begins with a detailed description of the data I used. Subsequently, the measurements of relevant variables are defined, and summary statistics are presented.

2.4.1 Data description

My empirical analysis is based on data from the Russia Longitudinal Monitoring Survey (RLMS), a nationally-representative household-based dataset designed and organized jointly by the Carolina Population Center at the University of North Carolina at Chapel Hill and the Demoscope team in Russia. The RLMS is designed to monitor the effects of Russian reforms on the health and economic welfare of households and individuals in the Russian Federation. Data have been collected annually (with two exceptions - 1997 and 1999) since 1994. 3975 households and 11290 individuals from 38 primary sampling unit (PSU) (containing 95.6% of the population) are surveyed. The survey includes a variety of questions related to individual demographic and socio-economic characteristics, individuals' health behaviours and status, household-level expenditures and service utilization, and relevant community-level data, including community-specific price data. The survey employs a multistage, random cluster sampling approach to get a nationally representative sample for Russia.

My study utilizes unbalanced panel data from round 10 to round 23 of the RLMS over the time span 2002-2016. I do not utilize data on earlier rounds because the questions related to the optimism/pessimism variables were not available.

I have made the following adjustments to the original sample: First, for questions related to this study, respondents who answer "refuse to answer" "do not know" or "answer not given" are

dropped. Second, respondents aged below 18, the legal age of employment in Russia, are dropped. Next, to prevent outliers from affecting the results, I have excluded the top 1% of all the continuous and discrete variables (e.g., income, consumption, household size). In order to eliminate the inflation effects, I have inflated households' income and consumption to the real value of 2016, using the consumer price index (CPI) from the Federal State Statistics Service of Russia.

This chapter analyses an unbalanced panel of data, which consists of 25,837 individuals and 114,707 observations. All individuals in the same household are included in the final sample, not just the head of household. The mean number of observation per individual is 4.4, with the maximum (minimum) being 15 (1). Only 313 individuals participated in all 15 waves.

Table 2.1 presents the round-by-round sample size of the unbalanced panel employed in this chapter. The sample frame was based on dwellings, where interviewers from both urban and rural areas attempted to conduct interviews in the same dwellings since 1994. Considering that there are attrition and drop-outs of the original sample of 1994, the sample was replenished several times. Most notably, larger replenishments took place in 2006 and 2010, with the latter replenishment increasing the sample size from 6,923 to 10,129 observations. Thus, the final sample is consisted of individuals residing at the addresses of 1994 sample, as well as addresses of replenishments (Kozyreva *et al.*, 2016).

2.4.2 Measurements of key variables

2.4.2.1 Errors in financial expectations

In the RLMS, there are two questions related to individuals' financial expectations and realisations, which I refer to as *finexp* and *finreal* for short:

Q(finexp):"Do you think that in the next 12 months you and your family finance will live better than today or worse?"

A(finexp): [1] much worse; [2]somewhat worse; [3]no change; [4]somewhat better; [5]much better.

Q(finreal): "How has the financial situation of you and your family changed in the last 12 months?"

A(finreal): [1]greatly worsened; [2]slightly worsened; [3]no change; [4]slightly improved; [5]greatly improved.

Following Das and van Soest (1997, 1999) and Cocco *et al.* (2019), I employ three types of variables (continuous, dummy and ordinal) to measure individuals' errors in financial expectations.

Error type 1 (continuous variable)

I first compute a continuous measure for the expectation error at time *t* ($error1_{it}$) by taking the difference between individuals' financial expectations at time *t-1* ($finexp_{it-1}$) and financial realisations at time *t* ($finreal_{it}$). The resulting variable ($bias_{it}$) ranges from -4 to 4. If the error score is positive then respondents are overestimating their future income growth, and if it is negative then respondents are underestimating income growth.

However, there is a limitation in regarding the expectation error as a continuous variable. A continuous variable assumes the difference between each spacing is the same (linear).Yet, Puri and Robinson (2007) and Massenot and Pettinicchi (2019) find that there is a distinction between extreme and moderate optimists (pessimists). They indicate that extreme optimists have a much shorter planning horizon and more imprudent financial habits. Thus, it is important to

acknowledge the differences and distinguish between extreme and moderate values of expectation errors.

Extreme/moderate optimism/pessimism (dummy variables)

I construct four binary errors in financial expectations indicators: extreme optimism (*eopt*), moderate optimism (*mopt*), extreme pessimism (*epes*) and moderate pessimism (*mpes*). An individual is classified as being extremely optimistic if the error score equals to 2/3/4; moderately optimistic if the error score equals to 1; extremely pessimistic if the error score equals to -2/-3/-4; moderately pessimistic if the error score equals to -1 (Puri and Robinson, 2007).

Error type 2 (ordinal variable)

Finally, I construct an ordinal variable (*error2*) denoting the five classifications of the expectation error: [1] extreme pessimism, [2] moderate pessimism, [3] correct expectation, [4] moderate optimism and [5] extreme optimism. Table 2.2 presents the distribution and classification of the expectation errors.

2.4.2.2 Past income changes

Turning to past income changes, as in Massenot and Pettinicchi (2019), I construct four dummy variables (*gimproved/ simproved/ gworsened/ sworsened*) in correspondence to the four responses to the *finreal* question. *Gimproved/simproved* is a dummy variable that equals to 1 if an individual has experienced a greatly/slightly improvement in his/her financial situation over the past 12 months, and 0 otherwise. *Gworsened/sworsened* is a dummy variable equaling to 1 if an individual answered his/her financial situation has greatly/slightly worsened in the past 12 months, and 0 otherwise. Table 2.3 presents the distribution of past income changes.

To assess whether individuals over-extrapolate their past income experience, Table 2.4 describes individuals' expectation errors conditional on past income changes in the last 12 months. For those who indicate a greatly improved financial situation in past 12 months, 57.4 % (35.83+21.57) make optimistic forecast errors. Among those denoting a slightly improved situation, 41.41% (31.43+9.98) show either moderate or extreme optimism. These findings suggest that individuals are prone to exaggerate the persistence of past income experience, expecting future income changes will be the same as previous income changes. This phenomenon is referred to as over-extrapolation (Massenot and Pettinicchi, 2019). Similar findings are observed for those who have experienced a great (slight) deterioration in past financial situation: 37.52% (29.25%) over-extrapolate the negative income changes and make pessimistic forecast errors.

2.4.2.3 Job insecurity

Following previous research (Linz and Semykina, 2010; Bissonnette and Soest, 2015), I make use of the question (*findjob*), regarding individuals' confidence in finding a new job in the case of sudden laid-off, as a proxy for self-perceived risk of job insecurity. In the RLMS questionnaires, the respondent was asked:

Q(findjob): "For some reason, the enterprise or organization where you work closes down tomorrow and consequently, all of the workers will be laid off. How certain are you that you will be able to find equally good work as your present job?"

A(findjob): [1] absolutely uncertain;[2] fairly uncertain; [3] both yes and no; [4] fairly certain; [5] absolutely certain.

I construct four dummy variables representing individuals' different degrees of confidence in finding a new job. *Hsecure* (highly secure) is a binary variable that equals to 1 if an individual is absolutely certain in finding a new job, and 0 otherwise. *Rsecure* (relatively secure) is a binary variable that equals to 1 if an individual is fairly certain in finding a new job, and 0 otherwise. *Hinsecure* (highly insecure) is a binary variable that equals to 1 if an individual is absolutely uncertain in finding a new job, and 0 otherwise. *Rinsecure* (relatively insecure) is a binary variable that equals to 1 if an individual is fairly uncertain in finding a new job, and 0 otherwise. Table 2.5 reports the distribution of the *findjob.*

2.4.3 Summary statistics

Table 2.6 presents the summary statistics and definitions of all variables employed in this chapter. Looking at expectation errors indicators, I find that on average, the error score is around 0.151, which indicates that individuals are more prone to make optimistic, instead of pessimistic, forecast errors. Specifically, 8.1% and 22.3% of the sample are extreme and moderate optimism, while only 4% and 16.6% of the sample are extreme and moderate pessimism, respectively.

As for past income changes dummies, I find that in general, 25% of the sample report that their financial situation has improved over the past year, while 19.4% of the sample report having a deteriorated financial situation. Among those who experienced an improvement in financial situation, only 2% of the sample suggest that they have experienced a great improvement, while the 23% of the sample report only experienced a slight improvement. Around 5.5% /13.9% of the sample indicate that their financial situation has greatly/slightly deteriorated. I can see that the bulk (55.6%) of the sample answer there is no change in their financial situation at all.

Turning to job insecurity, the observations have fallen from 114,707 to 60,280 because this question is restricted to employed respondents only. I find that the frequency of the answers to this question are much more evenly spread. I observe that 14.1% and 25.1% of the sample are highly and slightly confident in finding an alternative job in case of sudden laid-off, while 19.5% and 24% of the sample are highly and slightly insecure at job finding.

In terms of control variables, I find that 40% of the sample are males, and 66% of the sample are married or cohabitating, while 13% of the sample are single. Individuals are separated into five age groups: *age1830*, *age3040*, *age4050*, *age5065* and *age65,* which represents 22.2%, 19%, 17.2%, 23.8% and 17.7% of the sample respectively. The average household income in the last month is 44636 Rouble (RUB) and average household consumption is 31572 RUB. The proportion of savers in the sample is 15.1%. As for household composition, the average household size is 3. Over half the sample's highest education is secondary school or less. More precisely, 33% (4.2%) of the sample have completed secondary school (primary school), while 14.7% of the sample dropped out of secondary school. 72.8% of the sample are living in urban areas. 56.5% of the sample are employed, while 25.3% of the sample are retired. The average self-reported health status is 3.16, which suggests that respondents are likely to perceive their health status as good/excellent.

Sample means by past income changes

Panel A and B of Table 2.7 provide the means of each outcome variable by past income changes and job insecurity. In panel A of Table 2.7, I divide the sample into five categories based on the five responses of past income changes. As in regressions, the *nochange* categoty is treated as the omitted category and conclusions are drawn by comparing all four past income changes

indicators with the *nochange* category. Thus, to provide some preliminary evidence, I use the *nochange* category as the reference group to perform the two sample equality of means tests for *gimproved*, s*improved, sworsened* and *gworsened*. On average, I find significant evidence that respondents whose financial situation has greatly (slightly) improved show a 0.592 (0.276) higher *error* score relative to those who observed no change. Respondents whose financial situation has greatly/slightly improved, also exhibit higher (lower) prevalence of being extremely/moderately optimistic (pessimistic). This provides preliminary evidence for H1a.

Compared to those who observed no change, I find significant evidence that on average, respondents whose financial situation has greatly/slightly deteriorated has a lower *error* score (-0.232/-0.033) and a higher probability of being extremely and moderately pessimistic. This is in line with H1b, which suggests that after experiencing a (great/slight) deterioration in financial situation, individuals are more prone to make pessimistic forecast errors. However, I also observe that respondents whose financial situation has slightly deteriorated exhibits higher, rather than lower, prevalence of being extremely and moderately optimistic. This is in contrast to H1b, but because these are just descriptive statistics that do not control for other factors, a more sophisticated analysis is warranted. The differences between those who observed no change (*nochange*) and those whose financial situation has greatly/slightly improved/deteriorated (*gimproved/simproved/sworsened/gworsened*) are all statistically significant.

Sample means by job insecurity

Panel B of Table 2.7 presents the difference in means by job insecurity. Five categories of job insecurity are employed. I use the *indifferent* category as the reference group to perform the two samples equality of means tests for *hsecure*, *rsecure*, *rinsecure* and *hinsecure*. I find significant

evidence that individuals who are highly or slightly confident at finding a new job have, on average, a higher *error* score and higher (lower) chance of being extremely and moderately optimistic (pessimistic). Consistent evidence is also obtained focusing on *hinsecure* and *rinsecure*. First, compared to those who are indifferent at finding a new job, on average I observe significant evidence that those who are highly unconfident in finding a new job have a lower *error* score and lower (higher) chance of being extremely and moderately optimistic (pessimistic). Second, I find significant evidence that those who are relatively unconfident in finding a new job have a lower *error* score and a higher chance of being extremely and moderately pessimistic. These findings are generally consistent and give preliminary evidence for H2a and H2b.

2.5 Model specifications and estimation strategy

This section begins with a discussion of the three baseline specifications employed in this study. Subsequently, an overview of the estimation methods is presented.

2.5.1 The role of past income changes

Errors in financial expectations are defined as the miscalculation between individuals' financial expectations at time *t-1* and realisations at time *t*. In order to disclose the origin of the question why individuals make forecast errors, following Cocco *et al.* (2019) and Massenot and Pettinicchi (2019), I lag all independent and control variables to examine the determinants of errors in financial expectations. Focused on the role of past income changes in predicting the errors, I hereby estimate the following equation:

Equation (2.1)

$$Error_{it} = \beta_0 + \beta_1 gimproved_{it-1} + \beta_2 simproved_{it-1} + \beta_3 sworsened_{it-1}$$

$$+ \beta_4 gworsened_{it-1} + \beta_5 X_{it-1} + \mu_i + \mu_p + \mu_t + \varepsilon_{it}$$

where the subscript *i* indexes individuals; *t*, year (2002-2016); *p*, provinces. $Error_{it}$ denotes one of five errors in financial expectations indicators of individual *i* at time *t*. The error indicator is either a continuous variable (*error1*) or a binary variable (*eopt*, *mopt*, *mpes* or *epes*). *Gimproved*/*simproved* is a binary variable equal to 1 if an individual reports that his/her financial situation has greatly/slightly improved over the past year, and 0 otherwise. *Gworsened*/s*worsened* is a binary variable equal to 1 if an individual reports that his/her financial situation has shown a great/slight deterioration in the past year, and 0 otherwise.

The expected direction of the β coefficients vary with different dependent variables. Consistent with H1a, I expect the coefficients β_1 and β_2 to be positive when the dependent variable is *error1*/*eopt*/*mopt*, and to be negative when the dependent variable is *epes*/*mpes.* In line with H1b, the coefficients β_3 and β_4 are expected to be negative when the dependent variable is *error1*/*eopt*/*mopt*, and to be positive when the dependent variable is *epes*/*mpes.* This is in line with the over-extrapolation phenomenon (Das and van Soest, 1999; Massenot and Pettinicchi, 2019), which suggests that individuals whose financial situation has shown a positive (negative) change are likely to over-extrapolate this positive (negative) experience and make optimistic (pessimistic) forecast errors.

X is a vector of control variables, which includes individuals' age group, gender, marital status, education level, household income and consumption, saving, household size, employment status and health status. Firstly, when the dependent variable is *error1*/*eopt*/*mopt* (*epes*/*mpes*), I expect the coefficients on the younger age groups to be positive (negative) as younger individuals are more impatient and tend to overestimate future income growth (Modigliani and Brumberg, 1954; Das and van Soest, 1997). Next, when the dependent variable is *error1*/*eopt*/*mopt* (*epes*/*mpes*), I expect the coefficient on *male* to be positive (negative) as males are found to be more optimistic biased than females (Jacobsen *et al.*, 2014). Third, the coefficient on *cohabit* is also expected to be positive (negative) when the dependent variable is *error1*/*eopt*/*mopt* (*epes*/*mpes*) as married individuals have more optimistic expectations as a couple (Khodarahimi, 2015).

Fourth, as for education, I expect the coefficients on primary and secondary education (*edup*, *edus*) to be negative (positive) when the dependent variable is *error1*/*eopt*/*mopt* (*epes*/*mpes*) as

the less-educated are likely to suffer from less income and less job opportunities (Duval and Wolff, 2012). Fifth, I expect the coefficients associated with income and saving to be positive (negative) when the dependent variable is *error1*/*eopt*/*mopt* (*epes*/*mpes*) as higher income and saving encourage imprudent financial expectations (Carroll, 1997). Sixth, in line with Das and van Soest (1999), the coefficient on the employed dummy is expected to be positive (negative) when the dependent variable is *error1*/*eopt*/*mopt* (*epes*/*mpes*). Seventh, I expect the coefficient on the retired dummy to be negative (positive) when the dependent variable is *error1*/*eopt*/*mopt* (*epes*/*mpes*) as retired people have a relatively stable income source (pension) and are therefore less likely to over-estimate their financial situation (Das and van Soest, 1997). Lastly, I expect the coefficient on health status to be positive (negative) when the dependent variable is *error1*/*eopt*/*mopt* (*epes*/*mpes*). This could be explained by the fact that individual with a better health condition is more likely to have a proper financial planning horizon and over-estimate future financial situation (Bissonnette and Soest, 2015). The definitions of all variables used are provided in Table 2.6.

In equation (2.1), μ_i is an individual-specific component that captures time-invariant individual heterogeneity, such as fixed personal traits. I control for this component by using a fixed-effect estimator in all regressions. μ_p is a province-specific component that takes into account the heterogeneity across different provinces. I control for this by including province dummies. μ_t is a time-specific component that captures business cycle effects. I control for this component by including year dummies in all regressions. Finally, ε_{it} is an idiosyncratic component.

2.5.2 The role of job insecurity

To examine the role of job insecurity in determining errors in financial expectations, following Bissonnette and Soest (2015) and Cocco *et al.* (2019), I hereby estimate the following equation:

Equation (2.2)

$$Error_{it} = \beta_0 + \beta_1 hsecure_{it-1} + \beta_2 rsecure_{it-1} + \beta_3 rinsecure_{it-1}$$

$$+ \beta_4 hinsecure_{it-1} + \beta_5 X_{it-1} + \mu_i + \mu_p + \mu_t + \varepsilon_{it}$$

where $Error_{it}$ indexes the indicators of errors in financial expectations from equation (2.1). *Hsecure*/*rsecure* is 1 if an individual is highly/relatively secure at finding a new job, and 0 otherwise. *Hinsecure*/*rinsecure* is 1 if an individual is highly/relatively insecure at finding a new job, and 0 otherwise. *X* indexes the same vector of control variables as in equation (2.1). The subscripts and the corresponding terms are also the same as in equation (2.1).

In accordance with H2a, I expect the coefficients β_1 and β_2 to be positive when the dependent variable is *error1*/*eopt*/*mopt*, and negative when the dependent variable is *epes*/*mpes*. Consistent with H2b, the coefficients β_3 and β_4 are expected to be negative when the dependent variable is *error1*/*eopt*/*mopt*, and to be positive when the dependent variable is *epes*/*mpes*.

2.5.3 The roles of past income changes and job insecurity

Additionally, to examine the joint roles of past income changes and job insecurity in determining errors in financial expectations, I estimate the following equation:

Equation (2.3)

$$Error_{it} = \beta_0 + \beta_1 gimproved_{it-1} + \beta_2 simproved_{it-1} + \beta_3 sworsened_{it-1}$$

$$+ \beta_4 gworsened_{it-1} + \beta_5 hsecure_{it-1} + \beta_6 rsecure_{it-1} + \beta_7 rinsecure_{it-1}$$

$$+ \beta_8 hinsecure_{it-1} + \beta_9 X_{it-1} + \mu_i + \mu_p + \mu_t + \varepsilon_{it}$$

where $Error_{it}$ indexes the indicators of errors in financial expectations from equation (2.1). *Gimproved*/*simproved* is a binary variable equal to 1 if an individual reports that his/her financial situation has greatly/slightly improved over the past year, and 0 otherwise. *Gworsened*/s*worsened* is a binary variable equal to 1 if an individual reports that his/her financial situation has shown a great/slight deterioration in the past year, and 0 otherwise. *Hsecure*/*rsecure* is 1 if an individual is highly/relatively secure at finding a new job, and 0 otherwise. *Hinsecure*/*rinsecure* is 1 if an individual is highly/relatively insecure at finding a new job, and 0 otherwise. *X* indexes the same vector of control variables as in equation (2.1). The subscripts and the corresponding terms are also the same as in equation (2.1).

With reference to the discussion in section 2.5.1, I expect the coefficients β_1 and β_2 to be positive when the dependent variable is *error1*/*eopt*/*mopt*, and to be negative when the dependent variable is *epes*/*mpes.* The coefficients β_3 and β_4 are expected to be positive when the dependent variable is *epes*/*mpes,* and to be negative when the dependent variable is *error1*/*eopt*/*mopt.* Consistent with the discussion in section 2.5.2, I expect the coefficients β_1 and β_2 to be positive when the dependent variable is *error1*/*eopt*/*mopt*, and to be negative when the dependent variable is *epes*/*mpes*. The coefficients β_3 and β_4 are expected to be positive when the dependent variable is *epes*/*mpes*, and to be negative when the dependent variable is *error1*/*eopt*/*mopt.*

Table 2.8 reports the correlation matrix of all variables. For the past income changes dummies, the correlation coefficients range from -0.0321 to -0.2371, which implies that there are weak and negative correlations between g*improved, simproved, gworsened* and s*worsened*. I also find that the correlation coefficients between job insecurity dummies range from -0.2004 to -0.2855, which indicates a weak and negative correlation. I only observe one coefficient exceeding 0.5. The correlation coefficient between *lghhincome* and *lgconsumption* equals to 0.5938, which suggests a moderate positive correlation between household income and consumption. This, however, is below the rule of thumb threshold of 0.7[4]. Overall, I find that there are only weak correlations between the right-handed side variables, which suggests that the following empirical results are not affected by multicollinearity.

2.5.4 Estimation strategy

There are two main issues concerning the estimation of the above specifications: unobserved time-invariant individual heterogeneity and endogeneity.

Fixed-effect estimation

First, I make use of a fixed-effect (FE) estimator to control for time-invariant individual heterogeneity (e.g individuals' personality traits). For instance, an individual could be always optimistic and hopeful towards life challenges in general, which is considered as a fixed personality trait. This general positive outlook may affect how individuals perceive their past income changes and whether they are more prone to make optimistic forecast errors.

Fixed-effect instrumental variable estimation

4 According to Hinkle *et al.* (2003), the rule of thumb for correlation analysis suggests that only when the correlation coefficient exceeds 0.7, there is a strong correlation between variables, which should be treated with caution.

Second, there is a concern that the past income changes and job insecurity indicators are likely to be endogenous due to some unobserved time-varying variables. For example, confounding variables like financial literacy, cognitive ability or social connections may affect past income changes (job insecurity) and errors in financial expectations simultaneously. To motivate the need for instrumental variables estimation, consider the following model:

Equation (2.4)

$$Y_{it} = \beta_0 + \beta_1 X_{1it-1} + \beta_2 X_{2it-1} + \beta_3 X_{3it-1} + \ldots + \beta_{l-1} X_{jit-1}$$

$$+ \beta_l C_{it-1} + \mu_i + \mu_p + \mu_t + \varepsilon_{it}$$

where *Y* indexes either a continuous variable (*error1*) or a binary variable (*eopt*, *mopt*, *mpes* or *epes*). $X_1,\ldots, X_j$ are the potential endogenous variables. With reference to equation (2.1), there are four endogenous variables (*Gimproved*, *simproved, Gworsened*, s*worsened*). With regards to equation (2.2), there are four endogenous variables (*Hsecure*, *rsecure, Hinsecure*, *rinsecure*). In reference to equation (2.3), there are eight endogenous variables (*Gimproved*, *simproved, Gworsened*, s*worsened, Hsecure*, *rsecure, Hinsecure*, *rinsecure*). *C* indexes exogenous control variables. The subscript *j* indicates the number of endogenous variables in the equation. Other subscripts and the corresponding terms are also the same as in equation (2.1).

To address the issue of endogeneity, I employ a fixed-effect instrumental variable (FEIV) approach to control for the potential endogeneity of the key independent variables. Following Xu (2013) and Staudigel (2016), I use two sets of instruments. The first includes the interaction between the province-specific unemployment rate (provided by the Russian Federal Statistical Office) and age categories. The second set of instruments comprises community-level means of

the past income changes (job insecurity) indicators where each individuals' value has been excluded from the average community statistics. Staudigel (2016) suggests that regional unemployment rates and community-level means of variables are both appropriate instruments. This is because these instruments are presumably correlated with individuals' past income changes and job insecurity, but are not influenced by unobserved individual characteristics (e.g., financial planning ability, cognitive ability).

Specifically, I estimate the following first stage equation to obtain the instrumental variables:

Equation (2.5)

$$X_{mit-1} = \beta_0 + \beta_1 \overline{X}_{1kt-1} + \beta_2 \overline{X}_{2kt-1} + \beta_3 \overline{X}_{3kt-1} + \ldots + \beta_{l-2} \overline{X}_{jkt-1}$$

$$+ \beta_{l-1} UNEM_{pt-1} * AGE_{it-1} + \beta_l C_{it-1} + e_{it-1}$$

where the subscript *i* indexes individuals; *t*, year; *p*, provinces; *k*, community; *j*, number of endogenous variables; *m*, endogenous variable. X_{mit-1} stands for each endogenous variable of individual *i* at time *t-1*. $\overline{X}_{jkt-1}$ denotes the mean of endogenous variables of community k at time *t-1*. $UNEM_{pt-1}$ represents the unemployment rate of province *p* at time *t-1*. AGE_{it-1} denotes the age group of individual *i* at time *t-1*. e_{it-1} is an idiosyncratic term. Each endogenous variable is instrumented using the interaction between province-specific unemployment rate and age categories, and the community-level means of all of the endogenous variables.

The challenge of using instrumental variables estimation is to obtain unbiased estimates, which requires that the instruments are valid. There are two requirements for a valid instrument: 1) Relevance, i.e., the instrument is significantly correlated with the endogenous variable, and 2) Exogeneity, i.e., the instrument is uncorrelated with the error term from original specifications

(Wooldridge, 2010). To access whether the instruments implemented is valid, I next present the first stage results of all of the FEIV models and discuss a full set of tests of instrument validity.

Appendix 2.1 presents the first stage results and relevant instrument validity tests of past income changes for expectation error models (*error1*/*eopt*/*mopt*/*mpes*/*epes*). Since the instruments for each expectation model are the same, Panel A of Appendix 2.1 only reports the one and only set of first stage results of past income changes. I observe that the community level means of the past income dummy is always a significant determinant of the corresponding past income changes dummy. For instance, in column 1, I find that the *gimproved_mean* term is significantly and positively associated with the *gimproved* term, with a coefficient of 0.472 which is statistically significant at the 1 percent level. In addition, some of the interaction terms of the unemployment rate and age categories are also significant determinants of the past income changes indicators. The above evidence shows that the community-level mean of the past income changes indicators and interaction terms are important and significant predictors of the respective past income changes indicators.

To formally test the relevance of instruments, the weak identification test and under-identification test are conducted. For detecting weak instruments, the first stage F-statistics provided by Sanderson and Windmeijer (2016) are used to examine whether the coefficients of the excluded instrumental variables are significantly different from zero. The null hypothesis of weak instruments is rejected at the 1 percent level in all first stage regressions. For detecting under-identification, the Kleibergen-Paap rk LM statistic (Kleibergen and Paap, 2006) are used, testing the null hypothesis that the instrumental variables have insufficient explanatory power to predict the endogenous variables in the model for identification of the parameters. Panel B of

Appendix 2.1 reports the tests of instrument validity of past income changes. I observe that the null hypothesis of under-identification is rejected at the 1 percent level. Thus, the overall findings indicator that the instruments used satisfy the requirement of relevance.

Next, to examine the exogeneity of the instrumental variables, the Hansen J-test of over-identifying restrictions (Hansen, 1982) are proposed. I find that my instruments cannot reject the null of the Hansen J-test, which suggests that all instruments used are exogenous.

I find similar evidence in Appendix 2.2 and 2.3, showing that the instruments used are not under-identified, or weakly identified, or endogenous. To sum up, the first stage results of the FEIV models provide consistent evidence that the instruments utlised in this chapter are valid and the FEIV estimation are unbiased.

2.6 Empirical results

2.6.1 Empirical results - the role of past income changes

Expectation error (*error1****)***

Table 2.9 presents the FE and FEIV estimates of equation (2.1). Focusing on expectation error (*error1*), columns 1 and 2 show consistent evidence that those who experienced a positive (negative) income change show a significant and positive (negative) expectation error. More specifically, I find that individuals whose financial situation has shown a great (slight) improvement show a 0.783 and 2.474 (0.383 and 0.322) units higher expectation error in the FE and FEIV regressions respectively. These coefficients are significant at the 1% level. This is in line with H1a, which suggests that individuals over-extrapolate their positive experience and make over-optimistic forecast errors (Massenot and Pettinicchi, 2019).

Columns 1 and 2 of Table 2.9 also show that respondents who experienced a great (slight) deterioration in their financial situation display a 0.611 and 1.045 (0.213 and 0.278) units lower expectation error in the FE and FEIV models respectively. The coefficients are significant at the 10% level or beyond. This supports H1b, which suggests that respondents tend to become overpessimistic when their financial situation has deteriorated in the past year.

Furthermore, I examine whether individuals react asymmetrically to positive and negative income changes. Focusing on the FEIV results (column 2 of Table 2.9), I find that the coefficient associated with *gimrpoved* (*gworsened*) is significantly (Wald test, p=0.003) larger than that associated with *rimproved* (*sworsened*). This indicates that individuals whose financial situation has greatly improved (deteriorated) are more likely to make larger optimistic (pessimistic)

forecast errors than those with a slight improvement (deterioration). Comparing the coefficients associated with *gimrpoved* (*simproved*) to those associated with *gworsened* (*sworsened*), I observed that the former are always larger (in absolute value) than the latter (Wald test, p=0.000). This suggests that individuals over-extrapolate positive events more than negative ones. This is in line with Massenot and Pettinicchi (2019), who suggest that there is an asymmetry of over-extrapolation between positive and negative income changes. A similar pattern has also been found using FE estimates for the expectation error (*error1*) in column 1 of Table 2.9.

Extreme optimism (eopt)

Columns 3 and 4 of Table 2.9 present the FE and FEIV estimates of the determinants of extreme optimism (*eopt*). Consistent with H1a, I find that respondents whose financial situation has shown a great improvement have a 13.2 and 53.1 percentage point higher probability of being extreme optimists in the FE and FEIV regressions respectively. This is in line with Cocco *et al.* (2019) who find that individuals who experienced a great improvement in their financial situation expect their situation to keep improving at the same rate, which eventually leads them to make extreme optimistic forecast errors.

The FE results also suggest that respondent whose financial situation has slightly improved (greatly/slightly deteriorated) are more (less) likely to show moderate optimism, which is in line with H1a and H1b. However, I do not find the same significant association between *simproved/gworsened/sworsened* and *mopt* in the FEIV regressions. Hence, there is inconsistent evidence between the FE and FEIV estimates.

With reference to the discussion in section 2.5.4, the instruments I used in the FEIV regressions are shown to be exogenous (Hansen J-test) and not weakly identified (Sanderson-

Windmeijer F-stat). Thus, I suggest that the inconsistent evidence could arise for the following reasons. First, if the past income changes indicators are exogenous, then FE and FEIV estimates are both consistent but FEIV would be much less efficient (i.e generate larger standard errors). Secondly, if the past income changes indicators are in fact endogenous, then the FE estimates yield biased results and the FEIV results should be trusted. Therefore, I mainly discuss the FEIV estimates along with the FE results. All in all, I find supporting evidence that individuals are likely to over-extrapolate their positive income changes and make optimistic forecast errors.

Moderate optimism (mopt)

Columns 5 and 6 of Table 2.9 report results of the FE and FEIV estimates for moderate optimism (*mopt*). In line with H1a, I observe consistent evidence that individuals who experienced a great (slight) improvement in their financial situation over the past year show a 21 and 64.7 (14.9 and 21) percentage point higher probability of being moderate optimism in the FE and FEIV regressions respectively. The coefficients are significant at the 1% level.

Furthermore, consistent with H1b, I see that individuals whose financial situation has shown a great deterioration have a 4.2 and 15.5 percentage point lower probability of showing moderate optimism. The corresponding coefficients are highly significant at the 1% level. These findings are in line with Cocco *et al.* (2019) who suggest that individuals exaggerate the persistence of their past experience. As a result, those whose financial situation has improved (deteriorated) are more (less) likely to show optimistic forecast errors.

As for the asymmetry of the over-extrapolation of past experience, I find similar evidence as in the *error1* regressions. In other words, those whose financial situation has greatly improved show a higher probability of displaying moderate optimism than those with a slight improvement

(Wald test, p=0.000). As in Massenot and Pettinicchi (2019), in columns 5 and 6 of Table 2.9, I also observe that the coefficient associated with *gimproved* is significantly larger in absolute value than the coefficient on *gworsened* (Wald test, p=0.000). This shows that individuals over-extrapolate positive experiences more than negative ones.

Moderate and Extreme pessimism (mpes and epes)

Columns 7 and 8 of Table 2.9 present the estimates obtained using the FE and FEIV models for moderate pessimism (*mpes*), and columns 9 and 10 of Table 2.9 report the results for extreme pessimism (*epes*). In line with H1b, I find that respondents whose financial situation has greatly deteriorated show a 4.8 and 19.4 (16.1 and 29.5) percentage point higher probability of displaying moderate (extreme) pessimism in the FE and FEIV models respectively. For those who only experienced a slight deterioration, I find that *sworsened* is significantly and positively (0.05 and 0.155) associated with moderate pessimism in the FE and FEIV estimates respectively. I also observe a positive relationship between *sworsened* and extreme pessimism with estimated coefficients of 0.05 and 0.081 in the FE and FEIV regressions respectively. The above coefficients are all significant at the 5% level or beyond.

What's more, in line with H1a, I find that respondent whose financial situation has greatly (slightly) improved have a 15.2 and 57.8 (11.1 and 9.7) percentage point lower probability of showing moderate pessimism in the FE and FEIV models respectively. These coefficients are significant at the 1% level. These findings are in line with Massenot and Pettinicchi (2019), who suggest that after experiencing a negative income change, individuals are more likely to be over-pessimistic and less likely to be over-optimistic. This is because individuals with negative income

changes will expect to experience the same disappointing situation again in the next period and thus show pessimistic forecast errors.

To sum up, I find robust evidence that individuals whose financial situation has improved are more likely to make optimistic forecast errors (e.g to show extreme and moderate optimism) and less likely to make pessimistic forecast errors (e.g., to show extreme and moderate pessimism). A positive relationship between respondents with negative income changes and pessimistic forecast errors has also been documented. Additionally, I find that there is an asymmetry of the over-extrapolation of past income changes, which suggests that individuals who experienced a greater improvement (deterioration) in their financial situation are more likely to show a larger optimistic (pessimistic) forecast error.

Control variables

Moving to control variables, I find that younger individuals are more likely to overestimate their future finance and show optimistic forecast errors, while older individuals are more prone to be over-pessimistic. This is in line with the life-cycle theory (Modigliani and Brumberg, 1954), which suggests that young individuals at the beginning of their career are more impatient and more likely to anticipate income jumps and show over-optimism, while their older counterparts show the opposite (over-pessimism) (Puri and Robinson, 2007). Secondly, turning to wealth controls, I find that the coefficients on the wealth variables (income, consumption and saving) are mostly insignificant. This could be explained by the fact that the effects of individuals' wealth changes are already captured by the past income changes indicators, and thus wealth controls turn insignificant (Staugigel, 2016).

Third, I find that those who are married or living with their partners are more (less) likely to show optimistic (pessimistic) forecast errors. This is because cohabitating couples are more likely to share commitments and responsibilities to achieve some common goals, and then overestimate the chance to succeed along the way (Khodarahimi, 2015). Fourth, focusing on education, I find that compared to the higher-educated, the less-educated are less likely to make optimistic forecast errors. This is consistent with Duval and Wolff (2012), who indicate that the low-educated typically have lower permanent income and wealth and fewer job opportunities in the long run, and thus are more conservative regarding their future income growth.

Fifth, I find that individuals living in urban areas are more likely to make forecast errors (both optimistic and pessimistic). This could be explained by the fact that urban respondents face more income and job uncertainties in the cities and hence are more likely to make erroneous forecast compared to rural respondent who are more capable to correctly predict their future (Duval and Wolff, 2012). Sixth, in contrast to Das and van Soest (1999), I find that respondents who are employed are less (more) likely to make optimistic (pessimistic) forecast errors than those who are not in the labor force. A plausible explanation is that individuals who are not in the labor force are less informated regarding job and income perspectives and, as such, may become over-optimistic.

Lastly, I observe consistent evidence that individuals who are retired or report worse health status are more prone to show pessimistic forecast errors and less likely to report optimistic forecast errors. This is in line with Das and van Soest (1997), who show that retirees mostly rely on stable pension income and are thus less likely to overestimate their future finance situation.

Bissonnette and Soest (2015) also suggest that individuals with health problems are less likely to make proper financial planning and, as a result, overestimate their future income growth.

2.6.2 Empirical results - the role of job insecurity

Table 2.10 reports the FE and FEIV estimates of the determinants of errors in financial expectations as a function of job insecurity and other controls. Columns 1 and 2 present the FE and FEIV estimates of the determinants of expectation error (*error1*). Consistent with H2a, I observe that individuals who are highly confident at finding a new job if laid-off (*hsecure*) display a 0.124 and 0.336 units higher expectation error, respectively in the FE and FEIV models. The coefficients are significant at the 1% and 10% level respectively. This is in line with Bissonnette and Soest (2015) who suggest that individuals who perceive a stable job prospect are more likely to have imprudent financial expectations thus display a positive expectation error.

In line with H2b, respondents who are highly unconfident at finding an alternative job (*hinsecure*) show a significant and negative expectation error in the FE model. The coefficient is -0.042 and significant at the 5% level. However, the coefficients on *hinsecure* are insignificant in the FEIV models.

In reference to section 2.5.4, given that the instruments chosen are exogenous and not weakly identified, I mainly focus on the FEIV estimates along with the FE results. Moving to the determinants of extreme optimism (*eopt*), columns 3 and 4 of Table 2.10 show robust evidence that those who are highly confident at finding a new job show a 3.5 and 11.3 percentage point higher probability of displaying extreme optimism in the FE and FEIV models respectively. This is in line with H2a, which indicates that the more secure individuals are at work, the more

optimistic errors they show (Das and van Soest, 1999). Moreover, I find that respondents who are highly unconfident at job finding report a 12.4 percentage point lower probability of being extreme optimists. The above coefficients are significant at the 5 percent level or beyond.

Columns 5 and 6 of Table 2.10 present the FE and FEIV estimates of the determinants of moderate optimism (*mopt*). Consistent with H2a, I find that respondents who are highly (relatively) confident at job finding report a 1.9 (1.2) percentage point higher probability of being moderate optimists in the FE regression. The coefficient is significant at the 5% (10%) level. In addition, comparing the coefficient associated with *hsecure* to that associated with *rsecure*, I observed that the former is larger than the latter (Wald test, p=0.000). This indicates that the more confident individuals are, the higher the probability of displaying optimistic forecast errors. However, the FEIV estimates of the effect of *hsecure/rsecure* on *mopt* turn insignificant and thus the findings of the FE estimates of moderate optimism should be treated with caution.

Columns 7 and 8 of Table 2.10 report the results obtained using the FE and FEIV models for moderate pessimism (*mpes*). In line with H2a, I observe consistent evidence that respondents who are highly confident at finding a new job are characterized by a lower probability of showing moderate pessimism. The estimated coefficients are -0.028 and -0.133 in the FE and FEIV models respectively. Estimates of other job insecurity indicators (*rsecure/rinsecure/hinsecure*) are insignificant across the FE and FEIV models. Columns 9 and 10 of Table 2.10 present the FE and FEIV results of the determinants of extreme pessimism (*epes*). The coefficient associated with *rsecure* (*hinsecure*) is significant and negative (positive) in the FE models, while the coefficients associated with job insecurity indicators are insignificant in the FEIV models.

In conclusion, I find robust evidence that individuals who are over-confident in their ability to find a new job are more prone to make imprudent financial expectations. In other words, they are more likely to make larger optimistic forecast errors and to show extreme optimism. However, I also observe that the relationship between job insecurity indicators (except *hsecure*) and extreme and moderate pessimism are mostly insignificant.

There is a plausible explanation for this insignificance. Mandal *et al.* (2011) suggest that respondents may have some private information regarding the consequences of job volatility that is not observable to the researcher. For instance, following unemployment, individuals in Russia can sign up to receive unemployment benefits for up to 12 months and the amount of benefits makes up to 75 percent of the original wage during the first 3 months and 60% during the next four months (Pogosov and Sokolovskaya, 2016). Thus, individuals who feel insecure at finding a new job may not necessarily dramatically lower their financial expectations and show pessimistic forecast errors.

As for the coefficients associated with the control variables, most of them have the same signs and significance as discussed in Section 2.6.1. Notably, I find consistent evidence that wealthier individuals (with higher income and more saving) are more likely to show optimistic forecast errors given that they can afford to make some financial mistakes (Carroll, 1997). This finding supports the previous argument that the wealth controls are not significant in the past income changes models because the inclusion of past income indicators may already capture the effects of individuals' wealth changes on expectation errors. Thus, in the job insecurity models, I find that the effects of wealth controls (income and saving) on expectation errors are significant and positive.

2.6.3 Empirical results - the roles of past income changes and job insecurity

Table 2.11 presents the FE and FEIV estimates of equation (2.3). Including past income changes and job insecurity indicators in one estimation enables me to compare which determinant is most important. Columns 1 and 2 report the FE and FEIV estimates of the determinants of expectation error (*error1*). I find that individuals whose financial situation has greatly (slightly) improved show a 0.792 and 1.903 (0.382 and 0.411) units higher expectation error, and that individuals who experienced a great (slight) deterioration in their financial situation show a 0.553 and 0.641 (0.188 and 0.523) units lower expectation error in the FE and FEIV regressions respectively. These coefficients are significant at the 5% level or beyond. These findings are in line with findings in section 2.6.1, which suggests that the role of past income changes in determining expectation errors is not affected by the inclusion of job insecurity indicators.

Moving to the job insecurity indicators in columns 1 and 2 of Table 2.11, I observe that the estimated coefficients associated with *hsecure*, *rsecure* and *hinsecure* are only significant in the FE model. However, comparing to the results obtained in section 2.6.2, I find that the estimated coefficient associated with *hsecure* turns to insignificant in the FEIV model. This implies that the role of job insecurity in determining expectation error is comprised by the inclusion of past income changes.

Moreover, in column 1 of Table 2.11, comparing the coefficient associated with *gimproved* (*simproved*) to that associated with *hsecure* (*rsecure*), I observe that the former is always larger than the latter (Wald test, p=0.000). This suggests that individuals whose financial situation has shown improvements are more likely to make over-optimistic forecast errors than those who are

highly confident at finding a new job if laid-off. To sum up, the above findings indicate that both past income changes and job insecurity are significant determinants of expectation error. However, the effects of past income changes are stronger and more significant than those of job insecurity on expectation error. This suggests that past income changes are the most important determinant of expectation error.

Focusing on the determinants of extreme optimism (*eopt*), columns 3-4 of Table 2.11 show that those whose financial situation has greatly (slightly) improved have a 12.8 and 24.8 (4.2 and 9.3) percentage point higher probability of being extreme optimists in the FE and FEIV estimates respectively. The coefficients are significant at the 10 percent level or beyond. In addition, comparing the coefficients associated with *gimproved* (*rimproved*) to those associated with *hsecure* (*rsecure*), I find that the former is larger and more significant than the latter in the FE models. This is in line with previous findings in the *error1* regressions that past income changes are the more important determinant of expectation error.

Columns 5 and 6 of Table 2.11 report the FE and FEIV estimates of the determinants of moderate optimism (*mopt*). Consistent with findings in the *eopt* regressions, I find that individuals who experienced a great (slight) improvement in their financial situation display a 20.8 and 73.6 (13.6 and 16.1) percentage point higher probability of being moderate optimism in the FE and FEIV regressions respectively. The coefficients are significant at the 1 percent level. Furthermore, the estimates of all four job insecurity indicators are insignificant across FE and FEIV models.

Columns 7 and 8 of Table 2.11 report the results obtained using the FE and FEIV models for moderate pessimism (*mpes*), and columns 9 and 10 of Table 2.11 present the estimates for extreme pessimism (*epes*). I find that respondents whose financial situation has slightly worsened show a 0.034 and 0.175 (0.046 and 0.141) percentage point higher probability of displaying moderate (extreme) pessimism in the FE and FEIV estimates respectively. For those who experienced a great deterioration in financial situation, I observe that *gworsened* is significantly and positively (0.035) associated with moderate pessimism in the FE estimates. I also find a positive relationship between *gworsened* and extreme pessimism, with the estimated coefficients of 0.139 and 0.258 in the FE and FEIV models respectively. The above coefficients are significant at the 10 percent or beyond. Turning to job insecurity, the coefficient associated with *hsecure* (*hinsecure*) is significant and negative (positive) in the FE models, while the coefficients associated with all job insecurity indicators are insignificant in the FEIV models.

To sum up, by pooling past income changes and job insecurity into one estimation, I find consistent evidence that both past income changes and job insecurity are significant determinants of expectation error. Additionally, I find that the effects of past income changes on expectation error are stronger and more significant than those of job insecurity. This implies that past income changes are the more important determinant of errors in financial expectations than job insecurity.

2.7 Robustness checks

This section begins by using correlated random effects (CRE) ordered Probit models to estimate the determinants of expectation errors. Subsequently, I examine whether my main findings suffer from attrition bias.

2.7.1 Correlated random effects - ordered Probit model

Following Brown and Taylor (2006) and Cocco *et al.* (2019), I construct a five-point scale ordinal variable (*error*) to indicate individuals' errors in financial expectations. The ordinal variable (*error*) equals to 1 if an individual is an extreme pessimist (*epes*), 2 if the individual is a moderate pessimist (*mpes*), 3 if the individual correctly predicts his/her future finance (*correct*), 4 if the individual is a moderate optimist (*mopt*) and 5 if the individual is an extreme optimist (*eopt*).

In order to control for the unobserved time-invariant heterogeneity, following Wooldridge (2019), I employ the correlated random effects ordered Probit model to estimate the determinants of *error*. The CRE method parameterises the individual specific component of the error term as a function of the mean values of time-varying variables in the regressions. Suppose $\mu_i = \alpha_1 \overline{X}_i + e_i$, where μ_i is the unobserved time-invariant individual hetergogeneity, the overbar indexes the mean of time-varying independent variables, and e_i is the remaining individual-specific component that is unrelated to the independent variables (X_{it-1}). I estimate the CRE ordered Probit models by plugging in the μ_i equation into equation 2.1 (2.2, 2.3) and get the following equation 2.6:

Equation (2.6)

$$Error^{*}_{i,t} = \beta_0 + \beta_1 X_{i,t-1} + \beta_2 C_i + \alpha_1 \overline{X}_i + e_i + u_p + u_y + \varepsilon_{i,t}$$

$$Error_{i,t} = j \quad if \ \ \theta_{j-1} < Error^{*}_{i,t} \leq \ \theta_j \quad j = 1,\ldots, 5$$

The central ideal of the CRE ordered Probit model is that an individual's expectation error is a latent variable, $Error^{*}_{i,t}$. This latent variable is a function of observed and unobserved variables. Specifically, $X_{i,t-1}$indexes the time-varying independent variables, $\overline{X}_i$ represents the mean of time-varying independent variables, C_i indexes the time-invariant independent variables, other variables and subscripts are the same as quoted in equation (2.1). The same set of independent and control variables are employed as in equations (2.1), (2.2) and (2.3) respectively. Individuals' observed expectation error, $Error_{i,t}$ is an ordinal response variable and assumed to be a function of the latent variable $Error^{*}_{i,t}$. Each level of the latent expectation error ($Error^{*}_{i,t}$) fall within the cut-off range defined by two cut-off points θ_{j-1} and θ_j. Each cut-off range corresponds to one of the 5-possible observed expectation error response ($Error_{i,t}$). It is assumed that the cut-offs points are increasing in *j*.

Table 2.12 presents the coefficients of the equation 2.6, as well as marginal effects (MEs) for each level of expectation error (ordinal, *error2*). Although the coefficients of an ordered Probit model cannot be interpreted as predicted changes in the response variable, but they correctly provide an indication of statistical significance and the direction of the effect. The MEs are also presented for each level of *error2*. The MEs from ordered Probit regressions show the change in the probability of an individual reporting each of the five possible levels of *error2* associated with a unit change in the independent variables. In the case of a dummy independent

variable, the MEs display the difference in probability of reporting the given *error2* level compared to the reference group.

Panel A of Table 2.12 reports the results of the determinants of expectation error *(error2*) as a function of past income changes. Consistent with H1a, in columns 5 and 6 of Panel A of Table 2.12, I observe that individuals whose financial situation has greatly (slightly) improved are more likely to show moderate and extreme optimism, with estimated marginal effects of 0.168 and 0.125 (0.0853 and 0.0637) respectively. On the other hand, I find that respondents who experienced a great (slight) deterioration are less likely to display moderate and extreme optimism. The marginal effects associated with *gworsened* (*sworsened*) are -0.127 and -0.0949 (-0.0475 and -0.0355) in the *mopt* and *eopt* models respectively. The above marginal effects are all significant at the 1 percent level.

In line with H2b, in columns 2 and 3 of Panel A of Table 2.12, I observe consistent evidence that respondents whose financial situation has greatly (slightly) worsened show a 12.6 and 5.49 (4.7 and 2.05) percentage point higher probability of displaying moderate and extreme pessimism respectively. Meanwhile, individuals who experienced a great (slight) improvement in their financial situation report a 16.6 and 7.25 (8.45 and 3.69) percentage point lower probability of displaying moderate and extreme pessimism respectively. These marginal effects are all significant at the 1% level.

The results obtained from the CRE Probit models are similar to those obtained from the FE and FEIV models. They suggest that individuals are more likely to over-extrapolate their past experience and expect the same situation to repeat itself (Massenot and Pettinicchi, 2019). Thus,

those who experienced a positive income change are more likely to overestimate their financial situation over the next year, while those who experienced a negative income change are more prone to underestimate their future finances.

Panel B of Table 2.12 presents the results of the determinants of expectation error *(error2*) as a function of job insecurity. Consistent with H2a, in columns 11 and 12 of Panel B of Table 2.12, I find that individuals who are highly (relatively) confident at finding a new job show a 2.8 and 2.24 (1.06 and 0.85) percentage point higher probability of being moderately and extremely optimistic respectively. I also find that respondents who are highly (relatively) confident at finding an alternative job report a 2.82 and 1.12 (1.06 and 0.43) percentage point lower probability of being moderate and extreme pessimists respectively. The marginal effects are significant at the 1 percent level. These findings, which are in line with the results obtained using the FE estimates, indicates that individuals who perceive a more stable job prospect are more likely (less) to have imprudent financial expectations and display moderate and extreme optimism (pessimism) (Bissonnette and Soest, 2015).

Focusing on the marginal effects of the *hinsecure* and *rinsecure* of Panel B of Table 2.12, I only find weak association (significant at the 10%) between *hinsecure* and *eopt*/*mopt*/*mpes*/*epes*, while no significant relationship between *rinsecure* and *eopt*/*mopt*/*mpes*/*epes* is observed. This is in line with the previous FE and FEIV results, which indicates that insecure individuals are less likely to adjust their expectations and to display pessimistic forecast errors as they have unemployment benefits and insurance to fall back on (Pogosov and Sokolovskaya, 2016).

Panel C of Table 2.12 presents the results of the determinants of expectation error *(error2*) as a function of past income changes and job insecurity. In columns 17 and 18 of Panel C of Table 2.12, I observe that individuals whose financial situation has greatly (slightly) improved show a 16.9 and 13.5 (8.4 and 6.71) percentage point higher probability of displaying moderate and extreme optimism. I also find that individuals who are highly (relatively) confident at finding a new job show a 2.22 and 1.77 (0.73 and 0.58) percentage point higher probability of displaying moderate and extreme optimism. The above marginal effects are significant at the 5 percent level or beyond.

Turning to columns 14 and 15 of Panel C of Table 2.12, I find that those whose financial situation has greatly (slightly) are more likely to show moderate and extreme pessimism. The marginal effects associated with *gworsened* (*sworsened*) are 0.0453 and 0.1140 (0.0167 and 0.0418) in the *epes* and *mpes* models respectively. As for job insecurity indicators, the marginal effects associated with hinsecure are 0.0026 and 0.0067 in the *epes* and *mpes* models respectively. The above marginal effects are all significant at the 10 percent level or beyond. All in all, the inclusion of both determinants does not affect the significance and sign of each other, which suggests that both past income changes and job insecurity are significant determinants of expectation error.

2.7.2 Attrition in the sample

Jones *et al.* (2006) state that using an unbalanced panel data creates a risk that the empirical results will be biased by non-random sample attrition, sample selection and self-selection. In line with Wooldridge (2002), my sample could suffer from sample selection as certain individuals do

not appear in a random sample due to individual choices or behaviours. Specifically, because the errors in financial expectations variables in analysis are defined as changes in variables between adjacent waves of the survey, my sample only consists of individuals who have answered the two related questions in adjacent waves.

To examine whether our main results are robust to attrition bias, I apply the variable addition tests proposed by Nijman and Verbeek (1992) and Jones *et al.* (2006). Firstly, I construct a continuous measure denoting the number of waves (*num_wave*) each individual participated in. I include this variable as an additional control variable in the main regressions and the results are reported in Appendix 2.6. Columns 1-5 present the FE results of the over-extrapolation models for the determinants of expectation errors controlling for *num_wave*. I find that *num_wave* is only significant in the regressions for the probability of being extremely optimistic. Columns 6-10 report the FE results of the job insecurity models for the determinants of expectation errors controlling for *num_wave.* The coefficients associated with *num_wave* are only significant in the *error1* and *mopt* regressions. Among all models, I find that adding the *num_wave* term does not change the main effects of past income changes/job insecurity on individuals' errors in financial expectations.

Next, I construct a dummy (*next_wave*) equal to 1 if the respondent can be tracked in the following survey round, and 0 otherwise. I include this variable as an additional control and re-estimate all main models. Columns 1-5 of Appendix 2.7 presents the FE results of the over-extrapolation models for the determinants of expectation errors controlling for *next_wave*. The coefficients on the *next_wave* term are significant in the *error1, eopt* and *epes* regressions. In columns 6-10, the FE results of the job insecurity models for the determinants of expectation

errors controlling for *next_wave* are reported. I find that the *next_wave* terms are only significant in the *error1* and *eopt* regressions. For all models, despite the significance of the *next_wave* term, the coefficients associated with past income changes/job insecurity and other controls are still significant and of the same expected signs as in findings of my main results (section 3.6). Thus, by employing these two addition variable tests, I find that attrition bias does not seem to pose sever biases and distort the empirical results found in previous models.

2.8 Conclusion

This paper investigates the determinants of errors in financial expectations in Russia, with particular focus on the role of past income changes and job insecurity. To this end, I have used panel data on more than 25,000 individuals over the period 2002-2016.

Focusing on the role of past income changes, my empirical analysis provides support for the over-extrapolation of past income changes in determining forecast errors. More precisely, I document that after experiencing an improvement in their financial situation over the past year, individuals are more likely to over-extrapolate the positive income experience and make optimistic forecast errors (Massenot and Pettinicchi, 2019). Individuals also tend to become over-pessimistic regarding their future finances following a negative change in their financial situation. Furthermore, I observe that there is an asymmetry between the effects of positive and negative income changes on forecast errors. First, individuals whose financial situation has greatly improved (deteriorated) make larger optimistic (pessimistic) forecast errors than those with a slight improvement (deterioration). Second, individuals make larger forecast errors following a positive income change than a negative one, which suggests that they over-extrapolate positive experiences more than negative ones.

Moving to the role of job insecurity, it is also found to be important in determining expectation errors. Specifically, I find that respondents who feel secure at finding a new job are prone to make optimistic forecast errors and display extreme and moderate optimism, while those who are insecure are less likely to display extreme optimism. In addition, I find that the more confident an individual is about his/her ability to find a new job, the higher his/her probability of

showing optimistic forecast errors is. This confirms that individuals who are more secure in their job prospects are more likely to display imprudent financial expectations and thus more likely to make optimistic forecast errors (Bissonnette and Soest, 2015). However, I notice that the relationship between job insecurity and extreme and moderate pessimism is mostly insignificant. This could be explained by the fact that individuals who feel unsecure may use unemployment schemes as a shield against potential job volatility (Pogosov and Sokolovskaya, 2016). Thus, insecure respondents have a lower incentive to adjust their expectations and to display extreme/moderate pessimism.

In addition, I have included past income changes and job insecurity as joint determinants of expectation error. I find that both past income changes and job insecurity remain significant and substantial determinants of expectation error after pooling together. I also find that the effects of past income changes are stronger and more profound than the effects of job insecurity on expectation error. This suggests that past income changes are the more important determinant of expectation error.

Given that individuals living in developing/transitional countries face higher economic volatility and greater financial pressure, the consequences of displaying forecast errors may be devastating (no savings, more debt, growing concerns over getting necessities). Thus, the empirical findings from this paper may interest policy makers in the following three ways. First, individuals are found to be "experience-driven". For instance, they make future financial expectations and plans based on past financial experiences. Hence, policy makers can introduce financial education programmes and schemes in schools, companies and communities to educate individuals on how to make sophisticated financial planning and develop prudent financial

expectations, regardless of their experiences. Secondly, the results on job insecurity suggest that policies aimed at increasing the resources dedicated to developing professional skills aimed at finding adequate jobs can provide confidence for insecure workers. Thirdly, social unemployment benefits and insurances can play an important role in insuring workers against the uncertainties associated with finding a new job in case they are laid off. Thus, policy makers could encourage workers to participate in unemployment insurance schemes and make sure that sufficient funds are channeled towards these unemployment schemes.

One strength of this study is that it is based on a large panel data set collected over a longer period of time than most previous studies. I make use of the FE and FEIV estimators to rule out unobserved individual heterogeneity and endogeneity problems. Nevertheless, there are several limitations in this paper. Firstly, several observations were lost in the process of constructing the intertemporal variables of interest (errors in financial expectations). Next, this study only uses job confidence as a proxy for job insecurity. Yet, Dickerson and Green (2012) suggest that job insecurity consists of various other components (e.g., risk of job loss and uncertainty over job-content). Thus, more robust tests on the effect of job insecurity on expectation errors are needed. Future research could explore how different proxies of job insecurity affect individuals' expectation errors in alternative ways. Secondly, in the context of a dynamic model, future research could examine whether individuals' past forecast errors affect current forecasting.

Table 2.1 Structure of the unbalanced panel

Year	Frequency	Percent	Cumulative (%)
2002	5,697	4.97	4.97
2003	6,085	5.3	10.27
2004	5,831	5.08	15.35
2005	6,372	5.56	20.91
2006	5,678	4.95	25.86
2007	6,884	6	31.86
2008	6,927	6.04	37.9
2009	6,606	5.76	43.66
2010	6,923	6.04	49.69
2011	10,129	8.83	58.52
2012	10,168	8.86	67.39
2013	10,529	9.18	76.57
2014	8,907	7.77	84.33
2015	8,797	7.67	92
2016	9,174	8	100
Total	114,707	100	

Notes: This table presents the structure of the unbalanced panel used in this chapter.

Table 2.2 Distribution and classification of expectation errors

Error1	Frequency	Percent	Cumulative	Classification	*Error2*
-4	28	0.02	0.02		
-3	488	0.43	0.45	Extreme pessimism *(epes)*	*1*
-2	4,041	3.52	3.97		
-1	19,085	16.64	20.61	Moderate pessimism (*mpes)*	*2*
0	56,192	48.99	69.6	Correct expectation (*correct*)	*3*
1	25,546	22.27	91.87	Moderate optimism (*mopt)*	*4*
2	7,662	6.68	98.55		
3	1,497	1.31	99.85	Extreme optimism (*eopt*)	*5*
4	168	0.15	100		

Notes: Columns 1-4 present the distribution of the expectation error calculated by taking the difference between individuals' financial expectations and realisations. Column 5 reports the classification of the four dependent variables and column 6 presents the classification of *Error2*. Definitions of all variables are shown in Table 2.6.

Table 2.3 Distribution of past income changes

Category	Frequency	Percent	Cumulative	Classification	Abbreviation
1	6,330	5.52	5.52	greatly worsened	*gworsened*
2	15,907	13.87	19.39	slightly worsened	*sworsened*
3	63,790	55.61	75	no change	*nochange*
4	26,408	23.02	98.02	slightly improved	*simproved*
5	2,272	1.98	100	greatly improved	*gimproved*

Notes: Columns 1-4 present the distribution of past income changes. Column 5 reports the classification of the four independent variables related to past income changes. Definitions of all variables are shown in Table 2.6.

Table 2.4 Expectation errors conditional on past income changes

	Error2 (t)					
finreal (t-1)	1	2	3	4	5	Total
1	1,190	1,172	2,150	1,007	811	6,330
	18.8%	18.52%	33.97%	15.91%	12.81%	100%
2	1,149	3,505	6,196	3,564	1,493	15,907
	7.22%	22.03%	38.95%	22.41%	9.39%	100%
3	1,735	10,663	35,634	11,860	3,898	63,790
	2.72%	16.72%	55.86%	18.59%	6.11%	100%
4	428	3,488	11,556	8,301	2,635	26,408
	1.62%	13.21%	43.76%	31.43%	9.98%	100%
5	55	257	656	814	490	2,272
	2.42%	11.31%	28.87%	35.83%	21.57%	100%
Total	4,557	19,085	56,192	25,546	9,327	114,707
	3.97%	16.64%	48.99%	22.27%	8.13%	100%

Notes: This table reports the cross-tabulation between *finreal* and *error2*. Each row presents the frequency and the row cumulative contains the percentage of the corresponding *finreal* answer. Definitions of all variables are shown in Table 2.6.

Table 2.5 Distribution of job insecurity

Category	Frequency	Percent	Cumulative	Classification	Abbreviation
1	11,776	19.54	19.54	highly insecure	*hinsecure*
2	14,478	24.02	43.55	relatively insecure	*rinsecure*
3	10,323	17.13	60.68	both yes and no	*indifferent*
4	15,150	25.13	85.81	relatively secure	*rsecure*
5	8,553	14.19	100	highly secure	*hsecure*

Notes: Columns 1-4 present the distribution of job insecurity. Column 5 reports the classification of the four independent variables related to job insecurity. Definitions of all variables are shown in Table 2.6.

Table 2.6 Summary statistics

Variable	Description	Type	Obs	Mean	Std. Dev.
Dependent variables					
error1	Individuals' expectation error	Continuous	114,707	0.151	0.979
eopt	Extreme optimism [1]; other [0]	Dummy	114,707	0.081	0.273
mopt	Moderate optimism [1]; other [0]	Dummy	114,707	0.223	0.416
mpes	Moderate pessimism [1]; other [0]	Dummy	114,707	0.166	0.372
epes	Extreme pessimism [1]; other [0]	Dummy	114,707	0.040	0.195
Independent variables					
Past income changes:					
gimproved	Greatly improved financial situation [1]; other [0]	Dummy	114,707	0.020	0.139
simproved	Slightly improved financial situation [1]; other [0]	Dummy	114,707	0.230	0.421
sworsened	Slightly worsened financial situation [1]; other [0]	Dummy	114,707	0.139	0.346
gworsened	Greatly worsened financial situation [1]; other [0]	Dummy	114,707	0.055	0.228
Job insecurity:					
hsecure	Highly secure [1]; other [0]	Dummy	60,280	0.141	0.348
rsecure	Relatively secure [1]; other [0]	Dummy	60,280	0.251	0.433
rinsecure	Relatively insecure [1]; other [0]	Dummy	60,280	0.240	0.427
hinsecure	Highly insecure [1]; other [0]	Dummy	60,280	0.195	0.396

Control variables					
male	Male [1]; female [0]	Dummy	114,707	0.409	0.492
age1830	Respondent aged 18 to 30 [1]; other [0]	Dummy	114,707	0.222	0.416
age3040	Respondent aged 31 to 40 [1]; other [0]	Dummy	114,707	0.190	0.392
age4050	Respondent aged 41 to 50 [1]; other [0]	Dummy	114,707	0.172	0.378
age5065	Respondent aged 51 to 65 [1]; other [0]	Dummy	114,707	0.238	0.426
age65	Respondent aged above 65 [1]; other [0]	Dummy	114,707	0.177	0.381
hhincome	Household income in the previous month	Continuous	114,707	44,636.6	34,649.9
consumption	Household consumption in the previous month	Continuous	114,707	31,572.9	26,832.8
saving	Have saving [1]; other [0]	Dummy	114,707	0.151	0.358
hhsize	Household size	Continuous	114,707	3.249	1.577
cohabit	Married or living with a partner [1]; other [0]	Dummy	114,707	0.661	0.473
nmarried	Never married [1]; other [0]	Dummy	114,707	0.132	0.338
edup	Highest degree is primary education [1]; other [0]	Dummy	114,707	0.042	0.200
eduu	Highest degree is unfinished secondary education [1]; other [0]	Dummy	114,707	0.147	0.354
edus	Highest degree is secondary education [1]; other [0]	Dummy	114,707	0.330	0.470
urban	Living in urban area [1]; other [0]	Dummy	114,707	0.728	0.445
employed	Employed (including self-employed) [1]; other [0]	Dummy	114,707	0.565	0.496
retired	Retired [1]; other [0]	Dummy	114,707	0.252	0.434
hstatus	Self-reported health status	Ordinal	114,707	3.161	0.721

Notes: This table presents the definitions and summary statistics of all variables used. The number of observations for job insecurity is lower because it restricts the sample to employed population only. Income and consumption are in log form (*lghhincome* and *lgconsumption*) in all the regressions, and inflated into the real value of 2016 using the CPI from the Rosstat. Income is the total sum of money actually received by the individual in the last month, which includes salaries and wages from all jobs, incidental earnings, public and private transfers, retirement pensions and other receipts, including foreign currency (convert the value in rubles). Household consumption includes all goods and services consumed in the last 3 months, which includes expenditures on food at home or eating out, household appliances, entertainment supplies, fuel, services (postal, transportation, repair and internet services etc.), residential rent and utilities, medical treatment and private transfers. The top 1% of household income, consumption and household size are also dropped to avoid biased estimates due to extreme outliers.

Table 2.7 Sample means of errors in financial expectations by past income changes and job insecurity

Panel A: Sample means by past income changes									
Past income	*gimproved*		*simproved*		*nochange*	*sworsened*		*gworsened*	
	Mean	Diff1	Mean	Diff2		Mean	Diff3	Mean	Diff4
error1	0.684	0.592***	0.369	0.276***	0.093	0.06	-0.033***	-0.139	-0.232***
eopt	0.216	0.155***	0.1	0.039***	0.061	0.094	0.033***	0.128	0.067***
mopt	0.358	0.172***	0.314	0.128***	0.186	0.224	0.038***	0.159	-0.027***
mpes	0.113	-0.054***	0.132	-0.035***	0.167	0.22	0.053***	0.185	0.018***
epes	0.024	-0.003	0.016	-0.011***	0.027	0.072	0.045***	0.188	0.161***
N=	2,272		26,408		63,790	15,907		6,330	
Panel B: Sample means by job insecurity									
Job insecurity	*hsecure*		*rsecure*		*indifferent*	*rinsecure*		*hinsecure*	
	Mean	Diff5	Mean	Diff6	Mean	Mean	Diff7	Mean	Diff8
error1	0.374	0.203***	0.264	0.094***	0.17	0.139	-0.031**	0.03	-0.141***
eopt	0.141	0.063***	0.094	0.016***	0.078	0.076	-0.002	0.069	-0.009**
mopt	0.268	0.032***	0.264	0.028***	0.236	0.23	-0.006	0.2	-0.036***
mpes	0.14	-0.027***	0.155	-0.012**	0.167	0.182	0.015***	0.19	0.023***
epes	0.033	-0.001	0.025	-0.009***	0.034	0.035	0.001	0.06	0.026***
N=	8,553		15,150		10,323	14,478		11,776	

Note: Diff1/Diff2/Diff3/Diff4 reports differences in the equality of means between those who experienced *gimproved/ simproved/ sworsened/ gworsenedand* and those who experienced *nochange* in financial situation. Diff5/Diff6/Diff7/Diff8 reports differences in the equality of means between those who are *hsecure/ rsecure/ rinsecure/ hinsecure* and those who are *indifferent*. Statistical significance is indicated by * at the 10%, ** at the 5%, *** at the 1% level. Definitions of all variables are shown in Table 2.6.

Table 2.8 Correlation Matrix

	gimprove	*simproved*	*gworsened*	*sworsened*	*hsecure*	*rsecure*	*hinsecure*	*rsecure*	*male*	*age1830*	*age3040*	*age4050*	*age5065*
gimproved	1												
simproved	-0.0957	1											
gworsened	-0.0321	-0.1246	1										
sworsened	-0.061	-0.2371	-0.0795	1									
hsecure	0.0665	0.0342	0.0066	-0.0192	1								
rsecure	-0.0034	0.0524	-0.0441	-0.0202	-0.2356	1							
hinsecure	-0.0143	-0.0468	0.0663	0.0234	-0.2004	-0.2855	1						
rinsecure	-0.0236	-0.0205	-0.0165	0.0107	-0.2286	-0.3258	-0.277	1					
male	0.0035	0.0092	-0.0368	-0.0113	0.0241	0.0599	-0.0803	-0.0163	1				
age1830	0.0594	0.1053	-0.0357	-0.0399	0.1104	0.0804	-0.1296	-0.0638	0.0301	1			
age3040	0.0043	0.0056	0.0023	0.0112	0.0318	0.0512	-0.0861	-0.0166	0.007	-0.3443	1		
age4050	-0.0259	-0.0427	0.0175	0.0257	-0.0398	-0.025	0.0102	0.039	-0.0254	-0.3209	-0.3377	1	
age5065	-0.0332	-0.0606	0.0155	0.0025	-0.0979	-0.0974	0.1791	0.0479	-0.0158	-0.3088	-0.325	-0.303	1
hhincome	0.065	0.1059	-0.076	-0.0556	0.0636	0.1035	-0.1178	-0.0577	0.0549	0.0342	0.0093	-0.0495	0.0064
consumption	0.0498	0.078	-0.046	-0.0185	0.0887	0.104	-0.1273	-0.0657	0.0209	0.0335	0.0867	-0.0064	-0.0969
saving	0.0578	0.0886	-0.0507	-0.0546	0.0358	0.0026	0.0041	-0.0171	0.0013	-0.0069	-0.0218	-0.0313	0.0458
hhsize	-0.0006	0.0311	-0.0018	0.0132	0.0028	0.0416	-0.0508	-0.0049	0.0744	0.0646	0.1227	0.0176	-0.1793
cohabit	0.0118	0.0089	-0.0135	0.0058	-0.0045	-0.0086	-0.0009	0.0052	0.1817	-0.1432	0.0964	0.07	-0.0103
nmarried	0.0072	0.0353	-0.0217	-0.0262	0.0348	0.0494	-0.0691	-0.0221	0.0132	0.3698	-0.0692	-0.1374	-0.1512
edup	-0.001	-0.0061	0.0004	-0.0019	-0.0053	-0.0096	0.0196	0.0011	0.0309	-0.0012	-0.021	-0.0243	0.0223
eduu	-0.0033	-0.0059	0.0091	-0.0206	-0.0126	-0.0097	0.0166	0.009	0.1027	0.0368	0.0045	-0.0449	-0.002

edus	-0.0097	-0.0203	0.0061	-0.0068	0.0009	-0.0101	-0.009	0.0142	0.1442	-0.0165	-0.0231	0.0513	0.0074
urban	0.0014	-0.0054	0.018	0.0267	0.0947	0.0969	-0.1415	-0.0613	-0.0174	0.0146	0.0046	-0.0261	-0.006
employed	0.0016	0.0217	-0.0338	-0.0507	-0.058	-0.0155	0.044	0.0309	0.1608	-0.1845	-0.0281	0.1015	0.106
retired	-0.0009	-0.0035	-0.0012	-0.0022	0.0059	-0.0033	-0.0028	0.0035	0.0006	-0.0033	-0.0035	-0.0032	0.0107
hstatus	0.0495	0.0971	-0.0768	-0.0638	0.0687	0.1052	-0.1318	-0.0527	0.1305	0.2328	0.108	-0.0859	-0.2341

Notes: This table presents the correlation matrix of all independent variables. The correlation coefficient is on bold if the coefficient is larger than 0.5. Definition of all variables are shown in Table 2.6.

Table 2.8 Correlation matrix (continued)

	hhincome	*consumption*	*saving*	*hhsize*	*cohabit*	*nmarried*	*edup*	*eduu*	*edus*	*urban*	*employed*	*retired*	*hstatus*
hhincome	1												
consumption	**0.5938**	1											
saving	0.1379	0.0891	1										
hhsize	0.2915	0.2944	-0.0428	1									
cohabit	0.1086	0.1164	0.0036	0.2103	1								
nmarried	0.0119	-0.0189	-0.0012	-0.0371	-0.6258	1							
edup	-0.0451	-0.0501	-0.0049	-0.0187	-0.0003	-0.0073	1						
eduu	-0.0774	-0.0796	-0.0296	0.0447	-0.0057	0.0196	-0.0161	1					
edus	-0.1263	-0.102	-0.0387	0.0494	0.0323	-0.0201	-0.0343	-0.2298	1				
urban	0.199	0.1897	0.0404	-0.1047	-0.0473	0.0258	-0.0196	-0.1324	-0.072	1			
employed	-0.0012	-0.0241	0.0061	-0.0884	-0.0492	-0.0056	0.0098	0.0182	0.0345	-0.0163	1		
retired	0.0035	0.0029	-0.0025	-0.0015	-0.0033	-0.0021	-0.0003	-0.0019	-0.004	0.0031	-0.0288	1	
hstatus	0.0908	0.0698	0.0033	0.1125	0.0118	0.1081	-0.0164	0.0023	-0.0185	0.0062	-0.0583	-0.0081	1

Table 2.9 FE and FEIV results of past income changes

Variable	*error1*		*eopt*		*mopt*		*mpes*		*epes*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	FE	FEIV	FE	FEIV	FE	FEIV	FE	FEIV	FE	FEIV
gimproved	0.783***	2.474***	0.132***	0.531***	0.210***	0.647***	-0.152***	-0.578***	-0.051***	-0.000
	(0.028)	(0.258)	(0.010)	(0.081)	(0.012)	(0.109)	(0.009)	(0.089)	(0.005)	(0.043)
simproved	0.383***	0.322***	0.037***	0.009	0.149***	0.210***	-0.111***	-0.097***	-0.018***	-0.002
	(0.009)	(0.067)	(0.002)	(0.019)	(0.004)	(0.030)	(0.003)	(0.028)	(0.001)	(0.013)
sworsened	-0.213***	-0.278*	-0.018***	0.014	-0.024***	0.029	0.050***	0.155**	0.050***	0.081**
	(0.011)	(0.158)	(0.003)	(0.044)	(0.004)	(0.067)	(0.004)	(0.063)	(0.003)	(0.034)
gworsened	-0.611***	-1.045***	-0.075***	-0.026	-0.042***	-0.155***	0.048***	0.194***	0.161***	0.295***
	(0.021)	(0.156)	(0.005)	(0.041)	(0.007)	(0.054)	(0.006)	(0.054)	(0.006)	(0.039)
age1830	-0.008	-0.030	0.020	0.018	-0.041**	-0.047**	0.028	0.037*	-0.006	-0.005
	(0.047)	(0.050)	(0.013)	(0.013)	(0.020)	(0.020)	(0.018)	(0.019)	(0.009)	(0.010)
age3040	-0.023	-0.025	0.021*	0.020*	-0.038**	-0.038**	0.027*	0.028*	0.001	0.001
	(0.039)	(0.041)	(0.011)	(0.011)	(0.016)	(0.016)	(0.016)	(0.016)	(0.008)	(0.008)
age4050	-0.043	-0.031	0.008	0.008	-0.021*	-0.019	0.036***	0.032**	0.001	-0.001
	(0.030)	(0.032)	(0.008)	(0.008)	(0.012)	(0.013)	(0.012)	(0.013)	(0.007)	(0.007)
age5065	-0.032	-0.021	0.003	0.003	-0.002	0.001	0.028***	0.024***	0.003	0.001
	(0.020)	(0.022)	(0.005)	(0.005)	(0.008)	(0.009)	(0.009)	(0.009)	(0.005)	(0.005)
hhincome	0.018**	-0.028**	0.008***	0.007**	-0.003	-0.018***	-0.000	0.018***	-0.003*	0.002
	(0.008)	(0.013)	(0.002)	(0.004)	(0.003)	(0.005)	(0.003)	(0.005)	(0.002)	(0.003)

consumption	0.000	-0.010	-0.002	-0.002	0.004	-0.000	-0.000	0.003	0.000	0.002
	(0.006)	(0.007)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)
saving	0.012	-0.015	0.005**	0.004	-0.002	-0.011**	-0.005	0.006	0.002	0.004*
	(0.010)	(0.012)	(0.003)	(0.003)	(0.004)	(0.005)	(0.004)	(0.005)	(0.002)	(0.002)
hhsize	0.005	0.017***	-0.001	-0.000	0.004**	0.008***	-0.003	-0.008***	-0.000	-0.001
	(0.005)	(0.006)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
cohabit	0.093***	0.078***	0.027***	0.027***	0.015*	0.011	-0.012*	-0.006	-0.002	0.001
	(0.019)	(0.020)	(0.005)	(0.005)	(0.008)	(0.008)	(0.007)	(0.007)	(0.004)	(0.004)
nmarried	-0.011	-0.010	-0.005	-0.000	-0.005	-0.001	0.003	0.006	-0.003	0.001
	(0.028)	(0.030)	(0.008)	(0.009)	(0.012)	(0.012)	(0.011)	(0.011)	(0.006)	(0.006)
edup	0.089	0.079	-0.008	-0.008	0.028	0.034	-0.040*	-0.031	-0.022*	-0.020
	(0.057)	(0.059)	(0.015)	(0.015)	(0.023)	(0.024)	(0.024)	(0.025)	(0.013)	(0.013)
eduu	-0.024	-0.028	-0.008	-0.006	0.007	0.007	0.004	0.007	0.001	0.003
	(0.023)	(0.025)	(0.007)	(0.007)	(0.010)	(0.011)	(0.009)	(0.009)	(0.004)	(0.004)
edus	-0.015	-0.019	-0.012**	-0.011**	0.008	0.008	0.005	0.007	-0.004	-0.002
	(0.018)	(0.019)	(0.005)	(0.005)	(0.008)	(0.008)	(0.007)	(0.007)	(0.003)	(0.003)
urban	0.088	0.118	0.008	0.014	0.116***	0.127***	-0.029	-0.038	0.039*	0.041**
	(0.084)	(0.089)	(0.022)	(0.023)	(0.035)	(0.036)	(0.036)	(0.037)	(0.020)	(0.019)
employed	-0.065***	-0.110***	-0.017***	-0.012**	-0.002	-0.018**	0.003	0.024***	0.010***	0.019***
	(0.013)	(0.019)	(0.004)	(0.005)	(0.005)	(0.007)	(0.005)	(0.007)	(0.002)	(0.004)
retired	-0.064***	-0.093***	-0.021***	-0.021***	-0.014*	-0.025***	-0.007	0.002	0.005	0.009**
	(0.019)	(0.022)	(0.005)	(0.006)	(0.008)	(0.009)	(0.008)	(0.008)	(0.004)	(0.005)
hstatus	0.041***	0.026***	0.006***	0.006***	0.010***	0.006*	-0.012***	-0.006*	-0.003*	-0.000

	(0.007)	(0.008)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)	(0.002)
DIFF1 (P value)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.969
DIFF2 (P value)	0.000***	0.003***	0.000***	0.571	0.013**	0.069*	0.798	0.688	0.000***	0.000***
DIFF3 (P value)	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
DIFF4 (P value)	0.000***	0.000***	0.000***	0.195	0.000***	0.003***	0.000***	0.003***	0.000***	0.068*
Observations	114,707	109,012	114,707	109,012	114,707	109,012	114,707	109,012	114,707	109,012
R-squared	0.069	-0.000	0.016	-0.032	0.027	-0.000	0.023	-0.015	0.041	0.019
Number of idind	25,837	20,186	25,837	20,186	25,837	20,186	25,837	20,186	25,837	20,186

Notes: This table reports the FE and FEIV of the determinants of expectation errors indicators. DIFF1 is the Wald test performed to test whether the coefficient associated with *gimproved* is significantly different from that on *simproved*. DIFF2 is the Wald test performed to test whether the coefficient associated with *gworsened* is significantly different from that on *sworsened*. DIFF3 is the Wald test performed to test whether the coefficient associated with *gimproved* is significantly different from that on *gworsened*. DIFF4 is the Wald test performed to test whether the coefficient associated with *simproved* is significantly different from that on *sworsened*. Regional and Year dummies are included in all models, but their coefficients are not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Coefficients are presented, and robust standard errors are given in parentheses. A full tests of instrument validity of the FEIV models are presented in Appendix 2.1. Definitions of all variables are shown in Table 2.6.

Table 2.10 FE and FEIV results of job insecurity

Variable	*error1*		*eopt*		*mopt*		*mpes*		*epes*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	FE	FEIV	FE	FEIV	FE	FEIV	FE	FEIV	FE	FEIV
hsecure	0.124***	0.336*	0.035***	0.113**	0.019**	-0.011	-0.028***	-0.133**	-0.002	0.050
	(0.020)	(0.188)	(0.006)	(0.055)	(0.009)	(0.077)	(0.007)	(0.068)	(0.004)	(0.039)
rsecure	0.047***	-0.187	0.009**	-0.078	0.012*	-0.118	-0.007	0.015	-0.005*	-0.023
	(0.015)	(0.200)	(0.004)	(0.060)	(0.007)	(0.092)	(0.006)	(0.079)	(0.003)	(0.039)
rinsecure	0.005	0.103	0.004	0.035	0.010	-0.007	0.006	-0.004	0.003	0.024
	(0.015)	(0.235)	(0.004)	(0.067)	(0.006)	(0.109)	(0.006)	(0.098)	(0.003)	(0.048)
hinsecure	-0.042**	-0.291	0.001	-0.124**	0.001	-0.030	0.002	0.054	0.018***	-0.006
	(0.018)	(0.190)	(0.005)	(0.054)	(0.007)	(0.086)	(0.007)	(0.076)	(0.004)	(0.038)
age1830	-0.027	-0.025	-0.002	-0.001	-0.007	-0.006	0.061*	0.062*	-0.024	-0.025
	(0.079)	(0.080)	(0.023)	(0.023)	(0.033)	(0.033)	(0.033)	(0.033)	(0.017)	(0.017)
age3040	-0.070	-0.071	-0.006	-0.007	-0.017	-0.015	0.069**	0.070**	-0.016	-0.017
	(0.071)	(0.072)	(0.020)	(0.021)	(0.029)	(0.030)	(0.030)	(0.030)	(0.016)	(0.016)
age4050	-0.097	-0.102	-0.022	-0.026	-0.011	-0.009	0.070**	0.071***	-0.016	-0.017
	(0.063)	(0.064)	(0.018)	(0.019)	(0.026)	(0.026)	(0.027)	(0.027)	(0.015)	(0.015)
age5065	-0.081	-0.085	-0.025	-0.027	0.001	0.003	0.061**	0.062**	-0.016	-0.017
	(0.056)	(0.057)	(0.016)	(0.017)	(0.023)	(0.023)	(0.025)	(0.025)	(0.013)	(0.013)
hhincome	0.122***	0.123***	0.019***	0.019***	0.022***	0.022***	-0.023***	-0.023***	-0.018***	-0.017***
	(0.012)	(0.012)	(0.003)	(0.003)	(0.005)	(0.005)	(0.004)	(0.005)	(0.002)	(0.002)
consumption	0.022**	0.023**	0.002	0.002	0.010**	0.011***	-0.004	-0.004	-0.000	-0.001

	(0.009)	(0.009)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.002)	(0.002)
saving	0.075***	0.068***	0.014***	0.012***	0.010	0.008	-0.016***	-0.014**	-0.008***	-0.009***
	(0.014)	(0.015)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)	(0.003)	(0.003)
hhsize	-0.021***	-0.019***	-0.005**	-0.004**	0.001	0.001	0.001	0.001	0.003***	0.004***
	(0.007)	(0.007)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)	(0.001)
cohabit	0.128***	0.135***	0.036***	0.039***	0.025**	0.025**	-0.016	-0.018	-0.004	-0.003
	(0.027)	(0.028)	(0.008)	(0.008)	(0.012)	(0.012)	(0.011)	(0.011)	(0.006)	(0.006)
nmarried	-0.002	0.008	-0.001	0.004	-0.001	0.000	0.008	0.007	-0.004	-0.002
	(0.040)	(0.041)	(0.012)	(0.012)	(0.017)	(0.018)	(0.015)	(0.016)	(0.008)	(0.008)
edup	0.140	0.130	-0.016	-0.023	0.073	0.078	-0.137	-0.139	-0.002	-0.002
	(0.237)	(0.229)	(0.056)	(0.058)	(0.083)	(0.082)	(0.090)	(0.089)	(0.058)	(0.060)
eduu	-0.043	-0.035	-0.019*	-0.016	0.014	0.016	0.005	0.002	0.001	0.003
	(0.037)	(0.038)	(0.011)	(0.011)	(0.016)	(0.016)	(0.014)	(0.014)	(0.007)	(0.007)
edus	-0.032	-0.029	-0.022***	-0.021**	0.021*	0.024**	0.001	-0.000	-0.000	0.001
	(0.027)	(0.028)	(0.008)	(0.009)	(0.012)	(0.012)	(0.010)	(0.011)	(0.005)	(0.006)
urban	0.109	0.130	-0.026	-0.016	0.173***	0.165***	-0.012	-0.019	0.006	0.010
	(0.125)	(0.128)	(0.033)	(0.036)	(0.051)	(0.051)	(0.053)	(0.055)	(0.025)	(0.027)
employed	-0.052	-0.044	-0.026***	-0.024**	0.003	0.006	-0.001	-0.003	-0.004	-0.003
	(0.032)	(0.033)	(0.010)	(0.010)	(0.014)	(0.014)	(0.011)	(0.011)	(0.005)	(0.005)
retired	0.796***	0.608*	-0.016	-0.087**	0.472	0.454	-0.370***	-0.313***	0.003	-0.029
	(0.298)	(0.337)	(0.017)	(0.036)	(0.354)	(0.354)	(0.089)	(0.062)	(0.007)	(0.021)
hstatus	0.046***	0.047***	0.007**	0.007*	0.012***	0.014***	-0.009**	-0.008**	-0.005**	-0.005**
	(0.011)	(0.011)	(0.003)	(0.003)	(0.005)	(0.005)	(0.004)	(0.004)	(0.002)	(0.002)
DIFF1 (P value)	0.000***	0.012**	0.000***	0.002***	0.387	0.238	0.000***	0.058*	0.271	0.079*

DIFF2 (P value)	0.002***	0.053*	0.476	0.006***	0.188	0.801	0.567	0.500	0.000***	0.482
DIFF3 (P value)	0.000***	0.003***	0.000***	0.000***	0.054*	0.840	0.000***	0.024**	0.000***	0.201
DIFF4 (P value)	0.000***	0.164	0.237	0.069*	0.718	0.234	0.038**	0.821	0.007***	0.260
Observations	60,280	55,410	60,280	55,410	60,280	55,410	60,280	55,410	60,280	55,410
R-squared	0.013	-0.017	0.008	-0.051	0.003	-0.008	0.004	-0.007	0.005	-0.011
Number of idind	16,515	11,732	16,515	11,732	16,515	11,732	16,515	11,732	16,515	11,732

Notes: This table reports the FE and FEIV of the determinants of expectation errors indicators. DIFF1 is the Wald test performed to test whether the coefficient on *hsecure* is significantly different from that on *rsecure*. DIFF2 is the Wald test performed to test whether the coefficient on *hinsecure* is significantly different from that on *rinsecure*. DIFF3 is the Wald test performed to test whether the coefficient on *hsecure* is significantly different from that on *hinsecure*. DIFF4 is the Wald test performed to test whether the coefficient on *rsecure* is significantly different from that on *rinsecure*. Regional and Year dummies are included in all models, but their coefficients are not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Coefficients are presented, and robust standard errors are given in parentheses. A full tests of instrument validity of the FEIV models are presented in Appendix 2.2. Definitions of all variables are shown in Table 2.6.

Table 2.11 FE and FEIV results of past income changes and job insecurity

Variable	*error1*		*eopt*		*mopt*		*mpes*		*epes*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	FE	FEIV	FE	FEIV	FE	FEIV	FE	FEIV	FE	FEIV
gimproved	0.792***	1.903***	0.128***	0.248*	0.208***	0.736***	-0.161***	-0.433***	-0.059***	-0.004
	(0.037)	(0.445)	(0.012)	(0.126)	(0.016)	(0.205)	(0.012)	(0.162)	(0.006)	(0.084)
simproved	0.382***	0.411***	0.042***	0.093**	0.136***	0.161***	-0.112***	-0.082	-0.019***	0.003
	(0.012)	(0.131)	(0.003)	(0.038)	(0.005)	(0.061)	(0.005)	(0.052)	(0.002)	(0.025)
sworsened	-0.188***	-0.523**	-0.021***	-0.053	-0.018***	0.108	0.034***	0.175*	0.046***	0.141***
	(0.016)	(0.254)	(0.004)	(0.073)	(0.007)	(0.110)	(0.006)	(0.099)	(0.004)	(0.053)
gworsened	-0.553***	-0.641**	-0.081***	0.071	-0.028***	-0.205*	0.035***	0.092	0.139***	0.258***
	(0.034)	(0.300)	(0.009)	(0.087)	(0.011)	(0.120)	(0.010)	(0.106)	(0.009)	(0.069)
hsecure	0.101***	0.285	0.032***	0.080	0.012	-0.041	-0.023***	-0.117	0.000	0.016
	(0.020)	(0.194)	(0.006)	(0.058)	(0.009)	(0.085)	(0.007)	(0.072)	(0.004)	(0.040)
rsecure	0.033**	-0.230	0.008*	-0.103*	0.009	-0.146	-0.003	0.023	-0.003	-0.043
	(0.015)	(0.211)	(0.004)	(0.062)	(0.007)	(0.098)	(0.006)	(0.083)	(0.003)	(0.040)
rinsecure	0.000	0.052	0.003	-0.004	0.008	-0.044	0.007	0.007	0.003	-0.011
	(0.014)	(0.248)	(0.004)	(0.071)	(0.006)	(0.118)	(0.006)	(0.105)	(0.003)	(0.051)
hinsecure	-0.037**	-0.250	0.002	-0.151**	-0.001	-0.004	0.004	0.054	0.015***	-0.041
	(0.018)	(0.209)	(0.005)	(0.059)	(0.007)	(0.095)	(0.007)	(0.082)	(0.004)	(0.041)
age1830	-0.017	-0.014	-0.001	0.001	-0.005	-0.003	0.059*	0.060*	-0.025	-0.027
	(0.080)	(0.084)	(0.023)	(0.023)	(0.033)	(0.034)	(0.033)	(0.034)	(0.017)	(0.017)
age3040	-0.050	-0.036	-0.003	-0.002	-0.013	-0.008	0.065**	0.062**	-0.019	-0.022

	(0.071)	(0.074)	(0.020)	(0.021)	(0.030)	(0.030)	(0.030)	(0.031)	(0.016)	(0.016)
age4050	-0.076	-0.059	-0.020	-0.022	-0.007	0.001	0.066**	0.061**	-0.019	-0.023
	(0.063)	(0.065)	(0.018)	(0.018)	(0.026)	(0.027)	(0.027)	(0.028)	(0.014)	(0.015)
age5065	-0.078	-0.074	-0.024	-0.028*	0.000	0.005	0.062**	0.060**	-0.018	-0.020
	(0.056)	(0.057)	(0.016)	(0.016)	(0.023)	(0.023)	(0.025)	(0.025)	(0.013)	(0.013)
hhincome	0.023**	-0.033	0.006*	0.001	-0.002	-0.014	-0.002	0.013	-0.005**	0.002
	(0.012)	(0.021)	(0.003)	(0.006)	(0.005)	(0.009)	(0.005)	(0.008)	(0.002)	(0.004)
consumption	-0.001	-0.011	-0.001	-0.001	0.004	0.001	0.000	0.003	0.003	0.004*
	(0.009)	(0.010)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)	(0.002)	(0.002)
saving	0.023	-0.015	0.007*	0.001	-0.004	-0.011	-0.004	0.005	-0.002	0.000
	(0.014)	(0.018)	(0.004)	(0.005)	(0.006)	(0.008)	(0.006)	(0.007)	(0.003)	(0.003)
hhsize	0.003	0.020**	-0.002	-0.000	0.006**	0.010**	-0.003	-0.008***	0.000	-0.002
	(0.007)	(0.009)	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.003)	(0.001)	(0.002)
cohabit	0.107***	0.100***	0.033***	0.035***	0.021*	0.015	-0.012	-0.010	-0.002	0.000
	(0.027)	(0.029)	(0.008)	(0.008)	(0.012)	(0.012)	(0.011)	(0.011)	(0.006)	(0.006)
nmarried	-0.007	0.006	-0.001	0.005	-0.001	0.008	0.008	0.008	-0.002	0.003
	(0.040)	(0.044)	(0.012)	(0.013)	(0.017)	(0.018)	(0.015)	(0.016)	(0.008)	(0.008)
edup	0.066	0.001	-0.024	-0.035	0.062	0.098	-0.123	-0.099	0.013	0.035
	(0.210)	(0.210)	(0.055)	(0.060)	(0.077)	(0.081)	(0.089)	(0.087)	(0.055)	(0.056)
eduu	-0.042	-0.038	-0.019*	-0.016	0.015	0.019	0.005	0.003	0.002	0.004
	(0.037)	(0.040)	(0.011)	(0.011)	(0.016)	(0.017)	(0.014)	(0.015)	(0.007)	(0.007)
edus	-0.033	-0.028	-0.022***	-0.020**	0.022*	0.028**	0.001	0.001	0.001	0.003
	(0.028)	(0.030)	(0.008)	(0.009)	(0.012)	(0.013)	(0.010)	(0.011)	(0.005)	(0.006)
urban	0.187	0.253**	-0.016	0.005	0.199***	0.203***	-0.033	-0.048	0.002	0.008

	(0.118)	(0.125)	(0.033)	(0.036)	(0.049)	(0.050)	(0.054)	(0.057)	(0.024)	(0.026)
employed	-0.153***	-0.202***	-0.039***	-0.041***	-0.020	-0.023	0.020*	0.035***	0.010*	0.022***
	(0.032)	(0.039)	(0.010)	(0.012)	(0.014)	(0.016)	(0.011)	(0.013)	(0.005)	(0.007)
retired	0.869***	0.727*	-0.005	-0.048	0.507	0.562	-0.393***	-0.332***	0.008	0.018
	(0.311)	(0.424)	(0.020)	(0.051)	(0.357)	(0.415)	(0.089)	(0.060)	(0.007)	(0.032)
hstatus	0.024**	0.012	0.004	0.003	0.007	0.007	-0.005	-0.000	-0.002	0.000
	(0.011)	(0.012)	(0.003)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.002)	(0.002)
DIFF1 (P value)	0.000***	0.001***	0.000***	0.2390	0.000***	0.000***	0.000***	0.074*	0.000***	0.835
DIFF2 (P value)	0.000***	0.020**	0.000***	0.015**	0.000***	0.019**	0.000***	0.338	0.000***	0.356
DIFF3 (P value)	0.000***	0.108	0.000***	0.631	0.001***	0.357	0.001***	0.251	0.000***	0.044**
DIFF4 (P value)	0.000***	0.352	0.000***	0.063*	0.043**	0.257	0.012**	0.810	0.000***	0.001***
Observations	60,280	55,410	60,280	55,410	60,280	55,410	60,280	55,410	60,280	55,410
R-squared	0.068	-0.001	0.020	-0.058	0.024	-0.041	0.024	-0.014	0.032	-0.015
Number of idind	16,515	11,732	16,515	11,732	16,515	11,732	16,515	11,732	16,515	11,732

Notes: This table reports the FE and FEIV of the determinants of expectation errors indicators. DIFF1 is the Wald test performed to test whether the coefficient on *gimproved* is significantly different from that on *hsecure*. DIFF2 is the Wald test performed to test whether the coefficient on *simproved* is significantly different from that on *rsecure*. DIFF3 is the Wald test performed to test whether the coefficient on *sworsened* is significantly different from that on *rinsecure*. DIFF4 is the Wald test performed to test whether the coefficient on *gworsened* is significantly different from that on *hinsecure*. Regional and Year dummies are included in all models, but their coefficients are not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Coefficients are presented, and robust standard errors are given in parentheses. A full tests of instrument validity of the FEIV models are presented in Appendix 2.3. Definitions of all variables are shown in Table 2.6.

Table 2.12 CRE ordered Probit results

Panel A: The role of past income changes						
Variable	*Coefficient (S.E)*	*Error2=1*	*Error2=2*	*Error2=3*	*Error2=4*	*Error2=5*
	(1)	(2)	(3)	(4)	(5)	(6)
gimproved	0.902***	-0.0725***	-0.1660***	-0.0543***	0.1680***	0.1250***
	(0.0316)	(0.0026)	(0.0059)	(0.0022)	(0.0059)	(0.0044)
simproved	0.458***	-0.0369***	-0.0845***	-0.0276***	0.0853***	0.0637***
	(0.0102)	(0.0009)	(0.0019)	(0.0009)	(0.0019)	(0.0015)
sworsened	-0.682***	0.0205***	0.0470***	0.0154***	-0.0475***	-0.0355***
	(0.0237)	(0.0011)	(0.0024)	(0.0008)	(0.0024)	(0.0018)
gworsened	-0.255***	0.0549***	0.1260***	0.0411***	-0.127***	-0.0949***
	(0.0127)	(0.0021)	(0.0043)	(0.0016)	(0.0044)	(0.0033)
Observations	114,707	114,707	114,707	114,707	114,707	114,707
Panel B: The role of job insecurity						
Variable	*Coefficient (S.E)*	*Error2=1*	*Error2=2*	*Error2=3*	*Error2=4*	*Error2=5*
	(7)	(8)	(9)	(10)	(11)	(12)
hsecure	0.145***	-0.0112***	-0.0282***	-0.0109***	0.0280***	0.0224***
	(0.0219)	(0.0017)	(0.0043)	(0.0017)	(0.0042)	(0.0034)
rsecure	0.054***	-0.0043***	-0.0106***	-0.0041***	0.0106***	0.0085***
	(0.0169)	(0.0013)	(0.0033)	(0.0013)	(0.0033)	(0.0026)
rinsecure	-0.039*	-0.0007	-0.0018	-0.0007	0.0018	0.0015
	(0.0203)	(0.0013)	(0.0032)	(0.0013)	(0.0032)	(0.0026)
hinsecure	0.009	0.0030*	0.0075*	0.0029*	-0.0075*	-0.0060*
	(0.0166)	(0.0016)	(0.0039)	(0.0015)	(0.0039)	(0.0031)
Observations	60,280	60,280	60,280	60,280	60,280	60,280

Panel C: The roles of past income changes and job insecurity						
Variable	*Coefficient (S.E)*	*Error2=1*	*Error2=2*	*Error2=3*	*Error2=4*	*Error2=5*
	(13)	(14)	(15)	(16)	(17)	(18)
gimproved	0.906***	-0.0682***	-0.1710***	-0.0657***	0.1690***	0.1350***
	(0.0416)	0.0033	(0.0079)	(0.0034)	(0.0078)	(0.0063)
simproved	0.449***	-0.0338***	-0.0847***	-0.0326***	0.0840***	0.0671***
	(0.0135)	0.0012	(0.0026)	(0.0013)	(0.0026)	(0.0021)
sworsened	-0.603***	0.0167***	0.0418***	0.0161***	-0.0414***	-0.0331***
	(0.0373)	0.0014	(0.0034)	(0.0013)	(0.0034)	(0.0027)
gworsened	-0.221***	0.0453***	0.1140***	0.0437***	-0.1130***	-0.0900***
	(0.0181)	0.0030	(0.0070)	(0.0029)	(0.0070)	(0.0056)
hsecure	0.119***	-0.0089***	-0.0224***	-0.0086***	0.0222***	0.0177***
	(0.0222)	0.0017	(0.0042)	(0.0016)	(0.0042)	(0.0033)
rsecure	0.039**	-0.0029**	-0.0074**	-0.0028**	0.0073**	0.0058**
	(0.0170)	0.0013	(0.0032)	(0.0012)	(0.0032)	(0.0025)
rinsecure	-0.035*	-0.0002	-0.0006	-0.0002	0.0006	0.0005
	(0.0203)	(0.0013)	(0.0032)	(0.0012)	(0.0031)	(0.0025)
hinsecure	0.0033	0.0026*	0.0067*	0.0025*	-0.0066*	-0.0053*
	(0.0167)	0.0015	(0.0038)	(0.0015)	(0.0038)	(0.0030)
Observations	60,280	60,280	60,280	60,280	60,280	60,280

Notes: This table reports the CRE ordered Probit estimates of the determinants of expectation errors. Coefficients and Marginal effects are presented respectively, and standard errors are given in parentheses. The *error2* values correspond to the following responses: 1 = "*epes*"; 2 = "*mpes*"; 3 = "*correct*"; 4 = "*mopt*"; 5 = "*eopt*". Mean values of all time-varying variables are included, but not reported, in all regressions. Age, household income, consumption, saving, marital status, education level, urban, employment, retired and Regional and Year dummies are included in all regressions, and their marginal effects are reported in Appendix 2.4 and Appendix 2.5. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Definitions of all variables are shown in Table 2.6.

Appendix 2.1 First stage results of past income changes

Panel A: First stage results of past income changes for *error1/eopt/mopt/mpes/epe*				
Dependent variable (second stage)	*error1/eopt/mopt/mpes/epe*			
First stage result of	*gimproved*	*simproved*	*gworsened*	*sworsened*
	(1)	(2)	(3)	(4)
*gimproved_*	0.472***	0.019	0.061**	0.016
mean	(0.036)	(0.054)	(0.021)	(0.031)
*simproved_*	0.01	0.562***	0.013	-0.045***
mean	(0.006)	(0.017)	(0.007)	(0.011)
*gworsened_*	0.005	-0.001	0.534***	0.117***
mean	(0.008)	(0.025)	(0.025)	(0.024)
*sworsened_*	0.003	-0.023	0.064***	0.376***
mean	(0.007)	(0.021)	(0.014)	(0.021)
*UNEM**	0.071	-0.207	-0.265**	-0.253
age1830	(0.069)	(0.186)	(0.087)	(0.146)
*UNEM**	0.038	-0.338	-0.033	-0.224
age3040	(0.061)	(0.178)	(0.092)	(0.142)
*UNEM**	0.038	-0.23	-0.075	-0.156
age4050	(0.061)	(0.167)	(0.091)	(0.139)

*UNEM**	0.006	-0.097	0.102	-0.163
age5065	(0.044)	(0.146)	(0.094)	(0.138)
age1830	0.004	0.046	0.012	-0.007
	(0.009)	(0.025)	(0.013)	(0.021)
age3040	-0.001	0.023	0.002	0.017
	(0.007)	(0.022)	(0.013)	(0.019)
age4050	-0.006	0.001	0.014	0.027
	(0.006)	(0.018)	(0.011)	(0.017)
age5065	-0.002	-0.007	0.005	0.028
	(0.004)	(0.014)	(0.009)	(0.014)
hhincome	0.018***	0.097***	-0.039***	-0.058***
	(0.001)	(0.003)	(0.002)	(0.003)
consumption	0.003**	0.023***	-0.012***	-0.009***
	(0.001)	(0.002)	(0.001)	(0.002)
saving	0.012***	0.059***	-0.016***	-0.036***
	(0.002)	(0.004)	(0.002)	(0.003)
hhsize	-0.006***	-0.019***	0.006***	0.017***
	(0.001)	(0.002)	(0.001)	(0.002)
cohabit	0.004	0.012	-0.020***	-0.016*

	(0.002)	(0.007)	(0.005)	(0.007)
nmarried	-0.008*	-0.012	-0.020**	-0.037***
	(0.004)	(0.012)	(0.006)	(0.010)
edup	-0.003	-0.045*	-0.003	-0.045*
	(0.008)	(0.023)	(0.013)	(0.019)
eduu	0.001	0.006	-0.008	-0.020*
	(0.004)	(0.010)	(0.005)	(0.008)
edus	0.001	-0.002	-0.011**	-0.013*
	(0.003)	(0.007)	(0.004)	(0.006)
urban	-0.012	-0.033	-0.005	0.021
	(0.016)	(0.042)	(0.020)	(0.028)
employed	0.011***	0.124***	-0.067***	-0.073***
	(0.002)	(0.005)	(0.003)	(0.005)
retired	0.009***	0.050***	-0.037***	-0.007
	(0.002)	(0.008)	(0.005)	(0.007)
hstatus	0.003**	0.015***	-0.017***	-0.019***
	(0.001)	(0.003)	(0.002)	(0.002)
F-statistics	36.00	161.93	57.21	50.64
(first stage)	(0.000)	(0.000)	(0.000)	(0.000)

Observations	109,012	109,012	109,012	109,012	

Panel B: Tests of instrument validity of past income changes

	error1	*eopt*	*mopt*	*mpes*	*epes*
	(1)	(2)	(3)	(4)	(5)
Under-identification test	222.16 (0.000)	222.16 (0.000)	222.16 (0.000)	222.16 (0.000)	222.16 (0.000)
Over-identification test	2.70 (0.609)	5.02 (0.284)	4.71 (0.317)	6.03 (0.196)	3.17 (0.529)
Observations	109,012	109,012	109,012	109,012	109,012

Notes: This table reports the first stage results of past income changes for *error1*, *eopt*, *mopt*, *mpes* and *epes* models respectively. The Sanderson-Windmeijer multivariate F-statistic is used to examine whether the model is weakly identified. The null hypothesis of the Sanderson-Windmeijer F-test is that the instruments are weakly identified. The Kleibergen-Paap rk LM statistic of under-identification is used at testing the null hypothesis of under-identification. The Hansen J-test of over-identifying is aimed at testing the exogeneity of instruments. The null hypothesis of the Hansen J-test is that the instruments are not correlated with the error term. P values of F-statistics, under-identification test and over-identification test are in parenthesis. UNEM is the province-specific unemployment rate. Regional and Year dummies are included in all models, but their coefficients are not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Coefficients are presented, and robust standard errors are given in parentheses. Definitions of all variables are shown in Table 2.6.

Appendix 2.2 First stage results of job insecurity

Panel A: First stage results of job insecurity for *error1/eopt/mopt/mpes/epe*				
Dependent variable (second stage)	*error1/eopt/mopt/mpes/epe*			
First stage result of	*hsecure*	*rsecure*	*hinsecure*	*rinsecure*
	(1)	(2)	(3)	(4)
*hsecure_*	0.425***	-0.044	0.057*	-0.036
mean	(0.026)	(0.028)	(0.028)	(0.029)
*rsecure_*	0.048*	0.342***	-0.065**	-0.011
mean	(0.020)	(0.028)	(0.024)	(0.028)
*hinsecure_*	0.052**	-0.071**	0.366***	-0.017
mean	(0.018)	(0.024)	(0.026)	(0.028)
*rinsecure_*	0.038*	-0.025	0.052*	0.232***
mean	(0.017)	(0.024)	(0.025)	(0.029)
*UNEM**	-0.288	-0.246	0.083	0.093
age1830	(0.207)	(0.277)	(0.238)	(0.279)
*UNEM**	-0.152	-0.443	0.336	-0.149
age3040	(0.190)	(0.284)	(0.246)	(0.260)
*UNEM**	0.057	-0.545*	0.694**	-0.119
age4050	(0.190)	(0.260)	(0.265)	(0.288)

*UNEM**	-0.084	-0.027	0.58	-0.099
age5065	(0.206)	(0.310)	(0.318)	(0.332)
age1830	0.022	0.033	-0.007	0.013
	(0.028)	(0.038)	(0.037)	(0.040)
age3040	0.002	0.065	-0.057	0.041
	(0.025)	(0.035)	(0.036)	(0.037)
age4050	-0.015	0.068*	-0.098**	0.037
	(0.022)	(0.031)	(0.036)	(0.036)
age5065	-0.004	0.014	-0.066	0.035
	(0.021)	(0.032)	(0.036)	(0.036)
hhincome	-0.003	0.002	0.001	0.006
	(0.004)	(0.005)	(0.004)	(0.005)
consumption	0.004	0.005	-0.003	-0.003
	(0.003)	(0.004)	(0.003)	(0.004)
saving	0.016***	-0.013*	0.002	-0.002
	(0.005)	(0.006)	(0.005)	(0.006)
hhsize	-0.003	0.001	0.002	-0.003
	(0.002)	(0.003)	(0.003)	(0.003)
cohabit	-0.003	-0.005	0.019	-0.018

	(0.009)	(0.011)	(0.011)	(0.012)
nmarried	-0.003	0.023	0.008	-0.033
	(0.014)	(0.018)	(0.014)	(0.017)
edup	-0.06	0.055	-0.085	0.092
	(0.037)	(0.104)	(0.069)	(0.093)
eduu	-0.025*	0.026	-0.002	0.009
	(0.013)	(0.015)	(0.012)	(0.015)
edus	-0.013	0.032**	-0.013	0.009
	(0.009)	(0.012)	(0.009)	(0.011)
urban	-0.02	-0.037	0.063	-0.058
	(0.038)	(0.068)	(0.056)	(0.062)
employed	-0.016	0.021	0.001	0.006
	(0.011)	(0.013)	(0.009)	(0.011)
retired	0.437	-0.271***	-0.083	0.202
	(0.411)	(0.020)	(0.065)	(0.550)
hstatus	0.009*	0.019***	-0.012***	-0.010*
	(0.003)	(0.004)	(0.004)	(0.004)
F-statistics	64.32	35.52	25.92	37.79
(first stage)	(0.000)	(0.000)	(0.000)	(0.000)

Observations	55410	55410	55410	55410	

Panel B: Tests of instrument validity of job insecurity

	error1	*eopt*	*mopt*	*mpes*	*epes*
	(1)	(2)	(3)	(4)	(5)
Under-identification test	115.58 (0.000)	115.58 (0.000)	115.58 (0.000)	115.58 (0.000)	115.58 (0.000)
Over-identification test	4.70 (0.318)	3.05 (0.549)	1.25 (0.869)	2.65 (0.617)	1.12 (0.889)
Observations	55410	55410	55410	55410	55410

Notes: This table reports the first stage results of job insecurity for *error1*, *eopt*, *mopt*, *mpes* and *epes* models respectively. The Sanderson-Windmeijer multivariate F-statistic is used to examine whether the model is weakly identified. The null hypothesis of the Sanderson-Windmeijer F-test is that the instruments are weakly identified. The Kleibergen-Paap rk LM statistic of under-identification is used at testing the null hypothesis of under-identification. The Hansen J-test of over-identifying is aimed at testing the exogeneity of instruments. The null hypothesis of the Hansen J-test is that the instruments are not correlated with the error term. P values of F-statistics, under-identification test and over-identification test are in parenthesis. UNEM is the province-specific unemployment rate. Regional and Year dummies are included in all models, but their coefficients are not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Coefficients are presented, and robust standard errors are given in parentheses. Definitions of all variables are shown in Table 2.6.

Appendix 2.3 First stage results of past income changes and job insecurity

Panel A: First stage results of past income changes and job insecurity for *error1/eopt/mopt/mpes/epe*								
Dependent variable (second stage)	*error1/eopt/mopt/mpes/epe*							
First stage result of	*gimproved*	*simproved*	*gworsened*	*sworsened*	*hsecure*	*rsecure*	*hinsecure*	*rinsecure*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
*gimproved_*	0.443***	0.014	0.015	0.005	0.038	-0.136	0.082	-0.04
mean	(0.049)	(0.086)	(0.034)	(0.056)	(0.063)	(0.077)	(0.071)	(0.078)
*simproved_*	0.018	0.502***	0.017	-0.042*	0.017	-0.003	0.033	0.012
mean	(0.010)	(0.028)	(0.011)	(0.019)	(0.018)	(0.025)	(0.023)	(0.027)
*gworsened_*	0.002	-0.026	0.461***	0.186***	0.047	-0.117**	0.114**	0.034
mean	(0.014)	(0.040)	(0.032)	(0.036)	(0.035)	(0.038)	(0.039)	(0.042)
*sworsened_*	-0.007	-0.057	0.077***	0.382***	-0.029	0.006	-0.013	0.005
mean	(0.012)	(0.034)	(0.018)	(0.030)	(0.025)	(0.033)	(0.030)	(0.035)
*hsecure_*	0.001	0.026	0.029	0.001	0.419***	-0.033	0.045	-0.037
mean	(0.012)	(0.031)	(0.017)	(0.023)	(0.026)	(0.029)	(0.028)	(0.029)
*rsecure_*	0.001	0.051	0.02	0.016	0.046*	0.341***	-0.066**	-0.012
mean	(0.010)	(0.029)	(0.013)	(0.020)	(0.020)	(0.028)	(0.024)	(0.028)
*hinsecure_*	-0.004	0.058*	0.033*	-0.037	0.048**	-0.064**	0.358***	-0.02

mean	(0.009)	(0.028)	(0.013)	(0.020)	(0.018)	(0.024)	(0.026)	(0.028)
*rinsecure_*	0.001	0.088**	0.028*	0.003	0.036*	-0.022	0.049	0.231***
mean	(0.009)	(0.027)	(0.013)	(0.021)	(0.017)	(0.024)	(0.025)	(0.029)
*UNEM**	-0.012	-0.048	0.019	-0.042	-0.295	-0.229	0.069	0.091
age1830	(0.097)	(0.284)	(0.113)	(0.212)	(0.207)	(0.277)	(0.238)	(0.279)
*UNEM**	-0.054	-0.409	0.138	0.033	-0.157	-0.426	0.324	-0.15
age3040	(0.087)	(0.266)	(0.120)	(0.200)	(0.190)	(0.283)	(0.246)	(0.260)
*UNEM**	-0.123	-0.43	0.014	0.297	0.056	-0.533*	0.686**	-0.122
age4050	(0.097)	(0.274)	(0.127)	(0.206)	(0.190)	(0.260)	(0.265)	(0.288)
*UNEM**	-0.142	-0.441	0.116	0.600*	-0.083	-0.029	0.578	-0.102
age5065	(0.105)	(0.318)	(0.144)	(0.255)	(0.206)	(0.310)	(0.318)	(0.332)
age1830	0.005	-0.023	0.012	0.012	0.023	0.031	-0.006	0.013
	(0.013)	(0.040)	(0.019)	(0.033)	(0.028)	(0.038)	(0.037)	(0.040)
age3040	0.003	0.002	0.001	0.02	0.002	0.063	-0.055	0.042
	(0.011)	(0.036)	(0.018)	(0.031)	(0.025)	(0.035)	(0.036)	(0.037)
age4050	0.001	0.017	0.011	-0.001	-0.015	0.067*	-0.098**	0.038
	(0.010)	(0.034)	(0.017)	(0.029)	(0.022)	(0.031)	(0.036)	(0.036)
age5065	0.008	0.042	0.003	-0.038	-0.004	0.014	-0.066	0.035
	(0.010)	(0.034)	(0.018)	(0.030)	(0.021)	(0.032)	(0.036)	(0.036)

hhincome	0.022***	0.121***	-0.035***	-0.071***	-0.003	0.002	0.001	0.006
	(0.002)	(0.005)	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)
consumption	0.005***	0.027***	-0.012***	-0.010***	0.004	0.005	-0.003	-0.003
	(0.001)	(0.004)	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
saving	0.013***	0.070***	-0.013***	-0.040***	0.016***	-0.013*	0.002	-0.002
	(0.003)	(0.007)	(0.002)	(0.004)	(0.005)	(0.006)	(0.005)	(0.006)
hhsize	-0.006***	-0.024***	0.011***	0.019***	-0.003	0.001	0.002	-0.003
	(0.001)	(0.003)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
cohabit	0.007	0.018	-0.007	-0.013	-0.003	-0.005	0.019	-0.018
	(0.004)	(0.012)	(0.006)	(0.010)	(0.009)	(0.011)	(0.011)	(0.012)
nmarried	-0.011	0.001	-0.009	-0.025	-0.003	0.023	0.008	-0.034
	(0.006)	(0.018)	(0.008)	(0.013)	(0.014)	(0.018)	(0.014)	(0.017)
edup	-0.012	0.053	-0.025	-0.210***	-0.06	0.053	-0.084	0.092
	(0.029)	(0.092)	(0.064)	(0.059)	(0.038)	(0.104)	(0.069)	(0.093)
eduu	0.001	-0.008	0.001	-0.012	-0.025*	0.026	-0.002	0.009
	(0.006)	(0.016)	(0.007)	(0.012)	(0.013)	(0.015)	(0.012)	(0.015)
edus	-0.002	-0.002	-0.004	-0.01	-0.012	0.031**	-0.013	0.009
	(0.005)	(0.012)	(0.005)	(0.009)	(0.009)	(0.012)	(0.009)	(0.011)
urban	-0.029	-0.104	-0.004	0.021	-0.018	-0.041	0.067	-0.058

	(0.030)	(0.065)	(0.021)	(0.035)	(0.038)	(0.068)	(0.056)	(0.062)
employed	0.015***	0.123***	-0.042***	-0.098***	-0.016	0.02	0.001	0.006
	(0.005)	(0.013)	(0.007)	(0.011)	(0.011)	(0.013)	(0.009)	(0.011)
retired	-0.077	-0.14	-0.029***	-0.167*	0.438	-0.273***	-0.081	0.203
	(0.074)	(0.077)	(0.008)	(0.067)	(0.412)	(0.022)	(0.066)	(0.550)
hstatus	0.005*	0.022***	-0.010***	-0.019***	0.009*	0.019***	-0.012***	-0.010*
	(0.002)	(0.005)	(0.002)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
F-statistics	17.21	38.06	18.73	19.83	52.26	29.64	29.95	21.55
(first stage)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	55410	55410	55410	55410	55410	55410	55410	55410

Panel B: Tests of instrument validity of past income changes and job insecurity

	error1	*eopt*	*mopt*	*mpes*	*epes*
	(1)	(2)	(3)	(4)	(5)
Under-identification test	86.41 (0.000)	86.41 (0.000)	86.41 (0.000)	86.41 (0.000)	86.41 (0.000)
Over-identification test	3.74 (0.441)	2.65 (0.618)	1.15 (0.886)	3.18 (0.531)	3.34 (0.501)
Observations	55410	55410	55410	55410	55410

Notes: This table reports the first stage results of past income changes and job insecurity for *error1*, *eopt*, *mopt*, *mpes* and *epes* models respectively. The Sanderson-Windmeijer multivariate F-statistic is used to examine whether the model is weakly identified. The null hypothesis of the Sanderson-Windmeijer F-test is that the instruments are weakly identified. The Kleibergen-Paap rk LM statistic of under-identification is used at testing the null hypothesis of under-identification. The Hansen J-test of over-identifying is aimed at testing the exogeneity of instruments. The null hypothesis of the Hansen J-test is that the instruments are not correlated with the error term. P values of F-statistics, under-identification test and over-identification test are in parenthesis. UNEM is the province-specific unemployment rate. Regional and Year dummies are included in all models, but their coefficients are not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the

1% levels. Coefficients are presented, and robust standard errors are given in parentheses. Definitions of all variables are shown in Table 2.6.

Appendix 2.4 CRE ordered Probit results (other controls)

	The role of past income changes						The role of job insecurity					
Variable	*Coefficient* （*S.E*）	*Error2=1*	*Error2=2*	*Error2=3*	*Error2=4*	*Error2=5*	*Coefficient* （*S.E*）	*Error2=1*	*Error2=2*	*Error2=3*	*Error2=4*	*Error2=5*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
male	0.031***	-0.0025***	-0.0057***	-0.0018***	0.0057***	0.0043***	0.035***	-0.0027***	-0.0068***	-0.0026***	0.0067***	0.0054***
	(0.0081)	(0.0007)	(0.0015)	(0.0005)	(0.0015)	(0.0011)	(0.0108)	(0.0008)	(0.0021)	(0.0008)	(0.0021)	(0.0017)
age1830	-0.025	0.0020	0.0047	0.0015	-0.0048	-0.0036	0.197**	-0.0153**	-0.0382**	-0.0149**	0.0380**	0.0304**
	(0.0488)	(0.0039)	(0.0090)	(0.0029)	(0.0091)	(0.0068)	(0.0814)	(0.0063)	(0.0158)	(0.0061)	(0.0157)	(0.0125)
age3040	-0.038	0.0030	0.0070	0.0023	-0.0071	-0.0053	0.092	-0.0071	-0.0180	-0.0070	0.0179	0.0143
	(0.0415)	(0.0033)	(0.0077)	(0.0025)	(0.0077)	(0.0058)	(0.0755)	(0.0059)	(0.0146)	(0.0057)	(0.0145)	(0.0116)
age4050	-0.057*	0.0046*	0.0105*	0.0034*	-0.0106*	-0.0079*	0.003	-0.0002	-0.0007	-0.0002	0.0007	0.0005
	(0.0331)	(0.0027)	(0.0061)	(0.0020)	(0.0062)	(0.0046)	(0.0697)	(0.0054)	(0.0135)	(0.0053)	(0.0134)	(0.0107)
age5065	-0.039*	0.0032*	0.0073*	0.0024*	-0.0074*	-0.0055*	-0.037	0.0028	0.0072	0.0028	-0.0071	-0.0057
	(0.0234)	(0.0019)	(0.0043)	(0.0014)	(0.0044)	(0.0033)	(0.0642)	(0.0050)	(0.0124)	(0.0048)	(0.0124)	(0.0099)
hhincome	0.016*	-0.0013*	-0.0030*	-0.0010*	0.0031*	0.0023*	0.111***	-0.0085***	-0.0215***	-0.0083***	0.0214***	0.0171***
	(0.0088)	(0.0007)	(0.0016)	(0.0005)	(0.0016)	(0.0012)	(0.0122)	(0.0010)	(0.0024)	(0.0009)	(0.0024)	(0.0019)
consumption	-0.001	0.0001	0.0002	0.0001	-0.0002	-0.0002	0.018*	-0.0014*	-0.0035*	-0.0013*	0.0035*	0.0028*

	(0.0067)	(0.0005)	(0.0013)	(0.0004)	(0.0013)	(0.0009)	(0.0098)	(0.0008)	(0.0019)	(0.0007)	(0.0019)	(0.0015)
saving	0.012	-0.0010	-0.0023	-0.0007	0.0023	0.0017	0.083***	-0.0064***	-0.0161***	-0.0062***	0.0160***	0.0128***
	(0.0113)	(0.0009)	(0.0021)	(0.0007)	(0.0021)	(0.0016)	(0.0160)	(0.0012)	(0.0031)	(0.0012)	(0.0031)	(0.0025)
hhsize	0.008	-0.0006	-0.0015	-0.0005	0.0015	0.0011	-0.013*	0.0010*	0.0025*	0.0009*	-0.0025*	-0.0020*
	(0.0056)	(0.0005)	(0.0010)	(0.0003)	(0.0011)	(0.0008)	(0.0075)	(0.0006)	(0.0015)	(0.0006)	(0.0014)	(0.0012)
cohabit	0.105***	-0.0084***	-0.0193***	-0.0063***	0.0195***	0.0146***	0.152***	-0.0118***	-0.0295***	-0.0115***	0.0293***	0.0235***
	(0.0214)	(0.0017)	(0.0040)	(0.0013)	(0.0040)	(0.0030)	(0.0299)	(0.0023)	(0.0058)	(0.0023)	(0.0058)	(0.0046)
nmarried	-0.014	0.0011	0.0027	0.0008	-0.0027	-0.0020	0.025	-0.0020	-0.0050	-0.0019	0.0049	0.0039
	(0.0323)	(0.0026)	(0.0060)	(0.0020)	(0.0060)	(0.0045)	(0.0439)	(0.0034)	(0.0085)	(0.0033)	(0.0085)	(0.0068)
edup	0.112*	-0.0090*	-0.0207*	-0.0067*	0.0208*	0.0156*	0.220	-0.017	-0.0425	-0.0165	0.0423	0.0338
	(0.0681)	(0.0055)	(0.0125)	(0.0041)	(0.0127)	(0.0095)	(0.263)	(0.0204)	(0.0510)	(0.0198)	(0.0507)	(0.0405)
eduu	-0.022	0.0022	0.0050	0.0016	-0.0051	-0.0038	-0.040	0.0031	0.0078	0.0030	-0.0077	-0.0061
	(0.0271)	(0.0022)	(0.0050)	(0.0016)	(0.0050)	(0.0038)	(0.0406)	(0.0032)	(0.0079)	(0.0031)	(0.0078)	(0.0063)
edus	-0.019	0.0015	0.0035	0.0011	-0.0035	-0.0026	-0.020	0.0015	0.0038	0.0015	-0.0038	-0.0030
	(0.0205)	(0.0017)	(0.0038)	(0.0012)	(0.0038)	(0.0029)	(0.0301)	(0.0023)	(0.0058)	(0.0023)	(0.0058)	(0.0046)
urban	0.107	-0.0086	-0.0197	-0.0064	0.0199	0.0149	0.140	-0.0108	-0.0271	-0.0105	0.0269	0.0215
	(0.0997)	(0.0080)	(0.0184)	(0.0060)	(0.0186)	(0.0139)	(0.1380)	(0.0107)	(0.0268)	(0.0104)	(0.0267)	(0.0213)
employed	-0.068***	0.0054***	0.0126***	0.0041***	-0.0127***	-0.0094***	-0.047	0.0036	0.0092	0.0035	-0.0091	-0.0073
	(0.0150)	(0.0012)	(0.0028)	(0.0009)	(0.0028)	(0.0021)	(0.0347)	(0.0027)	(0.0067)	(0.0026)	(0.0067)	(0.0053)

retired	-0.068***	0.0055***	0.0127***	0.0041***	-0.0128***	-0.0095***	0.919***	-0.0711***	-0.1780***	-0.0692***	0.1770***	0.1410***
	(0.0223)	(0.0018)	(0.0041)	(0.0014)	(0.0042)	(0.0031)	(0.311)	(0.0027)	(0.0603)	(0.0236)	(0.0600)	(0.0479)
hstatus	0.048***	-0.0039***	-0.0089***	-0.0029***	0.0090***	0.0067***	0.053***	-0.0041***	-0.0103***	-0.0040***	0.0103***	0.0082***
	(0.0080)	(0.0007)	(0.0015)	(0.0005)	(0.0015)	(0.0011)	(0.0120)	(0.0009)	(0.0023)	(0.0009)	(0.0023)	(0.0018)
Observations	114,707	114,707	114,707	114,707	114,707	114,707	60,280	60,280	60,280	60,280	60,280	60,280

Notes: This table reports the CRE ordered Probit estimates of the determinants of expectation errors. Coefficients and Marginal effects are presented respectively, and standard errors are given in parentheses. The *error2* values correspond to the following responses: 1 = "*epes*"; 2 = "*mpes*"; 3 = "*correct*"; 4 = "*mopt*"; 5 = "*eopt*". Mean values of all time-varying variables, Regional and Year dummies are included, but not reported, in all regressions. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Definitions of all variables are shown in Table 2.6.

Appendix 2.5 CRE ordered Probit results (other controls)

The roles of past income changes and job insecurity

Variable	*Coefficient* *(S.E)*	*Error2=1*	*Error2=2*	*Error2=3*	*Error2=4*	*Error2=5*
	(1)	(2)	(3)	(4)	(5)	(6)
male	0.036***	-0.0027***	-0.0068***	-0.0026***	0.0068***	0.0054***
	(0.0111)	(0.0008)	(0.0021)	(0.0008)	(0.0021)	(0.0017)
age1830	-0.021	0.0016	0.0040	0.0015	-0.0039	-0.0031
	(0.0843)	(0.0063)	(0.0159)	(0.0061)	(0.0158)	(0.0126)
age3040	-0.060	0.0046	0.0115	0.0044	-0.0114	-0.0091
	(0.0779)	(0.0059)	(0.0147)	(0.0057)	(0.0146)	(0.0116)
age4050	-0.091	0.0068	0.0172	0.0066	-0.0170	-0.0136
	(0.0713)	(0.0054)	(0.0135)	(0.0052)	(0.0133)	(0.0106)
age5065	-0.093	0.0071	0.0177	0.0068	-0.0175	-0.0140
	(0.0653)	(0.0049)	(0.0123)	(0.0047)	(0.0122)	(0.0098)
hhincome	0.020*	-0.0015*	-0.0039*	-0.0015*	0.0039*	0.0031*
	(0.0126)	(0.0009)	(0.0024)	(0.0009)	(0.0024)	(0.0019)
consumption	-0.005	0.0004	0.0011	0.0004	-0.0011	-0.0009
	(0.00996)	(0.0008)	(0.0019)	(0.0007)	(0.0019)	(0.0015)
saving	0.023	-0.0018	-0.0044	-0.0017	0.0044	0.0035
	(0.0162)	(0.0012)	(0.0031)	(0.0012)	(0.0030)	(0.0024)
hhsize	0.008	-0.0006	-0.0016	-0.0006	0.0016	0.0013
	(0.00788)	(0.0006)	(0.0015)	(0.0006)	(0.0015)	(0.0012)
cohabit	0.124***	-0.0093***	-0.0234***	-0.0090***	0.0232***	0.0185***
	(0.0310)	(0.0023)	(0.0059)	(0.0023)	(0.0058)	(0.0046)
nmarried	-0.011	0.0009	0.0022	0.0008	-0.0022	-0.0017

	(0.0456)	(0.0034)	(0.0086)	(0.0033)	(0.0085)	(0.0068)
edup	0.127	-0.0095	-0.0239	-0.0092	0.0237	0.0189
	(0.243)	(0.0183)	(0.0459)	(0.0176)	(0.0455)	(0.0363)
eduu	-0.048	0.0036	0.0091	0.0035	-0.0091	-0.0072
	(0.0423)	(0.0032)	(0.0080)	(0.0031)	(0.0079)	(0.0063)
edus	-0.032	0.0025	0.0062	0.0024	-0.0061	-0.0049
	(0.0312)	(0.0024)	(0.0059)	(0.0023)	(0.0058)	(0.0047)
urban	0.245*	-0.0184*	-0.0462*	-0.0178*	0.0458*	0.0366*
	(0.134)	(0.0101)	(0.0252)	(0.0097)	(0.0250)	(0.0199)
employed	-0.166***	0.0125***	0.0314***	0.0121***	-0.0311***	-0.0249***
	(0.0362)	(0.0027)	(0.0068)	(0.0026)	(0.0068)	(0.0054)
retired	1.054***	-0.0793***	-0.1990***	-0.0764***	0.1970***	0.1570***
	(0.337)	(0.0254)	(0.0635)	(0.0245)	(0.0630)	(0.0503)
hstatus	0.028**	-0.0021**	-0.0054**	-0.0020**	0.0053**	0.0042**
	(0.0120)	(0.0009)	(0.0023)	(0.0009)	(0.0023)	(0.0018)
Observations	60,280	60,280	60,280	60,280	60,280	60,280

Notes: This table reports the CRE ordered Probit estimates of the determinants of expectation errors. Coefficients and Marginal effects are presented respectively, and standard errors are given in parentheses. The *error2* values correspond to the following responses: 1 = "*epes*"; 2 = "*mpes*"; 3 = "*correct*"; 4 = "*mopt*"; 5 = "*eopt*". Mean values of all time-varying variables, Regional and Year dummies are included, but not reported, in all regressions. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Definitions of all variables are shown in Table 2.6.

Appendix 2.6 Variable addition tests 1 (num_wave)

Variables	*error1*	*eopt*	*mopt*	*epes*	*mpes*	*error1*	*eopt*	*mopt*	*epes*	*mpes*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
num_wave	0.006	0.002*	-0.001	-0.001	-0.001	0.017***	0.002	0.004	-0.001	-0.005**
	(0.004)	(0.001)	(0.002)	(0.001)	(0.002)	(0.006)	(0.002)	(0.003)	(0.001)	(0.002)
gimproved	0.782***	0.132***	0.210***	-0.051***	-0.152***					
	(0.028)	(0.010)	(0.012)	(0.005)	(0.009)					
simproved	0.383***	0.037***	0.149***	-0.018***	-0.111***					
	(0.009)	(0.002)	(0.004)	(0.001)	(0.003)					
gworsened	-0.611***	-0.075***	-0.042***	0.161***	0.048***					
	(0.021)	(0.005)	(0.007)	(0.006)	(0.006)					
sworsened	-0.213***	-0.018***	-0.024***	0.050***	0.050***					
	(0.011)	(0.003)	(0.004)	(0.003)	(0.004)					
hsecure						0.124***	0.035***	0.019**	-0.002	-0.028***
						(0.020)	(0.006)	(0.009)	(0.004)	(0.007)
rsecure						0.047***	0.009**	0.012*	-0.005*	-0.007
						(0.015)	(0.004)	(0.007)	(0.003)	(0.006)
hinsecure						-0.042**	0.001	0.001	0.018***	0.002
						(0.018)	(0.005)	(0.007)	(0.004)	(0.007)
rinsecure						0.005	0.004	0.010	0.003	0.006
						(0.015)	(0.004)	(0.006)	(0.003)	(0.006)
age1830	-0.009	0.020	-0.041**	-0.006	0.028	-0.031	-0.003	-0.008	-0.023	0.062*
	(0.047)	(0.013)	(0.020)	(0.009)	(0.018)	(0.079)	(0.023)	(0.033)	(0.017)	(0.033)
age3040	-0.023	0.020*	-0.038**	0.001	0.027*	-0.073	-0.006	-0.018	-0.016	0.070**

	(0.039)	(0.011)	(0.016)	(0.008)	(0.016)	(0.071)	(0.020)	(0.029)	(0.016)	(0.030)
age4050	-0.042	0.008	-0.021*	0.001	0.036***	-0.098	-0.023	-0.012	-0.016	0.070***
	(0.030)	(0.008)	(0.012)	(0.007)	(0.012)	(0.063)	(0.018)	(0.026)	(0.015)	(0.027)
age5065	-0.032	0.004	-0.002	0.003	0.028***	-0.083	-0.025	0.001	-0.016	0.062**
	(0.020)	(0.005)	(0.008)	(0.005)	(0.009)	(0.056)	(0.016)	(0.023)	(0.013)	(0.025)
hhincome	0.018**	0.008***	-0.003	-0.003*	-0.000	0.122***	0.019***	0.022***	-0.018***	-0.023***
	(0.008)	(0.002)	(0.003)	(0.002)	(0.003)	(0.012)	(0.003)	(0.005)	(0.002)	(0.004)
consumption	-0.000	-0.002	0.004	0.000	-0.000	0.022**	0.002	0.010**	-0.000	-0.004
	(0.006)	(0.002)	(0.002)	(0.001)	(0.002)	(0.009)	(0.003)	(0.004)	(0.002)	(0.003)
saving	0.012	0.005**	-0.001	0.002	-0.005	0.075***	0.014***	0.010	-0.008***	-0.016***
	(0.010)	(0.003)	(0.004)	(0.002)	(0.004)	(0.014)	(0.004)	(0.006)	(0.003)	(0.006)
hhsize	0.005	-0.001	0.004**	-0.000	-0.003	-0.021***	-0.005**	0.001	0.003***	0.001
	(0.005)	(0.001)	(0.002)	(0.001)	(0.002)	(0.007)	(0.002)	(0.003)	(0.001)	(0.003)
cohabit	0.093***	0.027***	0.015*	-0.002	-0.012*	0.126***	0.036***	0.025**	-0.004	-0.015
	(0.019)	(0.005)	(0.008)	(0.004)	(0.007)	(0.027)	(0.008)	(0.012)	(0.006)	(0.011)
nmarried	-0.012	-0.006	-0.004	-0.003	0.003	-0.004	-0.001	-0.002	-0.004	0.008
	(0.028)	(0.008)	(0.012)	(0.006)	(0.011)	(0.040)	(0.012)	(0.017)	(0.008)	(0.015)
edup	0.089	-0.008	0.028	-0.022*	-0.040*	0.138	-0.016	0.073	-0.002	-0.136
	(0.057)	(0.015)	(0.023)	(0.013)	(0.024)	(0.237)	(0.056)	(0.083)	(0.058)	(0.090)
eduu	-0.024	-0.008	0.007	0.001	0.004	-0.042	-0.019*	0.014	0.001	0.005
	(0.023)	(0.007)	(0.010)	(0.004)	(0.009)	(0.037)	(0.011)	(0.016)	(0.007)	(0.014)
edus	-0.015	-0.012**	0.008	-0.004	0.005	-0.031	-0.022***	0.021*	-0.000	0.001
	(0.018)	(0.005)	(0.008)	(0.003)	(0.007)	(0.027)	(0.008)	(0.012)	(0.005)	(0.010)
urban	0.085	0.007	0.116***	0.039**	-0.028	0.103	-0.027	0.172***	0.006	-0.010
	(0.084)	(0.022)	(0.035)	(0.020)	(0.036)	(0.125)	(0.033)	(0.051)	(0.025)	(0.053)

employed	-0.065***	-0.017***	-0.002	0.010***	0.003	-0.053*	-0.027***	0.003	-0.004	-0.001
	(0.013)	(0.004)	(0.005)	(0.002)	(0.005)	(0.032)	(0.010)	(0.014)	(0.005)	(0.011)
retired	-0.064***	-0.021***	-0.014*	0.005	-0.007	0.789***	-0.017	0.471	0.003	-0.367***
	(0.019)	(0.005)	(0.008)	(0.004)	(0.008)	(0.302)	(0.017)	(0.355)	(0.007)	(0.087)
hstatus	0.041***	0.006***	0.010***	-0.003*	-0.012***	0.046***	0.007**	0.012***	-0.005**	-0.009**
	(0.007)	(0.002)	(0.003)	(0.001)	(0.003)	(0.011)	(0.003)	(0.005)	(0.002)	(0.004)
Obs	114,707	114,707	114,707	114,707	114,707	60,280	60,280	60,280	60,280	60,280
R-squared	0.069	0.016	0.027	0.041	0.023	0.013	0.008	0.003	0.005	0.004
Number of id	25,837	25,837	25,837	25,837	25,837	16,515	16,515	16,515	16,515	16,515

Notes: This table presents the FE results of the main models for the determinants of expectation errors controlling for *num_wave*. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Coefficients are presented, and robust standard errors are given in parentheses. Regional and Year dummies are included but not reported. Definitions of all variables are shown in Table 2.6.

Appendix 2.7 Variable addition tests 2 (next_wave)

Variables	*error1*	*eopt*	*mopt*	*epes*	*mpes*	*error1*	*eopt*	*mopt*	*epes*	*mpes*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
next_wave	-0.024***	-0.015***	-0.001	-0.004***	-0.001	-0.038***	-0.014***	-0.006	-0.002	0.005
	(0.008)	(0.002)	(0.003)	(0.002)	(0.003)	(0.012)	(0.003)	(0.005)	(0.002)	(0.004)
gimproved	0.783***	0.133***	0.210***	-0.051***	-0.152***					
	(0.028)	(0.010)	(0.012)	(0.005)	(0.009)					
simproved	0.383***	0.037***	0.149***	-0.018***	-0.111***					
	(0.009)	(0.002)	(0.004)	(0.001)	(0.003)					
gworsened	-0.611***	-0.075***	-0.042***	0.161***	0.048***					
	(0.021)	(0.005)	(0.007)	(0.006)	(0.006)					
sworsened	-0.213***	-0.018***	-0.024***	0.050***	0.050***					
	(0.011)	(0.003)	(0.004)	(0.003)	(0.004)					
hsecure						0.124***	0.035***	0.019**	-0.002	-0.028***
						(0.020)	(0.006)	(0.009)	(0.004)	(0.007)
rsecure						0.047***	0.009**	0.012*	-0.005*	-0.007
						(0.015)	(0.004)	(0.007)	(0.003)	(0.006)
hinsecure						-0.042**	0.001	0.001	0.018***	0.002
						(0.018)	(0.005)	(0.007)	(0.004)	(0.007)
rinsecure						0.005	0.004	0.010	0.003	0.006
						(0.015)	(0.004)	(0.006)	(0.003)	(0.006)
age1830	-0.010	0.019	-0.041**	-0.007	0.028	-0.028	-0.002	-0.007	-0.024	0.061*
	(0.047)	(0.013)	(0.020)	(0.009)	(0.018)	(0.079)	(0.023)	(0.033)	(0.017)	(0.033)
age3040	-0.024	0.020*	-0.038**	0.001	0.027*	-0.071	-0.006	-0.017	-0.016	0.070**

	(0.039)	(0.011)	(0.016)	(0.008)	(0.016)	(0.071)	(0.020)	(0.029)	(0.016)	(0.030)
age4050	-0.044	0.007	-0.021*	0.001	0.036***	-0.098	-0.023	-0.011	-0.016	0.070**
	(0.030)	(0.008)	(0.012)	(0.007)	(0.012)	(0.063)	(0.018)	(0.026)	(0.015)	(0.027)
age5065	-0.033	0.003	-0.002	0.003	0.028***	-0.081	-0.025	0.001	-0.016	0.061**
	(0.020)	(0.005)	(0.008)	(0.005)	(0.009)	(0.056)	(0.016)	(0.023)	(0.013)	(0.025)
hhincome	0.018**	0.008***	-0.003	-0.003*	-0.000	0.122***	0.019***	0.022***	-0.018***	-0.023***
	(0.008)	(0.002)	(0.003)	(0.002)	(0.003)	(0.012)	(0.003)	(0.005)	(0.002)	(0.004)
consumption	-0.000	-0.002	0.004	0.000	-0.000	0.022**	0.002	0.010**	-0.000	-0.004
	(0.006)	(0.002)	(0.002)	(0.001)	(0.002)	(0.009)	(0.003)	(0.004)	(0.002)	(0.003)
saving	0.012	0.005**	-0.002	0.002	-0.005	0.075***	0.014***	0.010	-0.008***	-0.016***
	(0.010)	(0.003)	(0.004)	(0.002)	(0.004)	(0.014)	(0.004)	(0.006)	(0.003)	(0.006)
hhsize	0.005	-0.001	0.004**	-0.000	-0.003	-0.021***	-0.005**	0.001	0.003***	0.001
	(0.005)	(0.001)	(0.002)	(0.001)	(0.002)	(0.007)	(0.002)	(0.003)	(0.001)	(0.003)
cohabit	0.094***	0.027***	0.015*	-0.002	-0.012*	0.128***	0.036***	0.025**	-0.004	-0.016
	(0.019)	(0.005)	(0.008)	(0.004)	(0.007)	(0.027)	(0.008)	(0.012)	(0.006)	(0.011)
nmarried	-0.012	-0.006	-0.005	-0.003	0.002	-0.003	-0.001	-0.001	-0.004	0.008
	(0.028)	(0.008)	(0.012)	(0.006)	(0.011)	(0.040)	(0.012)	(0.017)	(0.008)	(0.015)
edup	0.089	-0.008	0.028	-0.022*	-0.040*	0.140	-0.016	0.073	-0.002	-0.137
	(0.057)	(0.015)	(0.023)	(0.013)	(0.024)	(0.237)	(0.056)	(0.083)	(0.058)	(0.090)
eduu	-0.023	-0.007	0.007	0.001	0.004	-0.042	-0.018*	0.014	0.001	0.005
	(0.023)	(0.007)	(0.010)	(0.004)	(0.009)	(0.037)	(0.011)	(0.016)	(0.007)	(0.014)
edus	-0.014	-0.012**	0.008	-0.004	0.005	-0.032	-0.022***	0.021*	-0.000	0.001
	(0.018)	(0.005)	(0.008)	(0.003)	(0.007)	(0.027)	(0.008)	(0.012)	(0.005)	(0.010)
urban	0.088	0.008	0.116***	0.039**	-0.029	0.109	-0.026	0.173***	0.006	-0.012
	(0.084)	(0.022)	(0.035)	(0.020)	(0.036)	(0.126)	(0.034)	(0.051)	(0.025)	(0.053)

employed	-0.066***	-0.017***	-0.002	0.010***	0.003	-0.052	-0.026***	0.003	-0.004	-0.001
	(0.013)	(0.004)	(0.005)	(0.002)	(0.005)	(0.032)	(0.010)	(0.014)	(0.005)	(0.011)
retired	-0.064***	-0.021***	-0.014*	0.005	-0.007	0.788***	-0.019	0.471	0.002	-0.368***
	(0.019)	(0.005)	(0.008)	(0.004)	(0.008)	(0.288)	(0.015)	(0.353)	(0.007)	(0.090)
hstatus	0.041***	0.006***	0.010***	-0.003*	-0.012***	0.046***	0.006**	0.012***	-0.005**	-0.009**
	(0.007)	(0.002)	(0.003)	(0.001)	(0.003)	(0.011)	(0.003)	(0.005)	(0.002)	(0.004)
Obs	114,707	114,707	114,707	114,707	114,707	60,280	60,280	60,280	60,280	60,280
R-squared	0.069	0.017	0.027	0.041	0.023	0.013	0.009	0.003	0.005	0.004
Number of id	25,837	25,837	25,837	25,837	25,837	16,515	16,515	16,515	16,515	16,515

Notes: This table presents the FE results of the main models for the determinants of expectation errors controlling for *next_wave*. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Coefficients are presented, and robust standard errors are given in parentheses. Regional and Year dummies are included but not reported. Definitions of all variables are shown in Table 2.6.

Chapter Three: Errors in financial expectations and unhealthy behaviours - Evidence from Russia

3.1 Introduction

Preventing and reducing the health and social burden caused by the harmful use of cigarettes and alcohol has become a worldwide public health priority. According to the World Health Organization (WHO) (2016), Russia has the 2nd highest cigarette consumption in the world, averaging 2295 cigarettes consumed per capita per year. Meanwhile, Russia is also the 4th highest pure alcohol consumption country in the world (11.7 litres per capita). Zaridze *et al.* (2014) find that the high prevalence of smoking and drinking has become a major factor contributing to the Russian mortality crisis in the past three decades. For example, they find only 7% of UK males, but 37% of Russian males would die before the age of 55. In the face of Russia's catastrophic health profile and the very high rates of smoking and binge drinking, the Russian government has launched various anti-drinking (from the mid 2000s) and anti-smoking campaigns (since 2013) to address these unhealthy behaviours.

A variety of psychological and economic studies have examined the determinants of individuals' decisions of smoking and drinking. One potential determinant of smoking and drinking which has been neglected in the economics literature is the role of optimism and pessimism held by individuals. Puri and Robinson (2007) find that optimism and pessimism are both important determinants of individuals' decision-making.

According to Geers *et al.* (2013) and Jefferson *et al.* (2017), there are mainly two types of optimism/pessimism: dispositional optimism/pessimism and unrealistic optimism/pessimism. Dispositional optimism (pessimism) refers to individuals who hold generalized favorable expectation for their future. In other words, it refers to individuals holding a general belief that good (bad) events will happen more often than bad (good) events in a lifetime. This is generally considered to be a fixed personal trait (Scheier and Carver, 1985). Much of the literature has found a positive (negative) relationship between dispositional optimism (pessimism) and unhealthy behaviours (Giltay *et al.*, 2007; Wray and Dvorak, 2013).

While holding a positive expectation has been associated with better physical and mental health (Carver *et al.*, 2010; Scheier and Carver, 2018), one could easily fall into the trap of unrealistic optimism. In the psychology literature, unrealistic optimism (pessimism) is measured as the tendency of individuals to expect that they are more likely to experience positive (negative) outcomes than others. Shepperd *et al.* (2017) note that although research on unrealistic optimism/pessimism has expanded in a variety of disciplines including psychology, health and economics, only few studies have actually investigated the negative consequences of unrealistic optimism/pessimism.

There are essentially two negative consequences of unrealistic optimism/pessimism examined in the literature. Dillard *et al.* (2009) and Floyd *et al.* (2000) find that unrealistic optimists tend to underestimate the perceived risks of harmful behaviours, they are more likely to take unnecessary risks, to disengage from health screening and to abuse of risky health behaviours.

Studies that explore the negative consequences of unrealistic optimism have also reported that unrealistic optimism may lead to psychological ill-being. Unrealistic optimism is, by definition, overestimating the probability of a good outcome. People who display unrealistic optimism may experience emotional let-down or even anxiety and depressive syndromes when their expectations fall short (Dewberry *et al.*, 1990; Armor and Taylor, 1998). According to the Tension-Reduction Theory, individuals make use of alcohol (Conger, 1956) and cigarettes (Ikard *et al.*, 1969) as self-medication to relieve and regulate negative emotions and psychological distress. Thus, psychological ill-being could be one of the primary mechanisms through which unrealistic optimism/pessimism affects unhealthy behaviours.

However, due to the scarcity of data on individuals' expectations, there is a shortage of research exploring the relationship between unrealistic optimism/pessimism and unhealthy behaviours. In this study, following Balasuriya *et al.* (2010) and Cocco *et al.* (2017), I construct measures of unrealistic optimism and pessimism in the form of errors in financial expectations[5]. For instance, those who overestimate (underestimate) their financial situation in the next year are classified as being over-optimistic (over-pessimistic). Nevertheless, the concern over the validity of using errors in financial expectations as unrealistic optimism/pessimism has been raised. Brunnermeier and Parker (2005) find that standard economic models always take a traditional view of human behaviour as governed by rational expectation. However, they provide a structural model which incorporates irrational financial expectations to explain the irrational behaviours by

5 In this chapter, I define individuals with inaccurate financial expectations as being either unrealistic optimism or pessimism. I use 'errors in financial expectations', 'unrealistic optimism and pessimism' and 'extreme/moderate optimism and pessimism' interchangeably.

agents. They suggest that these irrational financial expectations are not much different from unrealistic optimism/pessimism.

By using errors in financial expectations as a measure of unrealistic optimism/pessimism, this study builds a bridge between the psychology, health and economics literature on unrealistic optimism/pessimism and unhealthy behaviours. The objective of this paper will be to provide an in depth analysis of the relationship between errors in financial expectations and unhealthy behaviours in Russia. A better understanding of these relationships will enable us to suggest policies targeted at minimizing the adverse health effects brought by the unhealthy behaviours.

This paper mainly fills the following four gaps in the literature. First, the association between errors in financial expectations and unhealthy behaviours in Russia has, to the best of my knowledge, never been analysed. It may provide a new perspective to understand the determinants of smoking and drinking. Second, to the best of my knowledge, no study has examined this relationship using panel data. Third, this study examines the channel through which errors in financial expectations may affect unhealthy behaviours in Russia. Fourth, this paper adds to the literature by adopting both static and dynamic models to examine the correlations between errors in financial expectations and smoking and drinking, taking into account the persistent nature of smoking and drinking.

The remainder of this paper proceeds as follows. Section 3.2 investigates the relevant literature on the association between optimism/pessimism and unhealthy behaviours, with specific emphasis on unrealistic optimism and pessimism. Section 3.3 states the main hypotheses. Section 3.4 presents the baseline specifications and methodology. Section 3.5 describes the data

and reports the descriptive statistics. Section 3.6 presents a discussion of the main empirical findings. Section 3.7 concludes.

3.2 Literature Review

The vast majority of literature spanning psychology, health and economics studies has examined the association between dispositional optimism/pessimism and unhealthy behaviours. Shepperd *et al.* (2017) note that few papers have looked at the relationship between unrealistic optimism/pessimism and unhealthy behaviours. In this section, I will investigate the literature linking various measures of optimism/pessimism to unhealthy behaviours.

The literature review proceeds as follows. Section 3.2.1 defines the concepts relating to dispositional and unrealistic optimism/pessimism. Section 3.2.2 starts by examining the relationship between dispositional optimism/pessimism and unhealthy behaviours. This is followed by a review of the empirical works that have investigated the proposed mechanism through which unrealistic optimism/pessimism may be related to unhealthy behaviours. Section 3.2.4 provides a summary of the contributions that this paper adds to the existing literature.

3.2.1 Definitions of optimism and pessimism

There are two main categories of optimism/pessimism identified in the literature spanning from psychology to economics: dispositional optimism/pessimism and unrealistic optimism/pessimism. According to the optimism literature pioneered by Scheier and Carver (1985), dispositional optimism (pessimism) refers to people having an enduring positive (negative) belief or expectation about future events. This has been considered to be a fixed personal trait.

Weinstein (1980) first identifies unrealistic optimism (pessimism) as the tendency of individuals to expect that they are more likely to experience positive (negative) outcomes for a

specific or isolated event. However, later studies (Shepperd *et al.*, 2013) further distinguish between unrealistic absolute optimism and unrealistic comparative optimism. Both forms can be estimated at the individual level or the group level.

Unrealistic absolute optimism is an erroneous belief that a positive outcome will be more likely to occur than it should be according to some objective standard such as the actual outcome (Shepperd *et al.*, 2017). Although finding an objective standard can be challenging at times, scholars have developed a variety of approaches to demonstrate unrealistic absolute optimism. At the individual level, one can compare individuals' predictions about a specific event to the actual outcome. For instance, Shepperd *et al.* (1996) find that individuals tend to underestimate the time it will take them to finish a task, or overestimate their essay marks. At the group level, researchers are interested in whether the group's mean of predicted scores is lower than the mean of actual scores.

Contrary to absolute optimism, unrealistic comparative optimism refers to individuals expecting a better outcome than others who display an identical or even lower risk profile. For example, at the individual level, a smoker displays unrealistic comparative optimism if he/she claims his/her risk of lung cancer is below that of the average smoker even when a comparative-risk calculator indicates otherwise. At the group level, a group of people displays unrealistic comparative optimism if the group's mean comparative risk estimate for an unfavorable outcome is below average. For instance, a group of college students are displaying unrealistic comparative optimism if they claim that their risk of developing substance abuse problems is lower than the average risk at their college (Shepperd *et al.*, 2013).

3.2.2 Optimism/pessimism and unhealthy behaviours

Davidson and Prkachin (1997) find that dispostional and unrealistic optimism/pessimism have both been thought to predict mental and physical health, albeit in opposite directions. Dispositional optimism is associated with better mental health (less anger, fatigue and distress), and better physical health (faster recovery time from health difficulties; more exercise and less substance abuse), while unrealistic optimism is found to be associated with deteriorated mental and physical health. The authors explain that opposite relationships found between dispositional and unrealistic optimists in predicting mental and physical health are due to the different coping strategies employed by dispositional optimists and unrealistic optimists.

Scheier and Carver (1992) conduct a experiment with college undergraduates in the U.S. On the one hand, the authors find that dispositional optimists are 'adaptive copers' who are well-adapted to stress, anxiety and any kinds of life events. When facing life challenges, they actively seek for social support, emphasize the positive aspects of stressful events and engage in health-promoting behaviours (healthier diet, less smoking and drinking).

On the other hand, the authors find that unrealistic optimists are 'maladaptive copers' who use a passive coping strategy such as denial. Unrealistic optimists believe that they are less likely to suffer the negative consequences from a series of life events. As a result of their decreased motivations to engage in health-promoting behaviours, unrealistic optimists are found to have higher chances of showing health difficulties (anxiety, distress, more time to recover from surgeries).

In the empirical studies, a vast body of research has primarily focused on the relationships between dispostional optimism/pessimism and smoking and drinking, yet to the best of my knowledge, no study has investigated the link between unrealistic optimism/pessimism and unhealthy behaviours. In this section, empirical studies investigating the association between dispositional optimism/pessimism and unhealthy behaviours are then explored.

The relationship between dispositional optimism (pessimism) and unhealthy behaviours was first considered by Kelloniemi *et al.* (2005). They use a cross-sectional population-based survey, which is carried out by sending postal questionnaires and the subjects were examined at a clinic in 1997-1998. The response rate is 74% (8690 individuals). Dispositional optimism is measured by the revised version of the Life Orientation Test (LOT-R) developed by Scheier and Carver in 1985. This test is the most common measure of dispositional optimism. It contains six items rated on a five-point scale ranging from 0 (strongly disagree) to 4 (strongly agree). Respondents scoring in the upper quartile are classified as optimists (18-24 points), while those scoring in the lower quartile are classed as pessimists (0-12). Others are defined as neutral (13-17). Measures of alcohol consumption include: during the last year, the usual amount of beer, wine and spirits consumed on one occasion, the average frequency of each consumption. Those who are among the highest alcohol consumption quartile are identified as heavy drinkers. Smoking behaviour is assessed by four categories: non-smoker, ex-smoker, occasional smoker and regular smoker. The authors employ multivariate logistic regression analyses and find that dispositional pessimism is positively associated with heavy drinking among males. They also find a negative (positive) relationship between dispositional optimism (pessimism) and smoking, which suggests that being optimistic (pessimistic) could lead to less (more) cigarettes consumed. The authors explain that

higher levels of smoking and drinking may be triggered by negative emotions and situational stress to which pessimists are particular more vulnerable than the others. Their findings provide one of the first insights into the negative (positive) correlation between dispositional optimism (pessimism) and unhealthy behaviours.

Following the work by Kelloniemi *et al.* (2005), subsequent research found consistent results that dispositional optimism (pessimism) is positively (negatively) associated with prudent health behaviours among the young and the elderly.

Steptoe *et al.* (2006) employ a community sample of 128 participants aged between 65 and 80 in the London area. They examine the relationships between dispositional optimism (pessimism) and prudent health behaviours. The authors suggest that little research on optimism and health behaviours has been carried out with elderly people because the relationships could be biased by health status. People with chronic disease or disabilities may in fact be less likely to engage in risky health behaviours, and may also be more pessimistic on account of their medical conditions. Their study avoids this problem by selecting patients who have no record of chronic illnesses and cancer in the last 5 years. Moderate drinking is defined as consuming more than 5 units of pure ethanol per week. Smokers are defined as those who have smoked within the last 5 years, and non-smokers are identified as those who haven't smoked in over 5 years. Optimism is measured using the LOT-R, and further divided into three categories: low, medium and high optimism. The authors find that the odds of smoking are 0.18 times lower for respondents in the high optimism tertile. This suggests that high optimists smoke less cigarettes than low/medium optimists. This may be explained considering that smoking has compensatory effects on negative moods, which are more commonly observed among low/medium optimists (Brody, 2006).

Wray and Dvorak (2013) make use of the National Longitudinal Study of Youth 1997 (NLSY97), which consists of 8984 youth aged 14 to 18. Their primary motivation is to identify whether dispositional optimism and pessimism are associated with adolescent alcohol use initiation and trajectory. Optimism (pessimism) is measured through question relating to "I'm always optimistic (pessimistic) about my future". Alcohol usage is collected by three related measures: frequency of alcohol use in the last month, average number of drinks per day in the past 30 days, and number of days in which they consumed 5 or more drinks in one occasion during the last month. The study shows that among boys, dispositional pessimism is positively associated with latent alcohol use, while optimism is negatively related to latent alcohol use. However, they only find a negative relationship between optimism and latent alcohol use in girls. Their findings suggest that for boys, those who have more negative expectations about the future may be more likely to report higher frequency and heavier alcohol use at younger ages. For boys and girls, those who are optimistic about the future are associated with less frequent alcohol usage and lower levels of alcohol consumed at a young age.

The results from Steptoe *et al.* (2006) and Wray and Dvorak (2013) are consistent with Kelloniemi *et al.* (2005), who also find that being optimism is associated with less smoking and drinking respectively. However, the above findings are somewhat ambiguous and limited by the fact that the sample is only restricted to young people and is therefore not representative (Wray and Dvorak, 2013), the small sample size (Steptoe *et al.*, 2006), and the fact that the analysis undertaken is cross-sectional (Kelloniemi *et al.*, 2005).

3.2.3 Mechanisms underlying the link between optimism/pessimism and unhealthy behaviours

According to Shepperd *et al.* (2017), there is a lack of empirical work on the direct link between unrealistic optimism/pessimism and unhealthy behaviours. However, a growing body of psychology and economics literature has amassed a wealth of evidence indicating that unrealistic optimism/pessimism has detrimental effects on mental well-being, and thus potentially on physical well-being too. This section highlights the theoretical and empirical evidence on why the link between unrealistic optimism/pessimism and unhealthy behaviours exists. The following studies suggest that unrealistic optimism/pessimism may be associated with smoking and drinking through psychological ill-being.

Armor and Taylor (1998) explain why unrealistic optimism is associated with psychological ill-being. For people who are unrealistically optimistic, the failure of goal-attainment may directly lead to disappointment, especially when people compare their actual outcomes to either what could happen or what should have happened. The emotional let-down is enlarged if optimistic predictions are too extreme (overly-optimistic), or pertinent to behaviours that have important consequences. For example, Armor and Taylor (1997) shows that people who are unrealistically optimistic prior to performing an important task show greater levels of disappointment after not meeting those expectations compared to those who make accurate predictions.

Cocco *et al.* (2017) use the British Household Panel Survey (BHPS) over the period 1991-2008 to examine the link between errors in financial expectations and emotional ill-being. Unlike

the previous study, they make use of two questions to construct the errors in financial expectations. In the questionnaires, participant are asked: 1) whether they are financially better, worse-off, or about the same as last year; 2) whether their financial situation would be better, worse, or the same in the next year. Unrealistic optimism (pessimism) refers to people who over-estimate (under-estimate) their financial situations. For example, people would be classified as unrealistic optimists if they answered 'they expect better finance situation' in year t-1, and then answer 'their financial situation has been worsened in the last year' in year t. The authors find that people who are more depressed or have difficulties facing problems show a lower probability of being unrealistically optimistic. Conversely, those who have less depressive syndromes are less likely to be unrealistically pessimistic. Unlike the prior research, this study controls for personality traits with individual fixed effects, and concludes that unrealistic optimistic/pessimistic is associated with psychological ill-being in a meaningful way.

The Decision Affect Theory (DAT) provides strong support for the association between unrealistic optimism/pessimism and psychological ill-being. Shepperd and McNulty (2002) suggest that the DAT has two main arguments: First, individuals' emotional response is determined by the counterfactual thinking, which is comparing what happened with what could happen. A good outcome is less pleasing when the conterfactual alternative is better, and a bad outcome is less displeasing when the conterfactual alternative is worse. For instance, Medvec *et al.* (1995) find that bronze medalists are happier than silver medalists in a study of medal winners at the Olympia games. The authors suggest that bronze medalists focused on the conterfactual alternative of winning no medal, while silver medalists focused on the conterfactual alternative of winning a gold medal. Their study suggests that silver medalists can be regarded as unrealistic

optimists, and brown medalists are unrealistic pessimists. They find that even though the final outcome (silver medals) of unrealistic optimists is objectively better than that (brown medals) of unrealistic pessimists, optimistic people can nonetheless feel worse off.

Second, the DAT proposes that, among all unexpected outcomes, only unexpected bad outcomes are unpleasant, but that unexpected good outcomes are pleasant. Shepperd and McNulty (2002) conduct an experiment to investigate how individuals respond to expected and unexpected good and bad outcomes. They ask 24 students to score four hypothetical outcomes from 1 ("this outcome would make you feel the best") to 4 ("this outcome would make you feel the worst"). The outcomes pertain to the following four settings: (1) predicting a $1,000 raise and receiving a $1,000 raise; (2) predicting a $1,500 raise and receiving a $1,500 raise, (c) predicting a $1,000 raise and receiving a $1,500 raise, and (d) predicting a $1,500 raise and receiving a $1,000 raise. Settings (1) and (2) are classified as making accurate predictions, setting (3) is being unrealistic pessimism and setting (4) is being unrealistic optimism. The results show that unrealistic optimists have the lowest scores (worst psychological responses), while unrealistic pessimists have the highest scores (best psychological responses). Those who make accurate predictions report scores that are in between unrealistic optimists and pessimists. Therefore, the DAT provides supporting evidence underlying the positive (negative) association between unrealistic optimism (pessimism) and psychological ill-being.

What's more, as Conger (1956) and Abbey *et al.* (1993) show that psychological distress is an important determinant of health-compromising behaviours, they find that psychological ill-being is one of the theoretical pathway through which unrealistic optimism/pessimism may affect smoking and drinking. Conger (1956) first reveals a psychological model of the Tension

Reduction Theory (TRT), which suggests that alcohol is sought and consumed to reduce tension states such as emotional distress and anxiety. Abbey *et al.* (1993) extend the TRT and find that people use alcohol and nicotine to escape, avoid and regulate negative affections and cope with psychological distress. Empirical studies also find a positive relationship between psychological ill-being and unhealthy behaviours. For example, Pearlin and Radabaugh (1976) use a representative sample of 2,300 people of the census-defined urbanized area of Chicago (U.S. Bureau of the Census 1972) and conclude that the positive relationship between psychological strains and drinking for the relief of distress is particularly strong among intensely anxious people who have low self-esteem and little control of their stress. Cockerham *et al.* (2006) employ a representative sample of 4,000 participants in Russia and find that the distress symptom items (loneliness, stress, anxiety and depression) are associated with a higher frequency and volume of drinking and a higher prevalence of smoking.

3.2.4 Contribution

My objective in this paper is to fill four gaps in the literature. First, this paper is aimed at understanding the relationship between errors in financial expectations and unhealthy behaviours. In particular, it examines whether unrealistic optimism and pessimism displayed by individuals regarding their financial situations are related to smoking and drinking. To the best of my knowledge, no study has investigated the association between errors in financial expectations and unhealthy behaviours in Russia. Since unrealistic optimism/pessimism could be somewhat an important determinant of the high prevalence of smoking and drinking in Russia, it is particularly interesting to investigate the relationship between errors in financial expectations and unhealthy behaviours in Russia (Stuckler and Basu, 2013; Shepperd *et al.*, 2017).

Next, to the best of my knowledge, no other study has examined the relationship between errors in financial expectations and unhealthy behaviours in a panel setting, focusing on the general population. Most existing studies focus on certain groups of the population, such as the young and the elderly people (e.g., Steptoe *et al.*, 2006; Giltay *et al.*, 2007). Moreover, smoking and drinking are lifelong habits that needs to be examined in a dynamic setting, which can only be done with an analysis based on panel data. Finally, using panel data enables us to control for unobserved individual heterogeneity, avoiding biases due to omitted variables.

Third, this paper adds to the literature by testing the mediating role of psychological ill-being, which can potentially explain the underlying association between errors in financial expectations and unhealthy behaviours. Referring back to Cocco *et al.* (2017), individuals overestimating their future finances are in danger of negative emotions or even psychological distress once their outcomes fall short of the expectations. The tension reduction model proposes that alcohol and cigarettes are consumed to overcome the negative mood states, such as anxiety, frustration and disappointment (Conger, 1956; Ikard *et al.*, 1969). By identifying the mechanism linking errors in financial expectations and unhealthy behaviours, greater insights into the determinants of smoking and drinking can be gained.

Last, unlike most prior research, which mostly uses cross-sectional analysis without controlling for the possible endogeneity of the regressors (Kelloniemi *et al.*, 2005; Wray and Dvorak, 2013), this study makes use of the generalised method of moments (GMM) estimator (Arellano and Bover, 1995; Blundell and Bond, 1998), which controls for both unobserved individual heterogeneity and possible endogeneity.

3.3 Hypotheses

In the sub-sections that follow, I will illustrate in turn the two main hypotheses related to the association between unrealistic optimism/pessimism and unhealthy behaviours, namely the general hypotheses and the psychological ill-being channel hypotheses.

3.3.1 General hypotheses

Dispositional and unrealistic optimism have been found to predict mental and physical health, albeit in opposite directions (Davidson and Prkachin, 1997). More precisely, dispositional optimism is associated with better mental and physical health (Iwanaga *et al.*, 2004; Kelloniemi *et al.*, 2005; Wray and Dvorak, 2013), while unrealistic optimism is associated with deteriorated mental and physical health (Armor and Taylor, 1998; Shepperd *et al.*, 2017).

Scheier and Carver (1992) explain the reason why dispositonal and unrealistic optimism have different effects on mental and physical health is due to the difference in adoption of coping strategies by dispositional and unrealistic optimists. Dispositonal optimists are active copers who actively seek for social support, emphasize the positive aspects of stressful events and engage in health-promoting behaviours (healthier diet habits and less smoking and drinking). Unrealistic optimists are avoidant copers who usually underestimate the importance and the negative consequences of certain events. Eventually, they have a higher likelihood of showing health problems (anxiety, distress) as a consequence of their decreased motivations to engage in health-promoting behaviours.

Therefore, I believe that a positive relationship between unrealistic optimism and unhealthy bheaviours can be observed in Russia. Following Puri and Robinson (2007), I further divided unrealistic optimism (pessimism) into two sub-categories: extreme and moderate optimism (pessimism). Thus, the optimism hypotheses read:

H1: Extreme/moderate optimism is associated with a higher probability of drinking and higher levels of alcohol consumption.

H2: Extreme/moderate optimism is associated with a higher probability of smoking and higher levels of cigarette consumption.

The DAT shows that unexpected good outcomes are pleasant, which suggests unrealistic pessimism is associated with positive emotions and psychological well-being (Shepperd and McNulty, 2002). In accordance with the TRT, alcohol and cigarettes are sought to regulate negative emotions and relieve stress (Conger, 1956; Cockerham *et al.*, 2006). Taken together, I believe that unrealistic pessimism is negatively associated with alcohol and cigarette consumption. To sum up, the pessimism hypotheses read:

H3: Extreme/moderate pessimism is associated with a lower probability of drinking and lower levels of alcohol consumption.

H4: Extreme/moderate pessimism is associated with a lower probability of smoking and lower levels of cigarette consumption.

3.3.2 Psychological ill-being channel hypotheses

Empirical studies by Cocco *et al.* (2017) and Shepperd *et al.* (2017) show that unrealistic optimism is positively associated with depressive syndromes and emotional distress. Specifically, Cocco *et al.* (2017) find that those unrealistically optimistic about their future finances show various depressive syndromes, such as difficulties facing problems, loss of sleep and report of depression. Shepperd *et al.* (2017) suggest that the association between unrealistic optimism and emotional distress is particularly strong among extreme optimists because they are more likely to experience realities that fall short of expectations.

This is in line with the DAT, which states that unexpected bad outcomes (unrealistic optimism) lead to negative emotions and even distress, while unexpected good outcomes (unrealistic pessimism) are elated and lead to positive emotions. In addition, the TRT suggests that individuals with high levels of psychological stress may attempt to relieve and regulate negative emotions through unhealthy behaviours, such as increases in drinking and smoking (Leonard and Blane, 1987). The above findings provide evidence that psychological ill-being may be one of the primary channel through which unrealistic optimism/pessimism affects unhealthy behaviours. Thus, the psychological ill-being channel hypotheses read:

H5: Negative psychological well-being may be a channel through which unrealistic optimism is positively associated with smoking and drinking.

H6: Positive psychological well-being may be a channel through which unrealistic pessimism is negatively associated with smoking and drinking.

3.4 Data and descriptive statistics

In the sub-sections that follow, I will provide a brief description of the data I used. Next, I will illustrate the measurements of dependent and independent variables. This section concludes with descriptive analyse of the sample.

3.4.1 Data

This study makes use of unbalanced panel data from round 9 to round 23 of the RLMS spanning from 2001 to 2016. During the process of data cleaning, I have made the following adjustments to the sample: Firstly, I have excluded those who answer “do not know”, “refuse to answer” or “answer not given” for questions related to this study. Next, I have restricted the sample to individuals aged above 18 because, in Russia, it is prohibited to drink under the age of 18. Third, I have excluded the top 1% of the continuous variables, such as income, consumption and cigarette and alcohol consumption. This is to prevent extreme outliers from affecting the results. Fourth, I use the Consumer Price Index (CPI) from the Federal State Statistics Service (Rosstat) of Russia to inflate the income, consumption and community prices. These variables are converted to the real value of 2016.

The primary sample of this study consists of 26,495 individuals and 120,100 observations. The mean number of years of participation per individual is 4.5, with the maximum (minimum) being 16 (1). Only 205 individuals participated in all 16 waves. Be aware that not only the head of the household but also individuals within the same household are included in the final sample. Since the conditional fixed effects Logit model drops any observation that does not vary across

time, thus a smaller sample is employed in the Logit model (binary variables that are always 1 or 0 are then dropped).

Table 3.1 displays the structure of the unbalanced panel used in this study. As the same dataset is utilised, the number of observations in Chapter 2 and 3 should be around the same. Compared Table 2.1 to Table 3.1, I can see that the number of observations are not the same in the corresponding years. This is because Chapter 3 used one more year of data for data cleaning. In addition to add one more year of data, Chapter 3 focuses on the issues of unhealthy behaviours. In the data cleaning process, this study has excluded outliers based on 16 waves of data instead of 15 waves, and dropped missing observations on smoking and drinking variables. Thus, the resulting data is different than the one used in Chapter 2[6].

3.4.2 Measures of errors in financial expectations and unhealthy behaviours

3.4.2.1 Measures of unhealthy behaviours

Binge drinking and alcohol consumption

According to Yakovlev (2018), I first calculate *pure alcohol* as the reported amount of pure ethanol consumed by each individual in each wave. Then, I make use of two binary variables related to binge drinking. *Drink200/drink150* is 1 if an individual drank more than 200/150 pure alcohol per occasion in the last 30 days, and 0 otherwise. *Drink200* and *drink150* correspond to the top 10% and top 20% of pure alcohol consumption (among drinkers)[7]. The threshold of 200

[6] A discussion regarding sample attrition can be found in section 3.6.3.

7 The reported amount of pure alcohol is calculated using the RLMS data on consumption of all types of alcohol including vodka and other hard drinks, beer, wine, champagne and samogon (home-made hard liquors) using the following formula proposed by Yakovlev (2018): Q(pure alcohol) = 0.05Q(beer) + 0.12Q(table wine) + 0.15Q(fortified wine) + 0.4Q(hard drinks) + 0.4Q(samogon).

(150) gram pure ethanol corresponds to a daily consumption of 0.5 (0.375) liters of vodka or 4 (3) liters (12 (9) bottles, 0.33 liters each) of beer.

Smoking and cigarette consumption

Based on Kelloniemi *et al.* (2005), I define *Smoke* as a dummy variable equal to 1 if an individual is a current smoker, and 0 otherwise. *Cigarette consumption* is the reported numbers of cigarettes consumed per day. It is 0 if an individual is a non-smoker. Table 3.2 presents a brief summary of the dependent variables used.

3.4.2.2 Measures of errors in financial expectations

Following Puri and Robinson (2007) and Cocco *et al.* (2017), I calculate individuals' expectation errors by taking the differences between the lagged *financial expectation (finexp)*and the contemporaneous *financial realisation (finreal)*. Four dummies for errors in financial expectations are then constructed based on the expectation errors. The related survey questions (*finexp* and *finreal*) and the four expectation errors dummies are the same as those discussed with reference to section 2.4.2.1 in Chapter 2. Table 3.3 presents the distribution of individuals' errors in financial expectations in this study.

3.4.2.3 Measures of subjective psychological ill-being

Subjective psychological ill-being usually refers to how individuals evaluate their life and is measured by variables like anxiety, depression, life dissatisfaction and emotions. Bertoni (2015) and Staudigel (2016) find that life dissatisfaction, as a component of subjective ill-being, is highly associated with the occurrence of psychological problems, and therefore has been widely used as a robust measure of psychological ill-being in empirical studies. Hence, I use life dissatisfaction (*LDS*) as a proxy for subjective psychological ill-being. In each wave of the

RLMS, individuals are asked to rate their overall life satisfaction on a 5-point scale: "To what extent are you satisfied with your life in general at the present time?". Those who answer "less than satisfied" and "not at all satisfied" are defined as experiencing some life dissatisfaction. I construct a dummy variable equal to 1 if an individual is experiencing some life dissatisfaction, and 0 otherwise.

3.4.2.4 Other control variables

Following Herzfeld *et al.* (2014), I control for other possible determinants of unhealthy behaviours in Russia. The controls include socio-demographic, health, policy and price characteristics. Socio-demographic and health controls comprise individuals' age, gender, marital status (married, cohabiting, divorced), degree of education (primary, unfinished secondary, secondary and vocational secondary school), retirement, income[8], household consumption[9], and self-reported health status.

Policy controls account for the anti-drinking and anti-smoking policies implemented by the Russian Government in recent years as an effort to cut down drinking and smoking. Radaev (2015) find that the Russian government has been launching a comprehensive anti-drinking campaigns by increasing alcohol taxation, imposing strict regulations regarding alcohol manufacturing, selling and purchasing and prohibiting alcohol consumption in public places since 2010. As for the anti-smoking campaign, Kossova *et al.* (2018) conclude that the government has

8 This is the total sum of money actually received by the individual in the last month, which includes salaries and wages from all jobs, incidental earnings, public and private transfers, retirement pensions and other receipts, including foreign currency (convert the value in rubles).
9 This represents household consumption on all goods and services in the last 3 month, including expenditures on food at home or eating out, household appliances, entertainment supplies, fuel, services (postal, transportation, repair and internet services and so on), residential rent and utilities, medical treatment and private transfers.

addressed the issues of smoking by passing a federal law aimed at regulating smoking in public places, media advertising, and preventive medical care in 2013. Therefore, a dummy variable *anti-drinking* (*anti-smoking*) is constructed and equal to 1 if the year dummy is 2010 (2013) and beyond, and 0 otherwise.

Price controls include the prices of different types of cigarettes (domestic or imported) and alcohol (vodka, cognac, table wine, fortified wine and beer) within each Primary Sampling Unit (PSU). To be more specific, I make use of the maximum and minimum prices of cigarettes and alcohol recorded through visits by the RLMS interviewers to local stores in the communities in which respondents live. However, due to some case of missing information in the community prices recordings, I compute the average prices for each PSU instead of for each community (Rizov *et al.*, 2012). Finally, I control for year and regional dummies. In line with the Federal State Statistics Service (Rosstat), the regional dummies categorise the 38 PSU into 8 regions. A full list of the variables used is presented in Appendix 3.1.

3.4.3 Descriptive statistics

According to Quirmbach and Gerry (2016) and Yakovlev (2018), Russian men are infamous for their heavy drinking and smoking. It is therefore worthwhile to present results for the full sample, as well as for the male and female sub-samples. In this section, I present summary statistics of relevant variables in the full sample, and the male and female sub-samples.

3.4.3.1 Descriptive statistics by samples

Table 3.4 presents the means and standard deviations of relevant variables for the full sample and sub-samples. I observe that the overall rates of smoking and binge (heavy) drinking are around 30.1% and 4.4% (9%). In the full sample, extreme optimism and extreme pessimism correspond

to 8.1% and 4.1% of observations respectively. This is similar to Puri and Robinson (2007)'s findings, who use the top 5% tail of the optimism/pessimism to classify extreme optimism/pessimism. The average age in the sample is 46 ,with a minimum of 18 and maximum of 102. The low proportion of males in the sample (39.5%) could be explained by the high mortality rate and short life expectancy among Russian males (Zaridze *et al.*, 2014). The sample generally supports the idea that the majority (32.1%+26.3%) of respondents hold a secondary degree. Inflated income and consumption (2016 level) are also consistent with findings of Yakovlev (2018).

Table 3.4 also presents the mean differences in relevant variables across the male and female samples. I find that the probabilities of smoking and binge drinkings are almost double (quadrupled) in the male sample compared to the full (female) sample. The differences between male and female samples are significant at the 1% level. This in line with Quirmbach and Gerry (2016) and Yakovlev (2018), who find that Russian males are more likely to engage in unhealthy behaviours than females.

On average, I find that males are significantly more optimistically biased, less pessimistically biased and less likely to make correct predictions than females. Males also consider themselves having less psychological distress than females by showing a lower level of life dissatisfaction. Although Russian men are less educated than Russian women, they generally have higher income and household consumption. Russian men also claim to be in better health status than women even though they are more likely to suffer from adverse health effects due to their high prevalence of smoking and drinking.

3.4.3.2 Descriptive statistics by errors in financial expectations

Panel A of Table 3.5 details the mean descriptive statistics of relevant variables by errors in financial expectations in the full sample. I observe that there is a general decreasing trend in the presence of the unhealthy behaviours from left to right (i.e. from extreme optimism to extreme pessimism). For example, the rates of smoking and binge drinking are 38% and 6.1% in extreme optimists, and 27% and 4.3% in extreme pessimists. In other words, extreme (moderate) optimists are far more likely to engage in smoking and binge drinking than extreme (moderate) pessimists, which is in line with H1 and H2. However, I find that the differences in the rates of smoking and drinking between moderate pessimism and correct prediction are rather small, which contradicts H3 and H4. I also observe that extreme optimists are by far the most unsatisfied group with a rate of 47.4%.

Panel B of Table 3.5 reports the mean statistics by optimism/pessimism indicators in the male sample. Compared to Panel A of Table 3.5, I observe that the overall probabilities of smoking and binge drinking in males are rather high in absolute terms, especially in extreme optimists with rates of 62% (smoking) and 12.6% (binge drinking). This is similar to Yakovlev (2018), who suggests that the problems of high mortality and strong alcohol dependency can be traced back to the Russian males. Extreme optimists also consume more cigarettes (10.07) and alcohol (87.21) than any other groups. However, unlike the prior findings in the full sample, I observe that among males, extreme pessimists have a higher prevalence of smoking and drinking than other groups. This contradicts H3 and H4, which suggest that extreme pessimists, compared to those who make correct prediction, should be associated with less smoking and drinking.

Panel C of Table 3.5 presents descriptive results for the female sub-sample. I find that, once again, extreme optimists have the highest rates of smoking (21.7%) and binge drinking (1.5%). However, I observe that the probabilities of smoking and binge drinking are a lot less compared to male counterparts. For instance, among extreme optimists, the probability of binge and heavy drinking are 1.5% and 3.7% in the female sample, while males have rates of 12.6% and 22.8% in the same categories. I also find that females are more dissatisfied towards their general life than males in all categories.

After investigating the three samples, I find consistent results that extreme (moderate) optimists have the highest probabilities of being smokers and binge drinkers. These optimistic biased people may contribute to the high mortality rate and low life expectancy in Russia (Zaridze *et al.*, 2014). I also observe that moderate pessimists and those people who make correct predictions show a lower prevalence of smoking and drinking. However, the aforementioned descriptive analyses only provide limited information as they do not control for other co-variates and thus a more sophisticated analysis is warranted.

3.5. Baseline specification and estimation strategy

This section begins with an explanation of the two main models related to the association between errors in financial expectations and unhealthy behaviours. Next, I will in turn analyse the two models testing the channel through which the association may exist. This section concludes by explaining the empirical strategy used.

3.5.1 Main empirical models

3.5.1.1 Errors in financial expectations: static

To investigate the association between errors in financial expectations and unhealthy behaviours in Russia, I initially estimate the following equation (Cocco *et al.*, 2017; Yakovlev, 2018):

Equation (3.1)

$$UB_{it} = \beta_0 + \beta_1 Extreme_optimism_{it} + \beta_2 Moderate_optimism_{it}$$

$$+ \beta_3 Extreme_pessimism_{it} + \beta_4 Moderate_pessimism_{it}$$

$$+ \beta_5 X_{it} + v_{pt} + v_i + v_t + v_r + \varepsilon_{it}$$

where UB_{it} is the unhealthy behaviours of individual *i* at time *t*, measured in turn by *drink200*, *drink150*, *smoke*, *pure alcohol consumption* and *cigarette consumption*. The subscripts *i, t, r, p,* refer to individual, year, region and PSU respectively.

In accordance with H1 and H2, the coefficients β_1 and β_2 associated with the *Extreme_optimism* and *Moderate_optimism* terms are expected to be significantly and positively associated with unhealthy behaviours in Russia. This is in line with Scheier and Carver (1992)

and Puri and Robinson (2007) who indicate that extreme and moderate optimists exhibit increasing engagement in risky behaviours due to their inactive coping strategies in response to stress.

Referring back to H3 and H4, the coefficients β_3 and β_4 associated with the *Extreme_pessimism* and *Moderate_pessimism* terms are expected to be significantly and negatively associated with smoking and drinking. This is similar to the findings of Shepperd and McNulty (2002), Kelloniemi *et al.* (2005) and Cockerham *et al.* (2006), which suggests that pessimism is associated with decreases in smoking and drinking.

X_{it} represents a vector of control variables, which includes age, gender, marital status, retirement, degree of education, individual income, consumption, self-reported health status and anti-smoking and anti-drinking policy dummies. Ruhm and Black (2002) and Yakovlev (2018) indicate that age, married and retirement result in less unhealthy behaviours, while cohabiting and divorce result in more smoking and drinking due to less responsibility bounded by the family. They also suggest that increased income and consumption are associated with higher levels of engagement of unhealthy behaviours.

Quirmbach and Gerry (2016) and Yakovlev (2018) find that males exhibit a higher prevalence of smoking and drinking compared to females. Steptoe *et al.* (2005) and Quirmbach and Gerry (2016) find that smoking and drinking rates are higher among the less educated. They also suggest that a better self-reported health status is associated with more cigarettes and alcohol consumed, due to confidence in health prospects. Radaev (2015) and Kossova *et al.* (2018) show that the anti-smoking and anti-drinking policies implemented have effectively reduced smoking

and drinking prevalence in Russia. They also find that increases in prices will increase the burden of consuming cigarettes and alcohol. Therefore, the coefficients associated with the male dummy, cohabiting, divorce, income, consumption, education variables are expected to be positive, while the signs associated with age, marriage, anti-smoking policy, anti-drinking policy, retirement and poor health status are expected to be negative.

In equation (3.1), v_{pt} controls for the PSU-level cigarette and alcohol prices. v_i controls for time-invariant unobserved individual heterogeneity. v_t controls for year fixed effects. v_r controls for regional fixed effects. ε_{it} is assumed to be an independent and identically distributed disturbance with zero mean. A full list of the variables used is defined in Appendix 3.1.

The correlation between the errors in financial expectations indicators are presented in Table 3.6. I observe that the correlations between all errors indicators are negative and significant at the 1% level. The correlation coefficient between *Extreme_optimism* and *Moderate_optimism*/*Moderate_pessimism/Extreme_pessimism* is -0.16/-0.13/-0.06 respectively. The correlation coefficient between *Moderate_optimism* and *Moderate_pessimism* is -0.23, and that between *Moderate_optimism* and *Extreme_pessimism* is -0.11, and that between *Moderate_pessimism* and *Extreme_pessimism* is -0.09. According to the rule of thumb[10], I find that the errors indicators are only negligible/weak correlated (Hinkle *et al.*, 2003).

10 According to Hinkle *et al.* (2003), the rule of thumb for correlation analysis suggests that only when the correlation coefficient exceeds 0.7, there is a strong correlation between variables, which should be treated with caution.

3.5.1.2 Errors in financial expectations: dynamic

Following Jiang and Livingston (2015), smoking and drinking have been regarded as addictive behaviours, which sometimes can be developed into a lifelong habit or even a dependence syndrome. Hence, I include the lagged dependent variables in my model to address this problem, and then examine the relationship between errors in financial expectations and unhealthy behaviours in a dynamic context. I therefore estimate the following equation, where all variables and subscripts are the same as those discussed with reference to equation (3.1).

Equation (3.2)

$$UB_{it} = \beta_0 + \beta_1 UB_{it-1} + \beta_2 Extreme_optimism_{it} + \beta_3 Moderate_optimism_{it}$$

$$+ \beta_4 Extreme_{pessimism_{it}} + \beta_5 Moderate_{pessimism_{it}} + \beta_6 X_{it}$$

$$+ v_{pt} + v_i + v_t + v_r + \varepsilon_{it}$$

3.5.2 Interaction models

3.5.2.1 Life dissatisfaction: static

In estimating whether psychological ill-being is the channel through which errors in financial expectations are associated with smoking and drinking. Following Wooldridge (2010), I estimate the following equation:

Equation (3.3)

$$UB_{it} = \beta_0 + \beta_1 Error_{it} * LDS_{it} + \beta_2 Error_{it} * (1 - LDS)_{it} + \beta_3 LDS_{it}$$

$$+ \beta_4 X_{it} + v_{pt} + v_i + v_t + v_r + \varepsilon_{it}$$

where $Error_{it}$ indexes a significant errors in financial expectations indicator from equation (3.1), and $Error_{it} * LDS_{it}$ and $Error_{it} * (1 - LDS)_{it}$ are interactions between the errors in financial expectations indicator and life dissatisfaction (LDS). X_{it} refers to a vector of control variables as well as the insignificant errors in financial expectations indicators from equation (3.1). Other variables and subscripts are the same as in equation (3.1).

When the errors in financial expectations indicator ($Error_{it}$) indexes extreme/moderate optimism, in accordance with H5, the coefficient β_1 associated with the $Error_{it} * LDS_{it}$ term is expected to be significantly and positively associated with unhealthy behaviours in Russia, while the coefficient β_2 associated with the $Error_{it} * (1 - LDS)_{it}$ term is expected to be non-significant. I also perform a Wald test to examine whether β_1 and β_2 are significantly different from each other. For H5 to hold, β_1 and β_2 are expected to be significantly different from each other, which indicates that only dissatisfied optimists exhibit increasing prevalence of smoking and drinking, while satisfied optimists show no sign of association with unhealthy behaviours. This is in line with the DAT (Shepperd and McNulty, 2002) and the TRT (Pearlin and Radabaugh, 1976), which indicate that extreme optimists who suffer from a higher level of psychological stress may attempt to relieve negative emotions through engagements in unhealthy behaviours.

When the errors indicator ($Error_{it}$) indexes extreme/moderate pessimism, for H6 to hold, the coefficient associated β_1 with the $Error_{it} * LDS_{it}$ term is expected to be insignificant, while the coefficient β_2 associated with the $Error_{it} * (1 - LDS)_{it}$ term is expected to be significantly

and negatively associated with smoking and drinking. This is similar to the findings of Shepperd and McNulty (2002) and Cocco *et al.* (2017), which suggest that satisfied pessimists show a lower prevalence of smoking and drinking.

3.5.2.2 Life dissatisfaction: dynamic

To examine H5 and H6 in the dynamic context, I estimate the following equation, where all variables and subscripts are the same with reference to equation (3.3):

Equation (3.4)

$$UB_{it} = \beta_0 + \beta_1 UB_{it-1} + \beta_2 Error_{it} * LDS_{it} + \beta_3 Error_{it} * (1 - LDS)_{it}$$

$$+ \beta_4 LDS_{it} + \beta_5 X_{it} + v_{pt} + v_i + v_t + v_r + \varepsilon_{it}$$

3.5.3 Estimation strategy

One of the known unobserved individual heterogeneity could be people's general expectancy towards life (i.e. dispositional optimism/pessimism). Since dispositional optimism/pessimism is considered as a fixed personal trait, I take advantage of the fixed effects estimator and thus prevent the time-invariant unobserved individual heterogeneity from affecting the underlying association between unrealistic optimism/pessimism and unhealthy behaviours.

Specifically, when estimating static models, I first make use of the Logit model with a fixed effects estimator for binary dependent variables (*drink200*, *drink150*, *smoke*). However, a fixed effects estimator in a non-linear model could pose the incidental parameters problem, and lead to biased estimates (Neyman and Scott, 1948). Following Chamberlain (1980), I use a conditional

fixed effects Logit estimator, which avoids the incidental parameter problems by removing the fixed effects estimation of the unobserved individual heterogeneity. Due to the non-linear nature of the method, I use odds-ratio (OR) to interpret my results (the computation of odds ratio is presented in equation (3.5)).

Equation (3.5)

$$Odds\ Ratio = \frac{Prob(y_{it} = 1|x_{it} = M)}{Prob(y_{it} = 0|x_{it} = M)} \Big/ \frac{Prob(y_{it} = 1|x_{it} = F)}{Prob(y_{it} = 0|x_{it} = F)}$$

where y_{it} denotes whether individual *i* in year *t* is a smoker or not. x_{it} could be a dummy or continuous variable. For simplicity, I will define that x_{it} equals to *M* if the individual is male and *F* if female.

Equation (3.6)

$$\frac{Prob(y_{it} = 1|x_{it} = M)}{Prob(y_{it} = 0|x_{it} = M)}$$

Equation (3.7)

$$\frac{Prob(y_{it} = 1|x_{it} = F)}{Prob(y_{it} = 0|x_{it} = F)}$$

Equation (3.6) is the odds of smoking by male *i* in year *t*, and equation (3.7) is the odds of smoking by female *i* in year *t*. Taken together, equation (3.5) represents the odds ratio denoting the odds of smoking by male over female[11].

[11] Buis (2010) concludes that since the interaction term is a multiplicative term, the odds ratio or incidence-rate ratio (IRR) has an edge over marginal effect in terms of interpretation of the interaction effect. One can directly

One major drawback of the conditional fixed effects Logit estimator is that it only uses individuals who switch state over time. As a result, several observations are lost and cannot be used in estimation. Hence, I use the linear probability model (LPM) with fixed effects as an additional method.

Subsequently, I use a fixed effects linear regression to estimate the association between cigarette consumption (count variable) and errors in financial expectations. Consider the non-negative, integer-valued nature of a count variable, similar to Teng *et al.* (2017) and Nketiah-Amponsah *et al.* (2018), I utilise a fixed effects negative binomial estimator for cigarette consumption as an additional robustness check.

Next, I use a fixed effects linear regression model to examine the correlation between pure alcohol consumption (continuous variable) and errors in financial expectations. In addition, I follow Yakovlev (2018) and Machado and Silva (2019) and make use of the fixed effects quantile regression, where I model different quantiles in the entire conditional distribution of y given x. Specifically, I estimate the differential effects of errors in financial expectations on different quantiles of the distribution of pure alcohol consumption.

Following Arellano and Bover (1995) and Blundell and Bond (1998), I employ a system generalized methods of moments (GMM) estimator to estimate the dynamic model. The GMM estimator can control for individual-specific and time-invariant heterogeneity, omitted variable bias, as well as the possible endogeneity of the independent variables[12]. This estimator is

---

interpret the interaction term as OR or IRR without going over the trouble to compute the cross partial derivative of E(y) with respect to x_1 and x_2.

[12] Maeshiro (1996, 1999) suggests that the presence of lagged dependent variable poses several issues. Firstly, a lagged dependent variable in the right-hand side leads to serial correlation of the error term. Secondly, the

designed for panel dynamic model with many individuals and few time periods (large N and small T). The system GMM builds up a system of two equations (the first-differenced equations and the level equations). Unlike the first-difference GMM estimator that suffers from finite sample properties, the system GMM estimator has shown a dramatic improvement in the precision and the efficiency of the estimator[13]. The system GMM estimator uses first-differencing to control for unobserved heterogeneity, and lagged values of the independent variable as instrumental variables to control for the potential endogeneity of the respective regressors (Roodman 2009).

There are two diagnostic tests utilised in GMM models to examine the validity of instruments and the correct specification of the model. The first is the Hansen J-test for over-identifying restrictions, which follows a *Chi-square* distribution with the value of degrees of freedom equal to the number of over-identifying restrictions (number of instruments less number of parameters). The null hypothesis of this test suggests that the instruments are uncorrelated with the error term or are being correctly included in the model. The second is the tests for the n^{th}-order serial correlation of the differenced residuals (*m(n)*). This test provides a further test for the legitimacy of the instruments and the validity of specification of the model. The null hypothesis of *m(n)* indicates that there is no n^{th}-order serial correlation of the differenced residuals. A

---

lagged dependent variable is as stochastic as the dependent variable, which violates the classic assumption of the linear models that both independent variables and the error term are uncorrelated with each other. Thus, using conventional linear models to estimate a dynamic model will lead to biased and inconsistent estimates.

[13] According to Blundell and Bond (1998) and Blundell *et al.* (2001), when an independent variable is persistent over time and the number of time periods is relatively small, the first-difference GMM estimator has been shown to have a large finite sample bias and perform poorly in simulation practices. On the other hand, the system GMM largely decreases the finite sample bias and is more asymptotically efficient as it is not dependent on relatively harsh restrictions on the initial condition.

rejection of the null hypothesis of *m(n)* implies that the instruments need to be lagged at least *n+1* times. In this study, GMM models generally reject the null hypothesis of no third-order autocorrelation, thus levels of endogenous variables dated *t-4* and further are used as instruments in the first-differenced equations, and first-differences of the endogenous variables dated *t-3* are used as additional instruments in the level equations (Roodman, 2009).

Following Wooldridge (2010), I perform a series of endogeneity tests regarding some potential endogenous variables, such as the extreme and moderate optimism/pessimism[14]. I use the first lag of these variables as self-instruments, and find no evidence of endogeneity in these variables[15][16]. For this reason, in my GMM models, I only instrument the lagged dependent variable with its own lags.

[14] The endogeneity test statistics used are the difference of two Sargan - Hansen statistics: one for the equation with the smaller set of instruments, where the suspect regressors are treated as endogenous, and one for the equation with the larger set of instruments, where the suspect regressors are treated as exogenous. The null hypothesis of this test is that the regressors deemed endogenous are in fact exogenous.

[15] The null hypothesis of endogeneity test is rejected, which suggests that the suspect endogenous variables can be treated as exogenous variables.

[16] Additionally, I use the first and second lags of the four key independent variables for the enogeneity tests. I find consistent evidence that the instruments for the four key independent variables (extreme/moderate optimism/pessimism) are not weakly-identified and valid. The null hypothesis of the endogeneity test is not rejected, which suggests that the four key independent variables can actually be treated as exogenous variables.

3.6 Empirical Results

In this section, I use a variety of estimation methods to document the relationship between errors in financial expectations and unhealthy behaviours. This section starts off with results for the probability of drinking and alcohol consumption, as well as results for the probability of smoking and cigarette consumption. It then reports results showing the interaction effects of life dissatisfaction on the association between errors in financial expectations and unhealthy behaviours. This section concludes with results testing whether this study suffers from sample attrition bias.

3.6.1 Main results

3.6.1.1 Regression results for drinking

Binge drinking: Results for static models

Table 3.7 presents the results for binge drinking. Columns 1-3 present the results from the fixed effects LPM models, and columns 4-6 present the conditional fixed effects Logit estimates.

Focusing on the LPM results, I find that the estimated coefficients on the extreme optimism are significant and positive (0.0055 and 0.0182) in both the full sample and male sample. In accordance with H1, I find consistent evidence that extreme optimism is associated with a higher probability of binge drinking in Russia, and it is significant at the 5% level in the full sample, and 1% level in the male sample.

Unlike previous literature in dispositional optimism (Kelloniemi *et al.*, 2005; Wray and Dvorak, 2013), I find consistent evidence that dispositional and unrealistic optimism have the opposite prediction power in predicting health behaviours. Specifically, I find a positive

relationship between unrealistic optimism (e.g., extreme optimism) and binge drinking, which is in line with Shepperd *et al.* (2017).

There are two possible explanations for this relationship. First, in line with Dillard *et al.* (2009) and Floyd *et al.* (2000), unrealistic optimists (extreme optimists) tend to underestimate the potential risks of unhealthy behaviours. Thus they are more likely to take unnecessary risks such as abuse of risky behaviours due to ignorance of serious consequences.

Second, this result could be explained by the Decision Affect Theory (Shepperd and McNulty, 2002) and the Tension Reduction theory (Conger, 1956; Ikard *et al.*, 1969). Decision Affect theory suggests that when extreme optimists find out that their outcomes fall short of expectations, it may leads to psychological ill-being in the form of anxiety and depressive syndromes. Furthermore, in conjunction with the Tension Reduction theory, depressed optimists are likely to turn to alcohol as self-medication to relieve and regulate negative emotions and psychological distress. These theories provide a plausible mechanism linking extreme optimists to binge drinking.

However, in columns 1-3, I find no evidence of negative associations between extreme/moderate pessimism and binge drinking, which contradicts H3. According to the DAT, people who experienced unexpected good outcomes (extreme pessimists) are generally associated with positive affections. As pessimists are satisfied with their final outcomes, they are likely to stay on their current alcohol consumption level with little fluctuating.

Columns 4-6 of table 3.7 present the estimated odds ratio obtained using the conditional fixed effects Logit models. The results show that extreme optimism is a significant determinant of binge drinking. Specifically, I find that for extreme optimists, the odds of binge drinking are

1.162 times greater than the odds of binge drinking for non-extreme optimists (the reference group) in the full sample. The significant and positive relationship between extreme optimism and binge drinking can also be found in the male sample, with an odds ratio of 1.171. In line with results of the LPM models, I find no significant relationships linking moderate optimism, extreme/moderate pessimism and binge drinking.

Turning to control variables, referring to columns 3 and 6 of Table 3.7, I observe that there are virtually no significant relationships except for retired in the female sample. This could be explained by the fact that only 759 out of a total of 71,903 females exhibit binge drinking behavior. In addition, I find that income and divorce are positive determinants of binge drinking, while the anti-drinking policy, higher education and poor health status have negative effects on binge drinking. These control variables have the same expected signs as discussed in section 3.5, with significance at the 5 percent level in the full and male samples.

Problem drinking: Results for static models

Table 3.8 presents the estimates when I use a dummy equal to 1 if the respondent consumes more than 150 gram pure ethanol per occasion in the last 30 days, and 0 otherwise, as a left-hand side variable. Consistent with H1, I find a positive and significant relationship between extreme optimism and problem drinking in the whole group as well as in the male sub-group. Columns 1 and 2 show that extreme optimism display a 0.65 and 2.07 percentage point higher chance of showing problem drinking in the full and male samples respectively.

In the Logit models, column 4 of Table 3.8 shows that for extreme optimists, the odds of problem drinking are 1.111 times higher compared to the odds of problem drinking by non-extreme optimists in the full sample. In the male sample, column 5 of Table 3.8 reports the odds

ratio associated with extreme optimism are 1.186. In contrast with H3, I find no significant relationships between moderate optimism and extreme/moderate pessimism and problem drinking across all columns in Table 3.8.

I observe that the association between extreme optimism and problem drinking is more profound in the male sample (significant at the 1% level) than in the full sample (significant at the 5% level) or in the female sample (non-significant). This could be explained by the findings in Yakovlev (2018), which suggests that alcohol abuse is a critical issue in Russia. In particular, there is a strong habit persistence in alcohol use among Russian male, which indicates that alcohol is the go-to behaviours in coping with financial stress and crisis (Stuckler and Basu, 2013). Precisely, the long history of heavy drinking among Russian men could explain why the association between optimism and drinking is stronger among male than female.

Looking at the controls, I find most of the control variables are significant determinants of problem drinking. More specifically, I find that income, consumption and divorce are all significantly and positively related to problem drinking, while the anti-drinking policy, retired and poor health are associated with a lower chance to show problem drinking. These relationships do not vary much across the LPM and Logit models.

Binge and problem drinking: Results for dynamic models

Table 3.9 reports the estimates of the dynamic models for drinking, obtained using the system GMM estimators[17]. Columns 1-3 report the results for binge drinking and columns 4-6, those for

[17] Unlike continuous variable, binary variable takes one of two values: 0 or 1 and is clearly not normally distributed. The non-linear nature of a binary variable may pose concern over its distributional assumption of the errors. Wooldridge (2010) suggests that GMM estimator does not pose distributional assumptions on the error term, and any dummy variable is treated as all other variables in the first-differenced equations without causing any further problems. This indicates that GMM is appropriate for modelling a binary outcome.

problem drinking. As can be seen from column 1, in line with H1, extreme and moderate optimism is significantly and positively associated with binge drinking in the full sample, with coefficients of 0.0087 and 0.0031 which are statistically significant at the 1 and 5 percent level respectively. In column 4, I find consistent evidence that extreme optimism is associated with a 0.96 percentage point higher chance of being a heavy drinker.

Comparing columns 2 and 5, I see that extreme optimism is a significant determinant of both binge and problem drinking in the male sample. In more detail, extreme optimism is positively associated with binge drinking, with a coefficient of 0.0206 which is statistically significant at the 1 percent level. Extreme optimism is also significantly and positively associated with problem drinking, with an estimated coefficient of 0.0242 and significance at the 1% level.

In column 2 of Table 3.9, I observe that moderate optimism is also a significant determinant of binge drinking in the male sample, with an estimated coefficient of 0.0067 which is significant at the 10 percent level. However, compared to extreme optimism, moderate optimism is a significant determinant of binge drinking though to a lesser degree. This is similar to findings from Armor and Taylor (1998), which suggests that extreme and moderate optimists are both subject to psychological ill-being after failing to reach their expectations. This emotional let-down would be enlarged if individuals' predictions are too extreme (e.g., extremely optimistic). Therefore, extreme optimists would show a greater level of disappointment and negative emotions than moderate optimists, which leads to engagement in greater amount of alcohol consumption to relieve negative emotions (Pearlin and Radabaugh, 1976).

The above estimates are based on GMM short-run estimation, measuring the immediate effects of expectation errors on binge and problem drinking. Additionally, the GMM models can

be used to estimate the long-run effects of errors in financial expectations on drinking[18]. More specifically, in Table 3.9, the long-run coefficients and standard errors are presented in the square bracket. In columns 1 and 2 of Table 3.9, I find consistent evidence that extreme (moderate) optimism has a positive long-run effect on binge drinking in the whole and male samples, with estimated coefficients of 0.0087 and 0.0206 (0.0031 and 0.0067) and significance at the 1 percent level. In columns 3 and 4 of Table 3.9, I also observe that in the long-run, the relationship between extreme optimism and problem drinking is significant at the 5 percent level or beyond, with estimated coefficients of 0.0096 and 0.0242 in whole and male samples respectively. Thus, I find consistent evidence that extreme and moderate optimism are significant determinants of binge and problem drinking both in the short-run and long-run.

Just like the prior findings, across all columns, extreme pessimism and moderate pessimism are not the significant determinants of binge and problem drinking, which contradicts H3. I further find that age, the anti-drinking policy, poor health and retirement have negative coefficients while male, income marital status have positive coefficients. These control variables are generally significant at the 1% level in all sampling groups. However, I observe that education variables have negative coefficients in static models, and positive coefficients in dynamic models. This could be attributed to the fact that static model fails to capture the habitual nature of drinking behaviour and is thus affected by this dynamic context.

To sum up, I find robust evidence suggesting that extreme optimism is positively and significantly associated with binge and problem drinkings in the full sample and male sample.

[18] I only present the long-run coefficients and standard errors for my four key independent variables for brevity.

Moderate optimism is also a positive determinant of the chance of showing binge drinking in the full sample, as well as in the male sample.

3.6.1.2 Regression results for alcohol consumption

Table 3.10 presents the estimated results of the fixed effects quantile regressions in the male sample. Initially, I use the fixed effects quantile model to examine the association between errors in financial expectations and alcohol consumption in the full, male and female samples respectively. I run separate regressions for the 10^{th}, 30^{th},50^{th}, 70^{th}, 90^{th} percentiles of the distribution of alcohol consumption. However, I only find significant coefficients of the errors variables in the male sample, and I report the quantile results of the males sample here. The quantile results of the full and female samples are presented in Appendix 3.2 and 3.3 respectively. In columns 1 and 2 of Table 3.10, I show that extreme optimism is not significantly associated with pure alcohol consumption at the 10^{th} and 30^{th} percentiles.

As I move along the distribution of pure alcohol consumption, I find that the association between extreme optimism and alcohol consumption is stronger in higher percentiles. At the 50^{th} percentiles, I find that extreme optimism is positively and significantly associated with alcohol consumption, with a coefficient of 5.242. This implies that extreme optimists are associated with a 5.242 units higher pure alcohol consumption compared to non-extreme optimists at the 50^{th} percentile. At the 70^{th} percentiles, extreme optimism is still highly significant at the 5% significance level with an estimated coefficient of 7.443.

Moving towards the 90^{th} percentile, the extreme optimism coefficient always remains highly significant at the 5% level. I find that extreme optimists are associated with almost twice the amount (9.170 units) of the alcohol consumption compared to the extreme optimists in the

median (5.242 units). This suggests that the associations between extreme optimism and alcohol consumption is stronger as I move from the lower to the higher end of the distribution of alcohol consumption. This could also explains why the extreme optimism coefficient is significant at the 1% level in the regressions for binge drinking, but only significant at the 5% level for problem drinking in the full sample. This finding is partially in line with H1. It is also consistent with with Steptoe *et al.* (2006), who suggest that the effects of extreme optimism could change dramatically by imposing different cutoff points in alcohol consumption.

However, I find no significant evidence linking extreme/moderate pessimism and alcohol consumption. This is in contrast to H3, which suggests that a negative relationship should be found. Looking at the controls, I find that as I move along the distribution of alcohol consumption, the coefficients associated with income, cohabiting and divorce are significant and increasing along the distribution, while the coefficients associated with poor health are significant and decreasing along the distribution. The coefficients associated with other control variables are not statistically significant.

Table 3.11 reports the estimates of the static and dynamic models of alcohol consumption. In columns 1-3, I report the fixed effects linear regression models. I find that extreme optimism is significantly and positively related to alcohol consumption in the male sample. To be more specific, I see that being an extreme optimist is associated with a 3.736 units higher pure alcohol consumption. This is in line with H1. However, I find no evidence of this association in the full samples.

Column 3 of Table 3.11 reports the fixed effects results of the female sample. In contrast to H1, extreme optimism is associated with a 1.310 units lower alcohol consumption, which is

significant at the 5% level. This is in line with Cockerham *et al.* (2006), who suggests that psychological stress does not translate into alcohol drinking in Russian women because of two reasons. First, there are gender differences in socially acceptable behaviours in post-Soviet societies. In particular, drinking (heavy drinking) is deemed as a masculine (not feminine) behaviour in the societies. Thus, the social norms proscribe female drinking (Van Gundy *et al.*, 2005). Second, women carry more social roles and family obligations than men in terms of continuing on with family daily life and taking care of family members in spite of personal hassles and distresses. Although female extreme optimists are likely to face negative emotions when their expectation fall short (Shepperd and McNulty, 2002), they are less likely to relieve their pressures by alcohol consumption. This is because alcohol consumption is more time-consuming and may lead to disability to perform their domestic obligations (e.g., childcare, errand). I also find that moderate pessimism is significant at the 10% level, with a coefficient of 0.743. This is in contrast to H3. However, the magnitude of the coefficient is rather small.

In columns 1 and 2 of Table 3.11, I find no significant relationships linking moderate optimism, moderate (extreme) pessimism and alcohol consumption in the full and male samples. As previously suggested, pessimists are satisfied with their financial situations, and are not likely to engage in negative emotions due to financial let-down. However, pessimists are not associated with significant decreases in alcohol consumption as H3 suggested. This maybe explained by the fact that drinking is often considered by Russian men as a positive social outlet and some of them even believe that it is beneficial for their health (Cockerham *et al.*, 2006). Yeltsin (2002) indicates that the mode of drinking is common to Russian male as part of their health lifestyle. This lifestyle is to drink with one's friends as long as one is able or until the supply of alcohol has

run-out. Van Gundy *et al.* (2005) point out that workers will not normally drink large amount of alcohol but rather drink to socialize. This lifestyle implies a steady amount of alcohol consumption throughout the year. However, the amount of alcohol consumption could be enlarged if one's affected by psychological distress as discussed in previous example (e.g., extreme optimism).

Most of the control variables are significant at the 5% level and beyond. I observe that married males show more alcohol consumption, while married females show less consumption. The anti-drinking policies are associated with large decreases in alcohol consumption (61.51) in the male sample. The corresponding figure for the female sample is 49.62.

In columns 4-6 of Table 3.11, I report the system GMM estimates of the dynamic models. In line with H1, I find that in short-run, extreme optimists are associated with a 4.552 units higher pure alcohol consumption in the male sample, which is significant at the 5% level. Extreme/moderate pessimism is insignificantly associated with alcohol consumption in all samples, which contradicts H3. In the long-run, I observe consistent evidence that extreme optimism has a positive and significant effect on pure alcohol consumption in the male sample. Control variables generally show similar significant levels and signs as in columns 1-3 of Table 3.11.

The above findings conclude that whether I use a static or a dynamic approach, I find consistent evidence that extreme optimism is associated with higher levels of alcohol consumption. This result is more profound in the male sample. This is in line with Leonard and Blane (1987) and Shepperd and McNulty (2002), who suggest that individuals may attempt to

relieve and regulate negative moods from fall-short of expectations through increases in alcohol consumption.

3.6.1.3 Regression results for smoking

Table 3.12 presents the estimates of equation 3.1 for the probability of smoking. In columns 1-3, I make use of the fixed effects LPM models. In contrast to H2, I find that extreme and moderate optimism is insignificantly associated with the probability of smoking in the full and male samples. Column 3 reports that extreme optimists are associated with a 0.64 percentage point higher chance of being a smoker in the female sample, with a 10% significance level. I also find that extreme and moderate pessimism is insignificantly associated with smoking in all samples, which contradicts H4.

Looking at columns 4-6 of Table 3.12, I find that extreme optimism is significantly and positively associated with the probability of smoking in the full and female samples. In column 4, in line with H2, I find that for extreme optimists, the odds of smoking are 1.165 times higher than the odds of smoking by non-extreme optimists in the full sample. The odds ratio associated with extreme optimism is significant at the 5% level. Column 6 also reports that the odds ratio associated with extreme optimism is 1.263 and significant at the 5% level in the female sample.

However, in contrast with H4, I find that extreme and moderate pessimism is insignificantly related to smoking in all samples, regardless of the estimator used. As for the control variables, I find that the anti-smoking policy, being married, poor reported health and retirement are negatively associated with smoking, while age, income and consumption are positively associated with smoking. These variables are generally significant at the 1% level. Others control variables are insignificant.

Table 3.13 presents the estimates of equation 3.2, obtained using the system GMM estimator. In column 1, I find that extreme optimism is significantly and positively associated with the probability of smoking in the full sample. Extreme optimists are associated with a 1.21 (4.13) percentage point higher probability of smoking in the short-run (long-run). The coefficient associated with extreme optimism is significant at the 1% level. This is in line with H2. Column 3 also shows that extreme optimism is highly significant at the 1% level, with estimated coefficients of 0.0136 (0.0472) in the female sample in the short-run (long-run). This is in line with Cockerham *et al.* (2006) who suggest that Russian women respond to psychological stress mainly through smoking instead of drinking due to social norms. Their study shows that Russian women always put the need of their family before their own, and perform almost all the domestic labor and childcare at home. Family obligations restrains the time Russian women have to properly deal with their personal strains. Hence, smoking as a less-time consuming activity, is preferred over alcohol drinking. This provides a potential explanation that in the face of a financial let-down, females are more prone to use cigarettes to regulate their negative emotions, and is in line with our findings that extreme optimism is positively associated with smoking in female sample.

Looking at the pessimism variables, I find that moderate pessimism and extreme pessimism is insignificantly associated with smoking in all samples, which contradicts H4. I also find that the lagged smoking is highly significant at the 1% significance level in all samples, which suggests persistence. Therefore, using a dynamic model may be a better fit for studies on smoking. As for control variables, the anti-smoking policy is significantly and negatively associated with smoking, which suggests that the policy does its job and prevents people from

smoking. The coefficients associated with divorce and education variables are positive and significant at the 1% level, which implies that smoking is more prevalence among the low educated and the divorced. Negative associations between poor health status and retirement and smoking are found and are consistent with results from the static models.

3.6.1.4 Regression results for cigarette consumption

Table 3.14 presents the estimates of the baseline specification of equation 3.1 including cigarette consumption as the unhealthy behaviour. In columns 1-3, I present regression results obtained from the fixed effects linear regressions, and I find that optimism and pessimism are not related to the number of cigarettes smoked in all samples. This contradicts H2 and H4. Most of the control variables are significant and have the same expected signs as discussed in section 3.5, except for the education variables whose coefficients are not significant.

Columns 4-6 of Table 3.14 present the estimates results for the fixed effects Negative Binomial models. Column 4 reports that extreme and moderate optimism is significantly and positively associated with cigarette consumption in the full sample. Specifically, the number of cigarettes smoked by extreme (moderate) optimists are 1.045 (1.022) times greater than the number of cigarettes consumed by non-extreme (non-moderate) optimists. This is in line with H2.

In columns 5 of Table 3.14, I find that optimism and pessimism are insignificantly associated with cigarette consumption in the male sample. This is consistent with findings for smoking using the LPM and the Logit models (Table 3.12, columns 2 and 5). In line with H2, column 6 reports that extreme optimism is positively and significantly associated with cigarette consumption in the female sample, with an estimated IRR of 1.089. I also find that the IRR associated with extreme pessimism is 0.903 and significant at the 5% level. This is in line with

H4, which implies that the number of cigarettes smoked by extreme pessimists is 0.903 times lower than the number of cigarettes consumed by non-extreme pessimists. Most of the other control variables are significant at the 1 percent level and have the same expected signs as I discussed in section 3.5.

Table 3.15 presents the regression results of cigarette consumption, obtained using the system GMM models. In column 3, I find that extreme optimism is significantly associated with a higher level of cigarette consumption in the female sample. Consistent with H2, I observe that being an extreme optimist is associated with a 0.178 (0.535) units higher cigarette consumption in the short-run (long-run), which is significant at the 1% level. In addition, I find no other evidence suggesting a significant relationship between pessimism variables and cigarette consumption, which contradicts H4. In line with previous findings, I see that the lagged cigarette consumption is a significant determinant of the current cigarette consumption, which suggests that smoking behaviour is habitual. As for control variables, I find that the coefficients associated with age, education, poor health and retired status are negative while the coefficients associated with income, consumption and divorced are positive. These control variables are significant at the 5% level.

To sum up, using a variety of static and dynamic models, I find consistent evidence that extreme optimists show a higher probability of smoking and high levels of cigarette consumption in the female sample. This finding is in line with Shepperd and McNulty (2002) and Cockerham *et al.* (2006), who suggest that extreme optimism is associated with higher level of stress, and in response to stress, Russian women tend to use cigarettes to relieve their pressures.

3.6.2 Interaction results

3.6.2.1 Interaction results for drinking and alcohol consumption

Binge and problem drinkings: Interaction results

Table 3.16 reports the estimation results of equation 3.3 for the probability of binge and problem drinkings[19]. Columns 1-3 of Panel A present the estimates of equation 3.3 using the LPM models. *EOPT_LDS* represents dissatisfied extreme optimists, while *EOPT_(1-LDS)* represents satisfied extreme optimists. In order to investigate the role of life dissatisfaction on the relationship between extreme optimism and drinking behaviours, I examine whether higher level of life dissatisfaction is associated with a higher prevalence of drinking among extreme optimists.

Columns 1 and 2 of Panel A show that the coefficients associated with *EOPT_LDS* are significantly and positively associated with binge drinking in the full and male samples, with estimated coefficients of 0.0075 and 0.0228 respectively. This is consistent with H5 suggesting that extreme optimists with higher level of life dissatisfaction are associated with a higher probability of binge drinking. Supporting H5, columns 4 and 5 also report that dissatisfied extreme optimists are associated with a 1.31 and 3.23 percentage point higher probability of problem drinking the the full and male sample respectively.

However, in columns 3 and 6 of Panel A of Table 3.16, the interaction terms *EOPT_(1-LDS)* are significant at the 10% level with estimated coefficients of -0.0042 and -0.0071, while the coefficients associated with *EOPT_LDS* are non-significant. In contrast to H5, this finding

19 The estimation results from equation (3.1) show that only *extreme optimism* is a significant determinant of all five measures of unhealthy behaviours in all samples, while *moderate pessimism* (*extreme pessimism*) is also a significant determinant of *pure alcohol consumption* (*cigarette consumption*) in the female sample. Therefore, I only focus on testing the channel through which a significant errors in financial expectations indicator is associated with unhealthy behaviours in the interaction models.

suggests that female dissatisfied optimists are not significantly and positively associated with binge and problem drinkings. This is in line with Cockerham *et al.* (2006) who suggest that Russian females do not translate stress into drinking behaviours due to social norms and family affairs.

Columns 1-6 of Panel B of Table 3.16 present the estimates obtained using the conditional fixed-effects Logit models. In columns 1 and 2, I find that the coefficients associated with *EOPT_LDS* are significant at the 10 percent level with odds ratios of 1.188 and 1.284 in the full and male samples. This is in line with H5, which implies that dissatisfied optimists display a higher probability of binge drinking. Looking at columns 4 and 5, I find that dissatisfied optimists are significantly and positively related to the probability of problem drinking in the full and male samples. Specifically, the odds of problem drinking for dissatisfied optimists are 1.221 (1.287) times as large as the odds of that for non-optimists in the full (male) sample. In line with H5, this result provides evidence showing that life dissatisfaction has a positive mediating effect on the relationships between extreme optimism and binge and problem drinkings.

In terms of dynamic models, Panel C of Table 3.16 reports the estimation results of equation 3.4 using the system GMM estimator. Columns 1 and 2 show that, in the full and male samples, the short-run (long-run) coefficients associated with *EOPT_LDS* always remain significant at the 5 percent level with coefficients of 0.0094 and 0.0212 (0.0097 and 0.0225). Consistent with H5, this finding suggests that dissatisfied optimists are associated with a higher probability of binge drinking when compared to non-optimists in the full and male samples. However, column 2 reports that both *EOPT_LDS* and *EOPT_(1-LDS)* have positive relationships with binge drinking. The short-run coefficients associated with *EOPT_LDS* and *EOPT_(1-LDS)* are not significantly

different from each other (Wald test, p=0.599), which implies that life dissatisfaction may not be the channel through which extreme optimism is associated with binge drinking in the male sample. This signals the need to find some other channels to explain the positive association between extreme optimism and binge drinking in the male sample.

Columns 4 and 5 of Panel C of Table 3.16 provide consistent evidence that *EOPT_LDS* is highly significant at the 1% level. In the short-run, I observe that dissatisfied optimists are associated with a 1.77 (3.49) percentage point higher chance of being a problem drinker in the full (male) sample. I also find that dissatisfied optimists are associated with a 2.26 (3.89) percentage point higher chance of being a problem drinker in the full (male) sample in the long-run. Compared to the static models, I find consistent evidence that life dissatisfaction is significantly associated with higher prevalence of binge and problem drinkings among extreme optimists both in the short-run and long-run, while life satisfaction plays no part in the associations between extreme optimism and binge and problem drinkings. This is in line with Armor and Taylor (1998) and the DAT, which implies that individuals are subject to greater level of disappointments and exposed to higher level of psychological distress when they find out their outcomes fall short of expectations. The TRT suggests that strained individuals tend to use alcohol to regulate negative emotions, and this is especially the case for Russian males considered their long history of binge drinking (Cockerham *et al.*, 2000).

Alcohol consumption: Interaction results

Table 3.17 presents the interaction results for alcohol consumption. Columns 1-3 of Panel A report the fixed effects linear models results. In columns 1 and 2, I find that dissatisfied optimists are significantly and positively associated with pure alcohol consumption in the full and male

samples. The coefficient associated with *EOPT_LDS* is 6.564 (1.985) and significant at the 1% (10%) level in the male (full) sample. These results are consistent with H5, however I find a violation of H5 in the female sample as a negative relationship between *EOPT_(1-LDS)* and alcohol consumption is observed. This is consistent with Van Gundy *et al.* (2005)'s finding, which suggests that Russian women do not turn to alcohol when facing financial or psychological constraints like men do.

Columns 4-6 of Panel A of Table 3.17 present the GMM estimates of equation 3.4 for alcohol consumption. In columns 4 and 5, similar results are observed in the short-run and long-run. Dissatisfied optimists are associated with a 3.333 and 9.019 (10.606 and 18.655) units higher alcohol consumption in the full and male samples in the short-run (long-run), with a significant level of 5 percent or beyond. This finding indicates that among extreme optimists, only those who are dissatisfied with their life display higher levels of alcohol consumption, while satisfied individuals are not associated with alcohol consumption at all. In line with H5, I find that life dissatisfaction may be one of the channel through which extreme optimism is positively associated with alcohol consumption.

Panel B, C and D of Table 3.17 present the interaction results of fixed effects quantile regressions for alcohol consumption. Panel B and D suggest that there are no significant relationships between *EOPT_LDS* and alcohol consumption in the full and female samples. However, focusing on Panel C, a positive relationship between *EOPT_LDS* and alcohol consumption is documented in the male sample. As I move along the distribution of alcohol consumption, I observe that the association between *EOPT_LDS* and alcohol consumption is only significant at the higher end (70^{th} and 90^{th} percentiles) of the distribution. At the 70^{th} percentiles,

dissatisfied optimists are associated with a 12.66 units higher alcohol consumption in the male sample. At the 90th percentiles, I find consistent evidence showing that dissatisfied optimists are associated with a 14.73 units higher alcohol consumption in the male sample. These results confirm H5, and suggest that as I move along the distribution of alcohol consumption, the association between *EOPT_LDS* and alcohol consumption only exists among heavy drinkers (who are at the 70th and 90th percentiles).

3.6.2.2 Interaction results for smoking and cigarette consumption

Smoking: Interaction results

Table 3.18 presents the interaction results for smoking. Columns 1-3 of Panel A of Table 3.18 report the estimation results of fixed effects LPM for smoking. Column 1 shows that the interaction term *EOPT_LDS* is significant at the 10 percent level, with an estimated coefficient of 0.0071. In support of H5, this result suggests that dissatisfied optimists are associated with a 0.71 percentage point higher chance of smoking in the full sample. A similar result is found in the female sample. Looking at column 3, dissatisfied optimists are associated with a 1.01 percentage point higher probability of smoking in the female sample. However, I fail to find a significant relationship between *EOPT_LDS* and smoking in the male sample. This could be explained by the findings of Van Gundy *et al.* (2005) and Quirmbach and Gerry (2016), which suggests that gender differences in socially acceptable behaviours may factor into people's decisions of unhealthy behaviours. In Russia, due to the social norms and family obligations, females tend to engage into smoking instead of drinking, and in reverse situation for males. Therefore, I find similar results indicating that the association between *EOPT_LDS* and smoking is observed in the female sample instead of in the male sample.

Columns 1-3 of Panel B of Table 3.18 report the Logit estimation results for smoking. In columns 1 and 3, the *EOPT_LDS* terms are significant at the 5 percent level in the full and female samples. The odds of smoking for dissatisfied optimists are 1.246 and 1.408 times higher than the odds of smoking for non-optimists in the full and female samples respectively. These results are in line with the aforementioned explanations, and with H5.

Columns 1-3 of Panel C of Table 3.18 present the system GMM estimates of equation 3.4 for smoking. Columns 1 and 3 show that the both the short-run and long-run coefficients associated with the *EOPT_LDS* terms are significant at the 5 percent level in the full and female samples. In the short-run (long-run), dissatisfied optimists are associated a 1.45 and 1.93 (5.25 and 6.96) percentage point higher probability of smoking in the full and female samples respectively. Supporting H5, I find consistent evidence showing that among extreme optimists, life dissatisfaction is associated with a higher probability of smoking. Therefore, this is in line with findings of Cockerham *et al.* (2006), Van Gundy *et al.* (2005) and Quirmbach and Gerry (2016), which suggest that Russian females translate their financial and psychological stress into smoking instead of drinking.

Cigarette consumption: Interaction results

Table 3.19 presents the interaction results for cigarette consumption. Panel A, B and C report the estimated results for the full, male, and female samples respectively. Column 1 of Panel A presents the fixed effects LR results in the full sample, and I find that the *EOPT_LDS* term is significant at the 5% level with an estimated coefficient of 0.152, while the *MOPT_LDS* term is significant at the 10% level with an coefficient of 0.103. In accordance with H5, I observe that

dissatisfied extreme and moderate optimists are significantly and positively associated with higher cigarette consumption.

Column 2 of Panel A of Table 3.19 presents the fixed effects NB results in the full sample, and shows that the number of cigarettes smoked by dissatisfied optimists are 1.060 times greater than the number of that by non-optimists. Column 3 of Panel A of Table 3.19 presents the GMM estimates and finds that there is a significant relationship between *EOPT_LDS* (*MOPT_LDS*) and cigarette consumption in the full sample in the long-run instead of in the short-run. This contradicts H5, and suggests that the positive relationships observed in the static model could be stronger once accounting for habit formation in the long-run.

Panel B of Table 3.19 presents the static and dynamic results of the male sample for cigarette consumption. In line with the previous findings, I observe no significant relationship between *EOPT_LDS* and cigarette consumption in the male sample as Russian men tend to translate their stress into drinking instead of smoking.

Panel C of Table 3.19 presents the estimated results of the female sample for cigarette consumption. In column 1, I find that dissatisfied pessimists are associated with a 1.28 percentage point lower cigarette consumption using the fixed effects LR models. Thus, H6 is supported as life dissatisfaction is the channel through which extreme pessimism is negatively associated with cigarette consumption.

Column 2 of Panel C of Table 3.19 reports the NB results, and shows that *EOPT_LDS* is significantly and positively associated with cigarette consumption. The estimated IRR is 1.109. This is in line with H5, and suggests that psychological ill-being (life dissatisfaction) is the channel through which extreme optimism is positively associated with cigarette consumption.

EPES_LDS is significantly and negatively associated with cigarette consumption. This finding supports H6, and suggests that negative well-being, instead of positive well-being, is the channel through which extreme pessimism is negatively associated with cigarette consumption.

In column 3 of Panel C of Table 3.19, in support of H5, I find that dissatisfied optimists are associated with a 2.32 (6.93) percentage point higher cigarette consumption in the short-run (long-run), using the system GMM estimators. This is consistent with the previous findings, which suggest that Russian women, opposite of men, translate their financial and psychological stress into smoking instead of drinking (Cockerham *et al.*, 2006).

3.6.3 Sample attrition

The problem of sample attrition could lead to severe biased estimation if sample attrition is correlated with the dependent variables. Following Nijman and Verbeek (1992) and Jones *et al.* (2006), I conduct the following two variable addition tests to examine whether my empirical results are robust to attrition bias. First, I construct a dummy variable (*next_wave*) equal to 1 if an individual is followed in the next round of survey, and 0 otherwise. I include this variable as an additional control and use a fixed-effects estimator to re-estimate the main models in the full, male and female samples respectively. The FE results for the full/male/female sample are reported in columns 1-5 of Table 3.20/3.21/3.22. I find that the coefficient associated with the *next_wave* term is only significant in the *Alcohol* regressions based on the male sample, while the coefficients on the *next_wave* term are significant in the *Drink200*, *Drink150* and *Smoke* regressions based on the female sample. In all other models, they are insignificant. Moreover, the sign and significance of the extreme optimism variable (as well as those of the control variables) generally do not change after controlling for this new variable. However, there is a concern of biased estimation in the *Smoke* model based on the female sample. In column 4 of Table 3.22, I find that the coefficient of *next_wave* is significant, but small in magnitude (0.0029), while that of extreme optimism turns insignificant.

Second, I examine whether my results are robust to including the number of waves each individual participated in (*num_wave*) as an additional control variable. I use a fixed-effects estimator and re-estimate the main models in the full, male and female samples respectively. The FE results for the full/male/female sample are reported in columns 6-10 of Table 3.20/3.21/3.22. I find that the coefficients associated with *num_wave* are only significant in the *Drink200* and

Smoke regressions based on the full sample, and in the *Smoke* and *Cigarettes* regressions based on the female sample. In addition, I find that adding the *num_wave* term generally does not change the effects of errors in financial expectations (and other control variables) on unhealthy behaviours. Once again, reported in column 9 of Table 3.22, I find that the coefficient of *num_wave* is significant, but small in magnitude (0.0030), while that of extreme optimism turns insignificant. Thus, by controlling for two addition variables, I find that sample attrition does not seem to pose bias estimation in *Drink200*, *Drink150*, *Alcohol* and *Cigarettes* models, while there is a small risk of biased estimation in *Smoke* models.

3.7 Conclusion

In this paper, I use a large dataset containing more than 26,000 individuals and 120,000 observations over the period 2001-2016 to investigate the relationship between errors in financial expectations and unhealthy behaviours in Russia. I primarily focus on two unhealthy behaviours - smoking and drinking. In order to better capture the association between errors in financial expectations and unhealthy behaviours, five measures of smoking and drinking are considered, which include binge and problem drinkings, pure alcohol consumption, smoking and cigarette consumption. As for the errors in financial expectations variables, a novel approach is adopted by dividing the errors variables into four categories: extreme optimism, moderate optimism, extreme pessimism and moderate pessimism (Puri and Robinson, 2007). This paper seeks to provide answers to two main questions: to what extent errors in financial expectations are associated with unhealthy behaviours, and whether psychological ill-being is the channel through which errors in financial expectations are associated with unhealthy behaviours.

The strength of this study is that I make use of a wide range of models regarding different dependent variables, which adds extra validity to my results. For binge drinking and problem drinking, I use the fixed effects LPM models and conditional fixed effects Logit models with the system GMM to account for habit formation. I observe that extreme optimsm is associated with higher probability of binge and problem drinkings in the full and male samples. The associations between extreme optimism and binge and problem drinkings are stronger and more profound in the male sample, which suggests that extreme optimism contributes to Russian men's notorious

reputation for their hard drinking and their high rates of death events related to alcohol abuse (Stuckler and Basu, 2013; Yakovlev, 2018).

My findings are enhanced by results from the alcohol consumption. I apply the fixed effects LR and quantile regressions and dynamic system GMM techniques, and document a positive relationship between extreme optimism and alcohol consumption in the full and male samples. What's more, I observe that as I move along the distribution of pure alcohol consumption, the association between extreme optimism and alcohol consumption becomes larger. This result indicates that extreme optimism is an important determinant of binge alcohol consumption.

Looking at smoking, two techniques are applied: static (fixed effects LPM models and conditional fixed effects Logit models) and dynamic (system GMM) techniques. The results show that extreme optimism is associated with a higher probability of smoking in the full and female samples. The associations found in the female sample could be explained by findings of Cockerham *et al.* (2006) and Quirmbach and Gerry (2016), which suggest that Russian females translate their financial and psychological stress into smoking instead of drinking due to social norms and family obligations.

As for cigarette consumption, I make use of the fixed effects LR and Negative Binomial models and system GMM models, and find consistent results with findings for smoking. I observe that extreme and moderate optimism is associated with more cigarette consumption in the full and female samples. I also document a negative relationship between extreme pessimism and cigarette consumption. These results imply that sometimes aiming low in your expectations could lower your risks of engaging in unhealthy behaviours.

To better understand why the association between errors variables and unhealthy behaviours exists, I use life dissatisfaction as a proxy for psychological ill-being and examine whether psychological ill-being is the channel linking errors variables to unhealthy behaviours. The results show that, for extreme optimists and pessimists, psychological distress (life dissatisfaction) is associated with higher probability of drinking and smoking and higher alcohol and cigarette consumption. This is in line with the DAT and the TRT, which suggest that extreme optimism is associated with psychological ill-being and consequently leads to higher engagements in drinking and smoking as an effort to relieve and regulate negative emotions (Conger, 1956; Ikard *et al.*, 1969; Shepperd and McNulty, 2002).

In summary, my results confirm that, in Russia, extreme optimism is a positive determinant of drinking (for men) and smoking (for women), while extreme pessimism is a negative determinant of smoking for women. Differentiating the associations by gender, I find that the associations with drinking are generally stronger and more profound in the male sample, which suggests that male extreme optimists are among the most prevalent group of alcohol abuses. Interestingly, I observe a positive (negative) relationship between extreme optimism (pessimism) and smoking behaviours in the female sample. These findings indicate that the go-to unhealthy behaviours for Russian men and women are very different and future study in unhealthy behaviours should take into account the gender differences. I also present a novel channel test-psychological ill-being, and find that life dissatisfaction plays a mediating role in the relationship between errors variables and unhealthy behaviours.

As for the policy implications that arises from this study, policy makers in Russia should pay attention to improve individuals' financial literacy and help them to form a good view of

financial expectations. First, policy makers could introduce educational courses and workshops in schools to help students form a prudent financial expectations, and also provide lectures in the communities and regions aimed at improving individuals' financial literacy. Next, based on results from the channel models, interventions and programmes on how to deal with and get a control of individuals' emotions and psychological problems are also needed as optimists are associated with psychological ill-being which in turns leads to higher alcohol and cigarette consumption.

There are several limitations of this work. First, in psychological studies, most of the research on unrealistic optimism use experimental data, instead of self-reported survey data. My work could be biased by inaccurate self-reported data. Second, since smoking and drinking are persistent behaviours, I lost a lot of observations in fixed effects technique due to insufficient variations on dependent variables. Future research on the underlying mechanisms between errors in financial expectations and unhealthy behaviours and the potential casual relationships behind these findings are warranted, and can help to uncover and understand the unknown drivers of smoking and drinking.

Table 3.1 Structure of the panel

Year	Frequency	Percent	Cum(%)
2001	4,943	4.12	4.12
2002	5,838	4.86	8.98
2003	5,731	4.77	13.75
2004	5,999	5	18.74
2005	6,178	5.14	23.89
2006	5,292	4.41	28.29
2007	6,736	5.61	33.9
2008	7,011	5.84	39.74
2009	6,721	5.6	45.34
2010	7,254	6.04	51.38
2011	9,842	8.19	59.57
2012	10,612	8.84	68.41
2013	10,871	9.05	77.46
2014	9,043	7.53	84.99
2015	8,640	7.19	92.18
2016	9,389	7.82	100
Total	120,100	100	

Notes: This table shows the structure of the unbalanced panel employed in this study.

Table 3.2 Summary of dependent variables

Variable	Type	Definition
Drink:		
Binge drinking (*drink200*)	Binary	[1] drank more than 200 grams of pure alcohol per occasion [0] otherwise
Problem drinking (*drink150*)	Binary	[1] drank more than 200 grams of pure alcohol per occasion [0] otherwise
Alcohol consumption	Continuous	the amount of pure alcohol consumed per occasion during the last month
Smoke:		
Smoking	Binary	[1] a current smoker [0] otherwise
Cigarettes	Count	numbers of cigarettes consumed per day during the last month
Notes: This table provides a brief summary of the dependent variables.		

Table 3.3 Distribution of errors in financial expectations

Error	Frequency	Percent	Cumulative	Classification
4	178	0.15	0.15	
3	1,557	1.3	1.44	Extreme optimism
2	8,141	6.78	8.22	
1	26,731	22.26	30.48	Moderate optimism
0	58,455	48.67	79.15	Correct expectation
-1	20,087	16.73	95.88	Moderate pessimism
-2	4,371	3.64	99.52	
-3	550	0.46	99.98	Extreme pessimism
-4	30	0.02	100	

Notes: Columns 1-4 present the distribution of the expectation errors calculated by taking the difference between individuals' financial expectations and realisations. The last column reports the classification of the four independent variables. A full list of the variables used is defined in Appendix 3.1

Table 3.4 Descriptive statistics and differences between male and female samples

Variable	Full sample		Male sample		Female sample		Diff
	Mean	S.D.	Mean	S.D.	Mean	S.D.	
Dependent variables:							
Drink200	0.044	0.205	0.094	0.292	0.011	0.102	0.084***
Drink150	0.09	0.287	0.186	0.389	0.026	0.159	0.161***
Alcohol Consumption	46.7	70.81	77.82	88.4	25.85	45.31	51.97***
Smoke	0.301	0.459	0.55	0.498	0.139	0.346	0.411***
Cigarette consumption	4.43	7.65	8.88	9.26	1.53	4.41	7.35***
Independent variables:							
Extreme Optimism	0.082	0.275	0.085	0.279	0.08	0.272	0.005***
Moderate Optimism	0.223	0.416	0.231	0.421	0.217	0.412	0.014***
Correct prediction	0.487	0.5	0.485	0.5	0.488	0.5	-0.003
Moderate Pessimism	0.167	0.373	0.162	0.369	0.171	0.376	-0.008***
Extreme Pessimism	0.041	0.199	0.037	0.189	0.044	0.205	-0.007***
Life Dissatisfaction	0.331	0.47	0.301	0.459	0.35	0.477	-0.049***
Age	46.95	17.62	44.58	16.54	48.54	18.14	-3.96***
Male	0.401	0.49	1	0	0	0	1
Income	17413	14984	20801	16750	15143	13194	5658***
Consumption	56419	43933	58342	43728	55129	44023	3212***
Married	0.555	0.497	0.656	0.475	0.486	0.5	0.170***
Cohabiting	0.105	0.307	0.119	0.324	0.096	0.294	0.023***
Divorced	0.082	0.275	0.044	0.205	0.108	0.311	-0.064***

Primary Education	0.043	0.203	0.033	0.179	0.049	0.217	-0.016***
Unfinished Education	0.145	0.352	0.166	0.372	0.131	0.337	0.036***
Secondary Education	0.321	0.467	0.387	0.487	0.277	0.447	0.110***
Vocational Education	0.261	0.439	0.209	0.407	0.296	0.457	-0.087***
Self_Health	0.151	0.358	0.111	0.314	0.178	0.382	-0.067***
Retired	0.268	0.443	0.192	0.394	0.318	0.466	-0.127***
Price_vodka	6.603	0.513	6.593	0.513	6.609	0.513	-0.016***
Price_fortified wine	6.102	0.610	6.093	0.614	6.108	0.607	-0.015***
Price_table wine	6.029	0.580	6.020	0.579	6.035	0.580	-0.015***
Price_beer	4.954	0.403	4.946	0.400	4.959	0.405	-0.012***
Price_cognac	7.788	0.782	7.773	0.782	7.798	0.782	-0.025***
Price_domestic cigarettes	4.048	0.468	4.044	0.468	4.050	0.468	-0.006**
Price_imported cigarettes	4.648	0.396	4.645	0.397	4.649	0.395	-0.004*
N =	120,100		48,197		71,903		

Notes: This table presents the summary statistics of all variables used. The Diff column reports the p-value of t-tests of the difference in each variable between male and female sub-groups. A full list of the variables used is defined in Appendix 3.1

Table 3.5 Descriptive statistics by errors in financial expectations in the full sample and sub-samples

Variable	Extreme Optimism	Moderate Optimism	Correct Prediction	Moderate Pessimism	Extreme Pessimism
			Panel A: Full sample		
Drink200	0.061	0.047	0.041	0.041	0.043
Drink150	0.117	0.095	0.086	0.084	0.087
Alcohol Consumption	54.45	49.22	44.73	45.82	44.66
Smoke	0.382	0.327	0.286	0.281	0.270
cigarette	5.51	4.74	4.24	4.18	4.09
Life Dissatisfaction	0.474	0.335	0.315	0.288	0.372
N=	9,876	26,731	58,455	20,087	4,951
			Panel B: Male sample		
Drink200	0.126	0.098	0.087	0.092	0.101
Drink150	0.228	0.193	0.178	0.180	0.190
Alcohol Consumption	87.21	80.07	75.24	77.03	79.51
Smoke	0.620	0.564	0.539	0.527	0.551
cigarette	10.07	9.02	8.70	8.60	9.04
Life Dissatisfaction	0.454	0.310	0.283	0.256	0.333
N=	4,109	11,120	23,359	7,820	1,789
			Panel C: Female sample		
Drink200	0.015	0.011	0.010	0.009	0.010
Drink150	0.037	0.026	0.025	0.023	0.029
Alcohol Consumption	31.10	27.25	24.42	25.92	24.95
Smoke	0.217	0.162	0.122	0.129	0.115
cigarette	2.38	1.76	1.35	1.43	1.37
Life Dissatisfaction	0.487	0.353	0.336	0.309	0.394
N =	5,767	15,611	35,096	12,267	3,162

Note: This table presents the summary statistics of relevant variables by errors in financial expectations. Statistical significance is indicated by * at the 10%, ** at the 5%, *** at the 1%. A full list of the variables used is defined in Appendix 3.1

Table 3.6 Correlation Matrix between errors in financial expectations indicators

	Extreme Optimism	Moderate Optimism	Moderate Pessimism	Extreme Pessimism
Extreme Optimism		-0.1602*	-0.1341*	-0.0621*
Moderate Optimism	-0.1602*		-0.2398*	-0.1109*
Moderate Pessimism	-0.1341*	-0.2398*		-0.0929*
Extreme Pessimism	-0.0621*	-0.1109*	-0.0929*	

Notes: Lower triangular cells present Spearman's rank correlation coefficients, upper triangular cells show the Pearson's correlation coefficients. * indicates statistical significance is at the 1% level.

Table 3.7 LPM and Logit models for binge drinking

Dependent variable	Binge Drinking (Drink200)					
	LPM			Logit		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables:	Full	Male	Female	Full	Male	Female
Extreme Optimism	0.0055**	0.0182***	-0.0023	1.162**	1.253***	0.866
	(0.0027)	(0.0061)	(0.0018)	(0.077)	(0.093)	(0.137)
Moderate Optimism	0.0019	0.0045	0.0001	1.059	1.064	1.044
	(0.0016)	(0.0035)	(0.0011)	(0.051)	(0.056)	(0.124)
Extreme Pessimism	0.0016	0.0039	0.0006	1.059	1.048	1.147
	(0.0030)	(0.0079)	(0.0019)	(0.106)	(0.115)	(0.275)
Moderate Pessimism	0.0001	0.0001	0.0003	1.020	1.006	1.130
	(0.0016)	(0.0038)	(0.0011)	(0.056)	(0.060)	(0.154)
Anti-Drinking Policy	-0.0595*	-0.1200**	-0.0091	0.130*	0.0755*	0.545
	(0.0343)	(0.0559)	(0.0418)	(0.159)	(0.103)	(1.264)
Age	0.0018	0.0035	-0.0002	1.075	1.119	0.961
	(0.0023)	(0.0036)	(0.0028)	(0.087)	(0.100)	(0.147)
Income (log)	0.0008**	0.0021***	-0.0001	1.020***	1.025***	0.997
	(0.0003)	(0.0007)	(0.0002)	(0.007)	(0.008)	(0.016)
Consumption (log)	0.0008	0.0019	0.0001	1.025	1.027	1.017
	(0.0009)	(0.0022)	(0.0006)	(0.032)	(0.035)	(0.079)
Married	-0.0035	0.0046	-0.0030	1.039	1.150	0.775
	(0.0030)	(0.0085)	(0.0022)	(0.122)	(0.159)	(0.184)
Cohabiting	0.0064	0.0260**	-0.0021	1.302**	1.470***	0.914
	(0.0044)	(0.0105)	(0.0037)	(0.154)	(0.208)	(0.206)

Divorced	0.0054	0.0146	0.0041	1.348**	1.366*	1.298
	(0.0043)	(0.0151)	(0.0033)	(0.197)	(0.246)	(0.333)
Primary Educ	-0.0107	-0.0106	-0.0080	0.707	0.875	0.253
	(0.0107)	(0.0211)	(0.0100)	(0.265)	(0.363)	(0.235)
Unfinished Secondary Educ	-0.0114*	-0.0179	-0.0032	0.749	0.794	0.672
	(0.0059)	(0.0115)	(0.0052)	(0.133)	(0.160)	(0.274)
Secondary Educ	-0.0057	-0.0057	-0.0054	0.823	0.905	0.565
	(0.0045)	(0.0094)	(0.0040)	(0.132)	(0.166)	(0.200)
Vocational Secondary Educ	-0.0090**	-0.0170*	-0.0028	0.770*	0.792	0.748
	(0.0041)	(0.0090)	(0.0037)	(0.113)	(0.135)	(0.218)
Self-reported Health	-0.0044**	-0.0128**	0.0001	0.817**	0.779**	1.029
	(0.0020)	(0.0055)	(0.0013)	(0.073)	(0.078)	(0.210)
Retired	-0.0023	-0.0005	-0.0034**	0.848	0.943	0.446***
	(0.0028)	(0.0070)	(0.0017)	(0.093)	(0.111)	(0.139)
Price_vodka	0.0013	0.0014	0.0011	1.051	1.035	1.107
	(0.0017)	(0.0039)	(0.0011)	(0.054)	(0.059)	(0.140)
Price_fortified wine	-0.0014	-0.0027	-0.0005	0.953	0.959	0.941
	(0.0013)	(0.0031)	(0.0009)	(0.038)	(0.042)	(0.089)
Price_table wine	0.0014	0.0019	-0.0013	1.008	1.035	0.91
	(0.0014)	(0.0033)	(0.0009)	(0.042)	(0.047)	(0.091)
Price_beer	0.0016	0.0017	0.0016	1.073	1.043	1.215
	(0.0019)	(0.0046)	(0.0012)	(0.069)	(0.073)	(0.189)
Price_cognac	-0.0025**	-0.0055**	-0.0005	0.917**	0.913**	0.914
	(0.0011)	(0.0025)	(0.0007)	(0.033)	(0.037)	(0.081)
Observations	120,100	48197	71,903	18,631	14933	3,698
R-squared	0.002	0.004	0.001			

Number of idind	26,495	11,408	15,087	2,768	2,272	496

Notes: This table reports the fixed-effects LPM and conditional fixed-effects Logit estimates for binge drinking. Coefficients are presented in LPM, while odds ratios are presented in Logit. Robust standard errors are given in parentheses. Regional and Time dummies are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Table 3.8 LPM and Logit models for problem drinking

Dependent variable	Problem Drinking (Drink150)					
	LPM			Logit		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables:	Full	Male	Female	Full	Male	Female
Extreme Optimism	0.0065*	0.0207***	-0.0025	1.117**	1.186***	0.924
	(0.0034)	(0.0074)	(0.0028)	(0.058)	(0.071)	(0.097)
Moderate Optimism	0.0007	0.0030	-0.0008	1.016	1.029	0.976
	(0.0020)	(0.0045)	(0.0016)	(0.037)	(0.043)	(0.077)
Extreme Pessimism	0.0017	-0.0014	0.0035	1.043	0.992	1.213
	(0.0041)	(0.0100)	(0.0032)	(0.079)	(0.087)	(0.184)
Moderate Pessimism	-0.0003	0.0008	-0.0005	1.000	1.005	0.987
	(0.0022)	(0.0049)	(0.0016)	(0.041)	(0.047)	(0.088)
Anti-Drinking Policy	-0.1070**	-0.1650*	-0.0619	0.141**	0.167*	0.043
	(0.0494)	(0.0855)	(0.0503)	(0.121)	(0.157)	(0.096)
Age	0.0049	0.0064	0.0032	1.096	1.084	1.182
	(0.0033)	(0.0056)	(0.0033)	(0.062)	(0.067)	(0.176)
Income (log)	0.0006	0.0010	0.0003	1.009*	1.008	1.012
	(0.0004)	(0.0008)	(0.0003)	(0.005)	(0.006)	(0.012)
Consumption (log)	0.0016	0.0014	0.00176**	1.032	1.012	1.120**
	(0.0012)	(0.0029)	(0.0009)	(0.024)	(0.026)	(0.059)
Married	0.0008	0.0275**	-0.0047	1.148	1.353***	0.826
	(0.0041)	(0.0116)	(0.0032)	(0.100)	(0.143)	(0.131)
Cohabiting	0.0103*	0.0375***	0.0019	1.297***	1.457***	1.048
	(0.0058)	(0.0136)	(0.0054)	(0.115)	(0.162)	(0.156)

Divorced	0.0133**	0.0535***	0.0030	1.410***	1.667***	1.100
	(0.0060)	(0.0203)	(0.0048)	(0.155)	(0.237)	(0.193)
Primary Educ	-0.0204	-0.0278	-0.0098	0.661	0.751	0.439
	(0.0144)	(0.0311)	(0.0100)	(0.187)	(0.231)	(0.337)
Unfinished Secondary Educ	-0.0111	-0.0082	-0.0075	0.863	0.960	0.599*
	(0.0080)	(0.0159)	(0.0070	(0.111)	(0.142)	(0.168)
Secondary Educ	-0.0052	0.0057	-0.0116**	0.912	1.064	0.519***
	(0.0063)	(0.0136)	(0.0053	(0.105)	(0.143)	(0.125)
Vocational Secondary Educ	-0.0051	-0.0063	-0.0029	0.928	0.965	0.866
	(0.0056)	(0.0125)	(0.0051	(0.099)	(0.123)	(0.170)
Self-reported Health	-0.0114***	-0.0300***	-0.0016	0.748***	0.712***	0.899
	(0.0028)	(0.0075)	(0.0020	(0.050)	(0.054)	(0.123)
Retired	-0.0152***	-0.0273***	-0.0079***	0.694***	0.747***	0.520***
	(0.0041)	(0.0099)	(0.0028	(0.055)	(0.066)	(0.099)
Price_vodka	-0.0010	-0.0036	0.0003	0.977	0.967	1.014
	(0.0023)	(0.0051)	(0.0017)	(0.038)	(0.043)	(0.085)
Price_fortified wine	-0.0046***	-0.0083**	-0.0020	0.919***	0.927**	0.902
	(0.0018)	(0.0039)	(0.0014)	(0.028)	(0.032)	(0.058)
Price_table wine	0.0018	0.0039	0.0001	1.032	1.033	1.017
	(0.0018)	(0.0041)	(0.0014)	(0.033)	(0.038)	(0.067)
Price_beer	0.0008	0.0007	0.0008	1.021	1.015	1.034
	(0.0026)	(0.0060)	(0.0020)	(0.049)	(0.056)	(0.103)
Price_cognac	-0.0036**	-0.0100***	0.0005	0.935**	0.909***	1.025
	(0.0015)	(0.0033)	(0.0011)	(0.025)	(0.028)	(0.058)
Observations	120,100	48,197	71,903	29,755	22,129	7,626
R-squared	0.002	0.006	0.001			

Number of idind	26,495	11,408	15,087	4,506	3,440	1,066

Notes: This table reports the fixed-effects LPM and conditional fixed-effects Logit estimates for problem drinking. Coefficients are presented in LPM, while odds ratios are presented in Logit. Robust standard errors are given in parentheses. Regional and Time dummies are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Table 3.9 GMM models for drinking

Dependent variable	Binge Drinking (Drink200)			Problem Drinking (Drink150)		
	GMM			GMM		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables:	Full	Male	Female	Full	Male	Female
Extreme Optimism	0.0087***	0.0206***	-0.0001	0.0096**	0.0242***	0.0015
	[0.0089***]	[0.0221***]	[-0.0001]	[0.0123***]	[0.0271***]	[0.0012]
	(0.0027)	(0.0062)	(0.0015)	(0.0038)	(0.0077)	(0.0031)
	[0.0027]	[0.0069]	[0.0014]	[0.0047]	[0.0090]	[0.0013]
Moderate Optimism	0.0031**	0.0067*	0.0004	-0.0001	0.0002	-0.0005
	[0.0032**]	[0.0071*]	[0.0004]	[-0.001]	[0.0002]	[-0.0004]
	(0.0015)	(0.0036)	(0.0009)	(0.0022)	(0.0048)	(0.0016)
	[0.0016]	[0.0039]	[0.0009]	[0.0028]	[0.0053]	[0.0013]
Extreme Pessimism	0.0018	0.0026	0.0007	0.0017	-0.0036	0.0039
	[0.0018]	[0.0028]	[0.0006]	[0.0021]	[-0.0040]	[0.0034]
	(0.0031)	(0.0082)	(0.0018)	(0.0044)	(0.0105)	(0.0029)
	[0.0031]	[0.0086]	[0.0017]	[0.0055]	[0.0119]	[0.0027]
Moderate Pessimism	-0.0001	-0.0001	0.0001	-0.0021	-0.0014	-0.0010
	[-0.001]	[-0.0001]	[0.0001]	[-0.0027]	[-0.0015]	[-0.0008]
	(0.0016)	(0.0040)	(0.0009)	(0.0024)	(0.0052)	(0.0016)
	[0.0016]	[0.0042]	[0.0009]	[0.0032]	[0.0059]	[0.0014]
Lagged Drink200	0.0315	0.0696	-0.0139			
	(0.1440)	(0.1530)	(0.0721)			
Lagged Drink150				0.2220	0.1080	-0.1520
				(0.1490)	(0.1500)	(0.2320)
Anti-Drinking Policy	-0.0246***	-0.0510***	-0.0085***	-0.0275***	-0.0575***	-0.0137***
	(0.0046)	(0.0100)	(0.0031)	(0.0063)	(0.0140)	(0.0045)
Male	0.0713***	0.1500***	0	0.1130***	0.2820***	0
	(0.0111)	(0.0403)	(0)	(0.0224)	(0.0641)	(0)
Age	-0.0003***	-0.0007***	-0.0003***	-0.0002**	-0.0005**	-0.0005***
	(0.0001)	(0.0002)	(0.0000)	(0.0001)	(0.0002)	(0.0001)

Income (log)	0.0006**	0.0012**	0.0001	0.0002	-0.0003	0.0003
	(0.0003)	(0.0006)	(0.0001)	(0.0004)	(0.0008)	(0.0003)
Consumption (log)	0.0007	0.0018	-0.0001	0.0014	0.0001	0.00125*
	(0.0008)	(0.0020)	(0.0005)	(0.0011)	(0.0027)	(0.0007)
Married	0.0033*	0.0224***	-0.0015	0.0093***	0.0486***	0.0004
	(0.0018)	(0.0073)	(0.0010)	(0.0031)	(0.0111)	(0.0016)
Cohabiting	0.0208***	0.0500***	0.0057**	0.0294***	0.0742***	0.0242***
	(0.0046)	(0.0118)	(0.0023)	(0.0076)	(0.0167)	(0.0055)
Divorced	0.0191***	0.0698***	0.00409**	0.0283***	0.1090***	0.0150***
	(0.0042)	(0.0158)	(0.0017)	(0.0065)	(0.0214)	(0.0044)
Primary Educ	0.0098***	0.0130	0.0043***	0.0136***	0.0226*	0.0093***
	(0.0033)	(0.0082)	(0.0016)	(0.0046)	(0.0123)	(0.0029)
Unfinished Secondary Educ	0.0162***	0.0242***	0.0053***	0.0238***	0.0430***	0.0151***
	(0.0034)	(0.0065)	(0.0016)	(0.0056)	(0.0105)	(0.0038)
Secondary Educ	0.0134***	0.0230***	0.0026**	0.0211***	0.0433***	0.0076***
	(0.0028)	(0.0056)	(0.0012)	(0.0049)	(0.0098)	(0.0026)
Vocational Secondary Educ	0.0037*	0.0055	0.0005	0.0071**	0.0159**	0.0035
	(0.0020)	(0.0049)	(0.0010)	(0.0030)	(0.0076)	(0.0021)
Self-reported Health	-0.0064***	-0.0185***	-0.0015*	-0.0108***	-0.0363***	-0.0018
	(0.0018)	(0.0049)	(0.0009)	(0.0031)	(0.0080)	(0.0018)
Retired	-0.0063**	-0.0113*	-0.0006	-0.0169***	-0.0336***	-0.0112***
	(0.0029)	(0.0067)	(0.0010)	(0.0052)	(0.0106)	(0.0034)
Price_vodka	0.0026*	0.0026	0.0024**	-0.0006	-0.0024	0.00187
	(0.0016)	(0.0038)	(0.0010)	(0.0023)	(0.0051)	(0.0018)
Price_fortified wine	-0.0013	-0.0015	-0.0005	-0.0051***	-0.0085**	-0.0018
	(0.0012)	(0.0029)	(0.0008)	(0.0018)	(0.0039)	(0.0013)

Price_table wine	-0.0012	-0.0002	-0.0014	0.0011	0.0023	-0.0014
	(0.0014)	(0.0032)	(0.0009)	(0.0020)	(0.0042)	(0.0014)
Price_beer	-0.0003	-0.0035	0.0008	0.0026	0.0031	0.0018
	(0.0019)	(0.0045)	(0.0012)	(0.0027)	(0.0060)	(0.0019)
Price_cognac	-0.0014	-0.0047**	0.0002	-0.0018	-0.0060*	0.0018*
	(0.0010)	(0.0024)	(0.0006)	(0.0015)	(0.0032)	(0.0011)
Hansen-J test (p value)	0.766	0.878	0.510	0.797	0.246	0.198
M3 (p value)	0.244	0.384	0.345	0.616	0.398	0.168
Observations	119,305	47,463	71,842	119,305	47,463	71,842
Number of idind	26,409	11,325	15,084	26,409	11,325	15,084

Notes: This table reports the system GMM estimates for binge and problem drinkings. M3 is a test for third-order serial correlation of the residuals in the differenced equations, under the null hypothesis that there is no third-order serial correlation. The Hansen J-test of over-identifying is aimed at testing the exogeneity of instruments. The null hypothesis of the Hansen J-test is that the instruments are not correlated with the error term. I use the third to seventh lag of dependent variable as instruments for drink200, and I use the fifth to seventh lag of dependent variable as instruments for drink150. Robust standard errors are given in parentheses. In addition, long-run coefficients and standard errors are included below the short-run coefficients and standard errors in square parentheses. Regional and Time dummies are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Table 3.10 Quantile regressions for alcohol consumption in the male sample

Dependent variable	Alcohol consumption				
	Quantile regressions				
	Male sample				
Variables:	10th	30th	50th	70th	90th
Extreme Optimism	2.475	3.708	5.242**	7.443**	9.170**
	(4.205)	(3.221)	(2.51)	(3.19)	(4.609)
Moderate Optimism	0.664	0.921	1.241	1.699	2.059
	(2.567)	(1.966)	(1.532)	(1.947)	(2.814)
Extreme Pessimism	-0.988	0.428	2.191	4.719	6.703
	(5.672)	(4.344)	(3.385)	(4.303)	(6.216)
Moderate Pessimism	1.562	1.315	1.008	0.568	0.223
	(2.817)	(2.158)	(1.682)	(2.137)	(3.088)
Anti-Drinking Policy	-22.550	-26.490	-31.400	-38.440	-43.960
	(74.91)	(57.38)	(44.71)	(56.83)	(82.11)
Age	3.901	4.283	4.759	5.441	5.977
	(4.958)	(3.798)	(2.959)	(3.761)	(5.435)
Income (log)	0.379	0.467	0.577**	0.734**	0.858*
	(0.452)	(0.346)	(0.27)	(0.343)	(0.495)
Consumption (log)	-0.532	-0.120	0.393	1.129	1.706
	(1.738)	(1.331)	(1.037)	(1.318)	(1.905)
Married	9.501	8.747*	7.809*	6.463	5.406
	(6.775)	(5.19)	(4.044)	(5.14)	(7.426)
Cohabiting	11.590	11.710**	11.850***	12.050**	12.210
	(7.248)	(5.552)	(4.326)	(5.499)	(7.945)
Divorced	16.050	16.570**	17.210***	18.130**	18.850

	(10.8)	(8.274)	(6.447)	(8.194)	(11.84)
Primary Educ	-7.478	-7.715	-8.011	-8.434	-8.767
	(20.51)	(15.71)	(12.24)	(15.56)	(22.48)
Unfinished Secondary Educ	-11.150	-10.480	-9.641*	-8.439	-7.496
	(9.258)	(7.092)	(5.526)	(7.023)	(10.15)
Secondary Educ	-8.859	-8.595	-8.265*	-7.793	-7.423
	(8.174)	(6.261)	(4.879)	(6.201)	(8.959)
Vocational Secondary Educ	-8.560	-8.274	-7.918*	-7.408	-7.008
	(7.177)	(5.498)	(4.284)	(5.445)	(7.867)
Self-reported Health	-6.506	-6.727*	-7.001**	-7.395**	-7.704
	(4.912)	(3.762)	(2.932)	(3.726)	(5.383)
Retired	-5.665	-5.048	-4.280	-3.178	-2.314
	(5.467)	(4.188)	(3.263)	(4.148)	(5.992)
Price_vodka	1.090	1.093	1.096	1.101	1.105
	(2.7790)	(2.1290)	(1.6590)	(2.1080)	(3.0460)
Price_fortified wine	-4.500**	-4.162**	-3.742***	-3.138*	-2.665
	(2.2070)	(1.6910)	(1.3180)	(1.6750)	(2.4190)
Price_table wine	2.191	1.653	0.984	0.025	-0.728
	(2.2460)	(1.7200)	(1.3410)	(1.7040)	(2.4620)
Price_beer	-1.797	-0.991	0.011	1.451	2.58
	(3.2570)	(2.4950)	(1.9440)	(2.4710)	(3.5700)
Price_cognac	-0.267	-0.239	-0.204	-0.155	-0.116
	(1.8760)	(1.4370)	(1.1200)	(1.4230)	(2.0560)
Observations	31,285	31,285	31,285	31,285	31,285
Number of idind	9,166	9,166	9,166	9,166	9,166

Notes: This table reports the fixed effects Quantile regression results for alcohol consumption in the male sample. Standard errors are given in parentheses. Regional and Time dummies are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Table 3.11 LR and GMM models for alcohol consumption

Dependent variable	Alcohol Consumption					
	LR			GMM		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables:	Full	Male	Female	Full	Male	Female
Extreme Optimism	0.613	3.736**	-1.310**	1.003 [3.295]	4.552** [9.522**]	-0.361 [-0.888]
	(0.733)	(1.558)	(0.645)	(1.013) [3.089]	(1.881) [4.235]	(0.879) [2.319]
Moderate Optimism	0.108	0.632	-0.290	0.009 [0.029]	0.344 [0.720]	0.139 [0.341]
	(0.428)	(0.905)	(0.381)	(0.578) [1.897]	(1.115) [2.293]	(0.477) [1.165]
Extreme Pessimism	0.581	0.614	0.623	0.019 [0.063]	0.045 [0.094]	0.197 [0.485]
	(0.841)	(1.984)	(0.714)	(1.068) [3.505]	(2.277) [4.758]	(0.888) [2.191]
Moderate Pessimism	0.514	0.371	0.743*	0.468 [1.538]	0.480 [1.003]	0.653 [1.606]
	(0.458)	(1.008)	(0.392)	(0.611) [2.081]	(1.249) [2.612]	(0.489) [1.369]
Lagged Alcohol Consumption				0.696***	0.522***	0.594***
				(0.118)	(0.136)	(0.149)
Anti-Drinking Policy	-53.590***	-61.510***	-49.620***	-5.891***	-10.95***	-5.256***
	(12.830)	(22.330)	(10.850)	(1.647)	(3.285)	(1.262)
Male				13.26**	48.04***	0
				(5.639)	(14.200)	(0)
Age	2.251***	2.086	2.374***	-0.0843***	-0.114***	-0.126***
	(0.848)	(1.470)	(0.720)	(0.019)	(0.041)	(0.035)
Income (log)	0.509***	0.684***	0.363***	0.421***	0.493***	0.398***
	(0.0855)	(0.174)	(0.0769)	(0.090)	(0.180)	(0.092)

Consumption (log)	0.849***	0.560	1.039***	0.796***	0.670	0.985***
	(0.275)	(0.636)	(0.216)	(0.286)	(0.594)	(0.282)
Married	0.586	10.58***	-2.201***	0.919	7.122**	0.227
	(0.976)	(2.563)	(0.848)	(1.065)	(3.285)	(0.519)
Cohabiting	3.102**	12.64***	-0.145	3.997*	11.57**	4.215**
	(1.272)	(2.864)	(1.245)	(2.247)	(4.573)	(1.892)
Divorced	2.996**	9.819**	1.478	3.273*	13.32**	2.590**
	(1.333)	(4.209)	(1.166)	(1.870)	(5.556)	(1.310)
Primary Educ	-7.848**	-9.857	-5.309*	0.217	-0.715	-0.480
	(3.307)	(6.733)	(2.814)	(0.878)	(2.363)	(0.775)
Unfinished Secondary Educ	-6.899***	-7.197**	-5.356***	2.273***	4.594***	1.140*
	(1.808)	(3.525)	(1.698)	(0.854)	(1.732)	(0.680)
Secondary Educ	-3.924***	-2.476	-4.654***	1.656**	4.203***	0.510
	(1.447)	(3.048)	(1.319)	(0.706)	(1.530)	(0.495)
Vocational Secondary Educ	-3.485***	-5.180*	-1.991	0.346	0.711	0.327
	(1.284)	(2.743)	(1.249)	(0.504)	(1.299)	(0.446)
Self-reported Health	-6.195***	-13.240***	-2.561***	-3.720***	-9.947***	-2.237***
	(0.672)	(1.674)	(0.535)	(0.989)	(2.260)	(0.759)
Retired	-4.970***	-7.532***	-3.318***	-2.821*	-5.920**	-2.797**
	(0.926)	(2.108)	(0.743)	(1.538)	(2.516)	(1.365)
Price_vodka	0.121	-0.221	0.244	0.0711	0.431	0.107
	(0.4670)	(1.0280)	(0.3860)	(0.5930)	(1.1770)	(0.5120)
Price_fortified wine	-1.002***	-1.838**	-0.435	-0.628	-1.544*	0.0341
	(0.3760)	(0.8030)	(0.3300)	(0.4420)	(0.8650)	(0.3790)
Price_table wine	0.526	0.903	0.19	-0.437	0.315	-0.983**
	(0.3760)	(0.8200)	(0.3250)	(0.4660)	(0.9190)	(0.4000)

Price_beer	0.784	1.804	0.124	-0.117	0.185	-0.454
	(0.5550)	(1.2310)	(0.4610)	(0.7020)	(1.4200)	(0.5900)
Price_cognac	-0.723**	-2.207***	0.264	-0.0488	-0.927	0.686**
	(0.3040)	(0.6550)	(0.2630)	(0.3670)	(0.7210)	(0.3200)
Hansen-J test (p value)				0.270	0.291	0.209
M3 (p value)				0.563	0.543	0.797
Observations	120,100	48197	71,903	119,305	47,463	71,842
R-squared	0.011	0.016	0.011			
Number of idind	26,495	11,408	15,087	26,409	11,325	15,084

Notes: This table reports the fixed effects LR and the system GMM estimates for alcohol consumption. M3 is a test for third-order serial correlation of the residuals in the differenced equations, under the null hypothesis that there is no third-order serial correlation. The Hansen J-test of over-identifying is aimed at testing the exogeneity of instruments. The null hypothesis of the Hansen J-test is that the instruments are not correlated with the error term. I use the third to sixth lag of dependent variable as instruments. Robust standard errors are given in parentheses. In addition, long-run coefficients and standard errors are included below the short-run coefficients and standard errors in square parentheses. Regional and Time dummies are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Table 3.12 LPM and Logit models for smoking

Dependent variable	Smoke					
	LPM			Logit		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables:	Full	Male	Female	Full	Male	Female
Extreme Optimism	0.0041	0.0000	0.0064*	1.165**	1.057	1.263**
	(0.0031)	(0.0060)	(0.0034)	(0.083)	(0.104)	(0.134)
Moderate Optimism	0.0016	0.0013	0.0012	1.041	1.019	1.057
	(0.0019)	(0.0038)	(0.0020)	(0.051)	(0.065)	(0.082)
Extreme Pessimism	-0.0036	-0.0010	-0.0050	0.904	0.967	0.792
	(0.0035)	(0.0071)	(0.0036)	(0.095)	(0.136)	(0.128)
Moderate Pessimism	0.0014	0.0041	0.0000	1.065	1.095	1.021
	(0.0020)	(0.0042)	(0.0020)	(0.059)	(0.078)	(0.092)
Anti-Smoking Policy	-0.216***	-0.433***	-0.0012	0.038***	0.014***	0.879
	(0.0789)	(0.0927)	(0.0569)	(0.037)	(0.018)	(1.761)
Age	0.00986*	0.0186***	-0.0010	1.110*	1.124	0.972
	(0.0052)	(0.0060)	(0.0038)	(0.070)	(0.091)	(0.129)
Income (log)	0.0013***	0.0008	0.0015***	1.035***	1.032***	1.039***
	(0.0004)	(0.0007)	(0.0004)	(0.007)	(0.010)	(0.011)
Consumption (log)	0.0001	-0.0047*	0.0031**	1.013	0.929*	1.187***
	(0.0013)	(0.0027)	(0.0012)	(0.030)	(0.035)	(0.065)
Married	-0.0190***	0.0071	-0.0174***	0.999	1.595***	0.685**
	(0.0052)	(0.0124)	(0.0053)	(0.107)	(0.242)	(0.106)
Cohabiting	0.0010	0.0217	0.0016	1.292**	1.716***	1.049
	(0.0068)	(0.0132)	(0.0080)	(0.149)	(0.296)	(0.165)
Divorced	0.0212***	0.0353*	0.0239***	2.093***	2.050***	1.972***

	(0.0073)	(0.0188)	(0.0075)	(0.281)	(0.427)	(0.353)
Primary Educ	0.0037	-0.0048	0.0085	0.903	0.769	8.648*
	(0.0136)	(0.0283)	(0.0100)	(0.307)	(0.287)	(10.950)
Unfinished Secondary Educ	-0.0040	-0.0078	0.0012	0.890	0.921	1.039
	(0.0090)	(0.0175)	(0.0090)	(0.137)	(0.175)	(0.290)
Secondary Educ	-0.0031	-0.0104	0.0045	0.920	0.911	1.187
	(0.0078)	(0.0162)	(0.0076)	(0.120)	(0.147)	(0.281)
Vocational Secondary Educ	-0.0121*	-0.0351**	0.0063	0.754**	0.611***	1.220
	(0.0070)	(0.0148)	(0.0068)	(0.093)	(0.099)	(0.252)
Self-reported Health	-0.0147***	-0.0314***	-0.0055**	0.562***	0.522***	0.679**
	(0.0031)	(0.0075)	(0.0026)	(0.051)	(0.060)	(0.106)
Retired	-0.0099**	-0.0221**	-0.0040	0.545***	0.604***	0.512***
	(0.0045)	(0.0098)	(0.0037)	(0.066)	(0.085)	(0.123)
Price_domestic cigarettes	-0.0032	-0.0053	-0.0019	0.898*	0.878	0.94
	(0.0023)	(0.0045)	(0.0024)	(0.055)	(0.071)	(0.089)
Price_imported cigarettes	0.0023	0.0045	0.0005	1.048	1.057	0.989
	(0.0027)	(0.0055)	(0.0027)	(0.068)	(0.089)	(0.106)
Observations	101,645	40,622	61,023	16,275	10,085	6,190
R-squared	0.011	0.030	0.004			
Number of idind	23,985	10,262	13,723	2,565	1,570	995

Notes: This table reports the fixed-effects LPM and conditional fixed-effects Logit estimates for smoking. Coefficients are presented in LPM, while odds ratios are presented in Logit. Robust standard errors are given in parentheses. Regional and Time dummies are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Table 3.13 GMM models by smoking

Dependent variable	Smoke		
	GMM		
	(1)	(2)	(3)
Variables:	Full	Male	Female
Extreme Optimism	0.0121*** [0.0413***] (0.0041) [0.0136]	0.0090 [0.0263] (0.0072) [0.0202]	0.0136*** [0.0472] (0.0047) [0.0176]
Moderate Optimism	0.0011 [0.0037] (0.0025) [0.0083]	-0.0004 [-0.0011] (0.0044) [0.0129]	0.0012 [0.0040] (0.0029) [0.0093]
Extreme Pessimism	-0.0068 [-0.0232] (0.0042) [0.0164]	-0.0071 [-0.0204] (0.0089) [0.0274]	-0.0049 [-0.0171] (0.0041) [0.0167]
Moderate Pessimism	-0.0017 [-0.0056] (0.0025) [0.0088]	-0.0066 [-0.0190] (0.0051) [0.0158]	0.0009 [0.0031] (0.0024) [0.0084]
lagged Smoke	0.707*** (0.0758)	0.655*** (0.0888)	0.713*** (0.1180)
Anti-Smoking Policy	-0.0155*** (0.0049)	-0.0400*** (0.0100)	-0.0012 (0.0059)
Age	-0.0011*** (0.0003)	-0.0015*** (0.0004)	-0.0012*** (0.0004)
Income (log)	0.0000 (0.0004)	-0.0011 (0.0007)	0.0006 (0.0005)
Consumption (log)	0.0015 (0.0012)	-0.0032 (0.0024)	0.00367** (0.0015)
Married	-0.0022	0.0234**	-0.0086***

	(0.0023)	(0.0108)	(0.0031)
Cohabiting	0.0290***	0.0571***	0.0228*
	(0.0107)	(0.0192)	(0.0137)
Divorced	0.0297***	0.0634***	0.0201**
	(0.0086)	(0.0204)	(0.0093)
Primary Educ	0.0341***	0.0439***	0.0306***
	(0.0088)	(0.0156)	(0.0109)
Unfinished Secondary Educ	0.0559***	0.0905***	0.0385**
	(0.0136)	(0.0231)	(0.0152)
Secondary Educ	0.0457***	0.0784***	0.0311***
	(0.0107)	(0.0194)	(0.0107)
Vocational Secondary Educ	0.0228***	0.0410***	0.0172***
	(0.0062)	(0.0125)	(0.0066)
Self-reported Health	-0.0122***	-0.0282***	-0.0028
	(0.0026)	(0.0066)	(0.0022)
Retired	-0.0144***	-0.0219***	-0.0079
	(0.0049)	(0.0082)	(0.0055)
Price_domestic cigarettes	0.0000	-0.0023	0.0018
	(0.0027)	(0.0050)	(0.0031)
Price_imported cigarettes	-0.0006	-0.0145***	0.0071**
	(0.0028)	(0.0056)	(0.0028)
Hansen-J test (p value)	0.314	0.190	0.140
M3 (p value)	0.750	0.848	0.478
Observations	97,370	38,109	59,261
Number of individual	23,688	10,045	13,643

Notes: This table reports the system GMM estimates for smoking. M3 is a test for third-order serial correlation of the residuals in the differenced equations, under the null hypothesis that there is no third-order serial correlation. The Hansen J-test of over-identifying is aimed at testing the exogeneity of instruments. The null hypothesis of the Hansen J-test is that the instruments are not correlated with the error term. I use the third to fifth lag of dependent variable as instruments. Standard errors are given in parentheses. In addition, long-run coefficients and standard errors are included below the short-run coefficients and standard errors in square parentheses. Regional and Time dummies are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Table 3.14 LR and Negative Binomial models by cigarette consumption

Dependent variable	Cigarette consumption					
	LR			NB		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables:	Full	Male	Female	Full	Male	Female
Extreme Optimism	0.0353	0.0074	0.0508	1.045***	1.026	1.089***
	(0.0561)	(0.1230)	(0.0451)	(0.016)	(0.018)	(0.033)
Moderate Optimism	0.0200	0.0202	0.0080	1.022**	1.018	1.024
	(0.0337)	(0.0735)	(0.0268)	(0.011)	(0.012)	(0.023)
Extreme Pessimism	-0.0525	-0.0548	-0.0493	1.004	1.026	0.903**
	(0.0623)	(0.1520)	(0.0477)	(0.022)	(0.025)	(0.043)
Moderate Pessimism	0.0231	0.0277	0.0181	1.011	1.007	1.012
	(0.0359)	(0.0834)	(0.0257)	(0.012)	(0.013)	(0.026)
Anti-Smoking Policy	-2.2090***	-4.8720***	0.2510	0.553***	0.544***	0.623***
	(0.7870)	(1.4970)	(0.5700)	(0.019)	(0.021)	(0.047)
Age	0.0957*	0.181*	-0.0127	1.011***	1.005***	1.030***
	(0.0515)	(0.0974)	(0.0377)	(0.001)	(0.001)	(0.002)
Income (log)	0.0379***	0.0513***	0.0246***	1.007***	1.006***	1.010***
	(0.0064)	(0.0132)	(0.0054)	(0.002)	(0.002)	(0.003)
Consumption (log)	0.0314	0.0143	0.0447***	1.023***	1.011	1.067***
	(0.0220)	(0.0518)	(0.0156)	(0.007)	(0.007)	(0.016)
Married	-0.0584	0.5130**	-0.0894	0.983	1.085***	0.909**
	(0.0806)	(0.2180)	(0.0675)	(0.022)	(0.030)	(0.037)
Cohabiting	0.1530	0.7810***	0.0085	1.190***	1.275***	1.135***
	(0.1080)	(0.2450)	(0.1080)	(0.028)	(0.036)	(0.046)
Divorced	0.3190***	0.7850**	0.2910***	1.306***	1.292***	1.306***
	(0.1090)	(0.3390)	(0.0956)	(0.037)	(0.048)	(0.058)

Primary Educ	-0.1360	-0.5570	0.194*	1.444***	1.395***	2.917***
	(0.2140)	(0.4780)	(0.1120)	(0.098)	(0.101)	(0.680)
Unfinished Secondary Educ	0.0780	0.0315	0.1960	1.463***	1.416***	1.698***
	(0.1570)	(0.3350)	(0.1210)	(0.046)	(0.052)	(0.107)
Secondary Educ	0.0654	0.0682	0.1130	1.327***	1.282***	1.549***
	(0.1320)	(0.3060)	(0.0898)	(0.039)	(0.044)	(0.086)
Vocational Secondary Educ	-0.1450	-0.3900	0.0807	1.156***	1.134***	1.247***
	(0.1130)	(0.2730)	(0.0823)	(0.033)	(0.038)	(0.067)
Self-reported Health	-0.2760***	-0.6950***	-0.0470	0.864***	0.851***	0.926**
	(0.0567)	(0.1490)	(0.0384)	(0.016)	(0.018)	(0.036)
Retired	-0.3970***	-0.9380***	-0.0977*	0.820***	0.857***	0.750***
	(0.0805)	(0.1880)	(0.0527)	(0.018)	(0.021)	(0.040)
Price_domestic cigarettes	-0.0369	-0.1520*	0.0337	1.034***	1.031**	1.049*
	(0.0389)	(0.0894)	(0.0285)	(0.013)	(0.014)	(0.027)
Price_imported cigarettes	0.0903*	0.2860***	-0.0396	0.999	0.996	1.006
	(0.0464)	(0.1050)	(0.0334)	(0.014)	(0.016)	(0.030)
Observations	101,645	40,622	61,023	36,638	25,652	10,986
R-squared	0.007	0.020	0.005			
Number of idind	23,985	10,262	13,723	7,086	5,013	2,073

Notes: This table reports the fixed-effects LR and the fixed effects Negative Binomial estimates for cigarette consumption. Coefficients are presented in LR, while incident rate ratio (IRR) are presented in Negative Binomial. Robust standard errors are given in parentheses. Regional and Time dummies are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Table 3.15 GMM models by cigarette consumption

Dependent variable	Cigarette consumption		
	GMM		
	(1)	(2)	(3)
Variables:	Full	Male	Female
Extreme Optimism	0.100 [0.313] (0.073) [0.249]	0.060 [0.186] (0.155) [0.460]	0.178*** [0.535***] (0.057) [0.188]
Moderate Optimism	-0.024 [-0.089] (0.042) [0.164]	-0.060 [-0.187] (0.089) [0.291]	0.021 [0.063] (0.033) [0.095]
Extreme Pessimism	-0.086 [-0.318] (0.075) [0.305]	-0.106 [-0.328] (0.188) [0.613]	-0.043 [-0.129] (0.053) [0.164]
Moderate Pessimism	-0.009 [-0.033] (0.045) [0.168]	-0.068 [-0.209] (0.105) [0.335]	0.043 [0.128] (0.032) [0.098]
lagged cigarette consumption	0.731*** (0.077)	0.679*** (0.113)	0.667*** (0.081)
Anti-Smoking Policy	-0.267*** (0.084)	-0.620*** (0.198)	-0.026 (0.078)
Age	-0.0109*** (0.003)	-0.0169** (0.007)	-0.0135*** (0.003)
Income (log)	0.008 (0.007)	0.004 (0.014)	0.007 (0.006)
Consumption (log)	0.024 (0.020)	-0.017 (0.046)	0.0493*** (0.016)
Married	-0.043	0.364	-0.0708**

	(0.053)	(0.289)	(0.035)
Cohabiting	0.381**	0.964**	0.323**
	(0.174)	(0.470)	(0.125)
Divorced	0.359**	1.044**	0.293***
	(0.141)	(0.490)	(0.092)
Primary Educ	0.520***	0.892***	0.415***
	(0.136)	(0.319)	(0.094)
Unfinished Secondary Educ	0.881***	1.641***	0.618***
	(0.227)	(0.541)	(0.134)
Secondary Educ	0.734***	1.402***	0.459***
	(0.184)	(0.462)	(0.096)
Vocational Secondary Educ	0.338***	0.730***	0.209***
	(0.097)	(0.262)	(0.058)
Self-reported Health	-0.170***	-0.494***	-0.016
	(0.049)	(0.134)	(0.033)
Retired	-0.311***	-0.601***	-0.176***
	(0.089)	(0.202)	(0.056)
Price_domestic cigarettes	-0.0102	-0.0653	0.0491
	(0.0463)	(0.1030)	(0.0369)
Price_imported cigarettes	0.0004	-0.0747	0.0461
	(0.0482)	(0.1080)	(0.0380)
Hansen-J test (p value)	0.822	0.626	0.568
M4 (p value)	0.587	0.702	0.602
Observations	97,370	38,109	59,261
Number of individual	23,688	10,045	13,643

Notes: This table reports the system GMM estimates for cigarette consumption. M4 is a test for fourth-order serial correlation of the residuals in the differenced equations, under the null hypothesis that there is no fourth-order serial correlation. The Hansen J-test of over-identifying is aimed at testing the exogeneity of instruments. The null hypothesis of the Hansen J-test is that the instruments are not correlated with the error term. I use the fourth to sixth lag of dependent variable as instruments. Standard errors are given in parentheses. In addition, long-run coefficients and standard errors are included below the short-run coefficients and standard errors in square parentheses. Regional and Time dummies are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Table 3.16 Interaction results for binge and problem drinkings

	Binge Drinking (Drink 200)			Problem Drinking (Drink150)		
Panel A: LPM						
	Full	Male	Female	Full	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
EOPT_LDS	0.0075*	0.0228**	-0.0011	0.0131***	0.0323***	0.0014
	(0.0040)	(0.0095)	(0.0028)	(0.0049)	(0.0111)	(0.0040)
EOPT_(1-LDS)	0.0017	0.0109	-0.0042*	-0.0017	0.0069	-0.0071*
	(0.0036)	(0.0077)	(0.0025)	(0.0047)	(0.0096)	(0.0041)
Diff (P value)	0.271	0.318	0.422	0.026	0.075	0.129
Obs	119,613	47,996	71,617	119,613	47,996	71,617
Number of ID	26,451	11,389	15,062	26,451	11,389	15,062
Panel B: Logit						
	Full	Male	Female	Full	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
EOPT_LDS	1.188*	1.284**	0.919	1.221***	1.287***	1.051
	(0.108)	(0.131)	(0.190)	(0.0879)	(0.108)	(0.150)
EOPT_(1-LDS)	1.068	1.157	0.745	0.987	1.057	0.783*
	(0.0969)	(0.115)	(0.169)	(0.0685)	(0.0843)	(0.114)
Diff (P value)	0.381	0.442	0.473	0.026	0.076	0.129
Obs	18,565	14,873	3,692	29,638	22,031	7,607
Number of ID	2,765	2,269	496	4,500	3,435	1,065
Panel C: GMM						
	Full	Male	Female	Full	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)

EOPT_LDS	0.0094**	0.0212**	0.00099	0.0177***	0.0349***	0.0044
	[0.0097**]	[0.0225**]	[0.0009]	[0.0226]	[0.0389***]	[0.0038]
	(0.0040)	(0.0099)	(0.0032)	(0.0053)	(0.0116)	(0.0039)
	[0.0042]	[0.1098]	[0.0030]	[0.0080]	[0.0147]	[0.0034]
EOPT_(1-LDS)	0.0055	0.0148*	-0.0021	-0.0009	0.0082	-0.0024
	[0.0056]	[0.0157**]	[-0.0020]	[-0.0011]	[0.0091]	[-0.0020]
	(0.0035)	(0.0077)	(0.0020)	(0.0054)	(0.0101)	(0.0047)
	[0.0034]	[0.0079]	[0.0019]	[0.0070]	[0.1098]	[0.0042]
Diff (P value)	0.449	0.599	0.459	0.013	0.078	0.246
Obs	118,821	47,265	71,556	118,821	47,265	71,556
Number of ID	26,366	11,307	15,059	26,366	11,307	15,059

Notes: This table presents the interaction results for binge and problem drinkings. Diff is the Wald test performed to examine whether the coefficient associated with *EOPT_LDS* is significantly different from the coefficient associated with *EOPT_(1-LDS)*. Coefficients are presented in LR and GMM, while odds ratio (OR) are presented in Logit. Robust standard errors are given in parentheses. In addition, long-run coefficients and standard errors are included below the short-run coefficients and standard errors in square parentheses. Other control variables include regressors listed in Appendix 3.1. Regional and Time dummies and PSU level of prices are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Table 3.17 Interaction results for alcohol consumption

Alcohol Consumption						
Panel A: LR and GMM						
	LR			GMM		
	Full	Male	Female	Full	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
EOPT_LDS	1.985*	6.564***	-0.726	3.333** [10.606**]	9.019*** [18.655**]	0.618 [1.3762]
	(1.0640)	(2.3750)	(0.9140)	(1.4200) [5.3843]	(2.8340) [7.6271]	(1.1840) [2.5434]
EOPT_(1-LDS)	-1.1850	0.5820	-2.265**	-1.7210 [-5.4837]	-0.3620 -0.7485	-1.4520 [-3.2348]
	(0.9810)	(1.9640)	(0.9020)	(1.3290) [5.2797]	(2.3790) [4.9746]	(1.1610) [3.1043]
Diff (P value)	0.025	0.045	0.225	0.006	0.009	0.190
Obs	119,613	47,996	71,617	118,821	47,265	71,556
Number of ID	26,451	11,389	15,062	26,366	11,307	15,059
Panel B: Quantile regressions for the full sample						
	10th	30th	50th	70th	90th	
	(1)	(2)	(3)	(4)	(5)	
EOPT_LDS	2.951	3.552	4.311	5.480	6.416	
	(31.87)	(24.95)	(16.26)	(3.798)	(8.828)	
EOPT_(1-LDS)	-2.517	-1.956	-1.248	-0.158	0.715	
	(28.41)	(22.24)	(14.49)	(3.386)	(7.870)	
Diff (P value)	0.894	0.864	0.792	0.252	0.618	
Obs	62,320	62,320	62,320	62,320	62,320	
Number of ID	18,719	18,719	18,719	18,719	18,719	
Panel C: Quantile regressions for the male sample						
	10th	30th	50th	70th	90th	

	(1)	(2)	(3)	(4)	(5)
EOPT_LDS	6.705	8.183	10.03	12.66**	14.73*
	(14.13)	(10.88)	(7.230)	(5.134)	(7.896)
EOPT_(1-LDS)	-1.642	-0.636	0.622	2.409	3.822
	(12.00)	(9.241)	(6.142)	(4.362)	(6.708)
Diff (P value)	0.638	0.520	0.302	0.113	0.2736
Obs	31,169	31,169	31,169	31,169	31,169
Number of ID	9,156	9,156	9,156	9,156	9,156
	Panel D: Quantile regressions for the female sample				
	10th	30th	50th	70th	90th
	(1)	(2)	(3)	(4)	(5)
EOPT_LDS	0.0175	-0.266	-0.619	-1.216	-1.705
	(3.129)	(2.444)	(2.294)	(3.771)	(5.577)
EOPT_(1-LDS)	-3.179	-3.104	-3.010	-2.850	-2.720
	(2.882)	(2.252)	(2.113)	(3.474)	(5.137)
Diff (P value)	0.441	0.381	0.432	0.744	0.890
Obs	31,151	31,151	31,151	31,151	31,151
Number of ID	9,563	9,563	9,563	9,563	9,563

Notes: This table presents the interaction results for alcohol consumption. Diff is the Wald test performed to examine whether the coefficient associated with *EOPT_LDS* is significantly different from the coefficient associated with *EOPT_(1-LDS)*. Coefficients are presented in all models. Robust standard errors are given in parentheses. In addition, long-run coefficients and standard errors are included below the short-run coefficients and standard errors in square parentheses. Other control variables include regressors listed in Appendix 3.1. Regional and Time dummies and PSU level of prices are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Table 3.18 Interaction results for smoking

Smoke			
Panel A: LPM			
	Full	Male	Female
	(1)	(2)	(3)
EOPT_LDS	0.0071*	-0.0007	0.0101**
	(0.0042)	(0.0083)	(0.0047)
EOPT_(1-LDS)	0.0014	0.0016	0.0016
	(0.0045)	(0.0083)	(0.0050)
Diff (P value)	0.335	0.834	0.204
Obs	97,889	38,699	59,190
Number of ID	23,794	10,160	13,634
Panel B: Logit			
	Full	Male	Female
	(1)	(2)	(3)
EOPT_LDS	1.246**	1.084	1.408**
	(0.134)	(0.165)	(0.218)
EOPT_(1-LDS)	1.112	1.084	1.097
	(0.105)	(0.140)	(0.155)
Diff (P value)	0.411	0.998	0.215
Obs	15,447	9,470	5,977
Number of ID	2,510	1,536	974
Panel C: GMM			
	Full	Male	Female
	(1)	(2)	(3)
EOPT_LDS	0.0145*** [0.0525**]	0.0055 [0.0164]	0.0193*** [0.0696**]
	(0.0053) [0.0207]	(0.0099) [0.0301]	(0.0057) [0.0283]
EOPT_(1-LDS)	0.0075 [0.0270]	0.0095 [0.0286]	0.0059 [0.0212]

	(0.0056) [0.0196]	(0.0093) [0.0277]	(0.0064) [0.0221]
Diff (P value)	0.333	0.756	0.092
Obs	97,049	37,982	59,067
Number of ID	23,657	10,035	13,622

Notes: This table presents the interaction results for smoking. Diff is the Wald test performed to examine whether the coefficient associated with *EOPT_LDS* is significantly different from the coefficient associated with *EOPT_(1-LDS)*. Coefficients are presented in LPM and GMM, while odds ratio (OR) are presented in Logit. Robust standard errors are given in parentheses. In addition, long-run coefficients and standard errors are included below the short-run coefficients and standard errors in square parentheses. Other control variables include regressors listed in Appendix 3.1. Regional and Time dummies and PSU level of prices are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Table 3.19 Interaction results for cigarette consumption

Cigarette consumption			
Panel A: Full sample			
	LR	NB	GMM
	(1)	(2)	(3)
EOPT_LDS	0.152**	1.060***	0.162 [0.6114*]
	(0.077)	(0.023)	(0.0997) [0.3684]
EOPT_(1-LDS)	-0.039	1.018	0.0190 [0.0716]
	(0.080)	(0.023)	(0.0932) [0.3475]
Diff 1.	0.076	0.180	0.264
MOPT_LDS	0.103*	1.023	0.0253 [0.0956]
	(0.053)	(0.018)	(0.0656) [0.2444]
MOPT_(1-LDS)	-0.021	1.018	-0.0616 [-0.2325]
	(0.043)	(0.014)	(0.0504) [0.2112]
Diff 2.	0.060	0.800	0.270
Obs	97,889	34,781	97,049
Number of ID	23,794	6,917	23,657
Panel B: Male sample			
	LR	NB	GMM
	(1)	(2)	(3)
EOPT_LDS	0.152	1.030	0.0961 [0.3018]
	(0.177)	(0.026)	(0.2190) [0.6760]
EOPT_(1-LDS)	-0.051	1.010	-0.0208 [-0.0653]
	(0.166)	(0.026)	(0.1880) [0.5983]

Diff.	0.379	0.576	0.666
Obs	38,699	10,665	37,982
Number of ID	10,160	2,043	10,035
Panel C: Female sample			
	LR	NB	GMM
	(1)	(2)	(3)
EOPT_LDS	0.0956	1.109***	0.232*** [0.6936***]
	(0.064)	(0.044)	(0.0794) [0.2619]
EOPT_(1-LDS)	0.003	1.040	0.0986 [0.2948]
	(0.065)	(0.046)	(0.0761) [0.2332]
Diff 1.	0.300	0.253	0.210
EPES_LDS	-0.128*	0.835**	-0.0820 [-0.2450]
	(0.072)	(0.063)	(0.0798) [0.2463]
EPES_(1-LDS)	-0.013	0.943	-0.0100 [-0.0298]
	(0.062)	(0.059)	(0.0718) [0.2156]
Diff 2.	0.216	0.208	0.505
Obs	59,190	10,665	59,067
Number of ID	13,634	2,043	13,622

Notes: This table presents the interaction results for cigarette consumption. Diff is the Wald test performed to examine whether the coefficient associated with *EOPT_LDS* is significantly different from the coefficient associated with *EOPT_(1-LDS)*. Coefficients are presented in LR and GMM, while incident rate ratio (IRR) are presented in NB. Robust standard errors are given in parentheses. In addition, long-run coefficients and standard errors are included below the short-run coefficients and standard errors in square parentheses. Other control variables include regressors listed in Appendix 3.1. Regional and Time dummies and PSU level of prices are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Table 3.20 Variable additional tests in the full sample (*num_wave* and *next_wave*)

VARIABLES	Drink200	Drink150	Alcohol	Smoke	Cigarettes	Drink200	Drink150	Alcohol	Smoke	Cigarettes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
next_wave	-0.0010	0.0009	0.4187	-0.0017	0.0173					
	(0.0015)	(0.0020)	(0.4197)	(0.0017)	(0.0302)					
num_wave						0.0014*	-0.0004	-0.1091	0.0030**	0.0080
						(0.0008)	(0.0011)	(0.2555)	(0.0013)	(0.0228)
Extreme Optimism	0.0054**	0.0065*	0.6262	0.0040	0.0736	0.0055**	0.0065*	0.6134	0.0040	0.0729
	(0.0026)	(0.0034)	(0.7331)	(0.0029)	(0.0518)	(0.0026)	(0.0034)	(0.7332)	(0.0029)	(0.0518)
Moderate Optimism	0.0019	0.0007	0.1133	0.0014	0.0341	0.0019	0.0007	0.1088	0.0014	0.0338
	(0.0016)	(0.0020)	(0.4285)	(0.0018)	(0.0309)	(0.0016)	(0.0020)	(0.4284)	(0.0018)	(0.0309)
Extreme Pessimism	0.0016	0.0017	0.5926	-0.0021	-0.0313	0.0016	0.0017	0.5795	-0.0021	-0.0317
	(0.0030)	(0.0041)	(0.8418)	(0.0032)	(0.0582)	(0.0030)	(0.0041)	(0.8414)	(0.0032)	(0.0582)
Moderate Pessimism	0.0001	-0.0003	0.5165	0.0004	0.0328	0.0001	-0.0003	0.5143	0.0004	0.0327
	(0.0016)	(0.0022)	(0.4579)	(0.0019)	(0.0336)	(0.0016)	(0.0022)	(0.4578)	(0.0019)	(0.0336)
Anti-Drinking Policy	-0.0595*	-0.1070**	-53.6092***			-0.0751**	-0.1028**	-52.3781***		
	(0.0343)	(0.0494)	(12.8312)			(0.0358)	(0.0511)	(13.1362)		
Anti-Smoking Policy				-0.1904***	-2.1932***				-0.2236***	-2.2791***
				(0.0684)	(0.7567)				(0.0695)	(0.7901)
Age	0.0018	0.0049	2.2527***	0.0081*	0.0970*	0.0017	0.0049	2.2530***	0.0080*	0.0967*
	(0.0023)	(0.0033)	(0.8481)	(0.0045)	(0.0496)	(0.0023)	(0.0033)	(0.8476)	(0.0045)	(0.0497)
Income (log)	0.0008**	0.0006	0.5083***	0.0012***	0.0363***	0.0008**	0.0006	0.5083***	0.0012***	0.0363***
	(0.0003)	(0.0004)	(0.0855)	(0.0004)	(0.0061)	(0.0003)	(0.0004)	(0.0855)	(0.0004)	(0.0061)
Consumption (log)	0.0008	0.0016	0.8462***	0.0004	0.0389*	0.0008	0.0016	0.8489***	0.0004	0.0390*
	(0.0009)	(0.0012)	(0.2751)	(0.0012)	(0.0213)	(0.0009)	(0.0012)	(0.2751)	(0.0012)	(0.0213)
Married	-0.0035	0.0008	0.5791	-0.0187***	-0.1127	-0.0034	0.0008	0.5762	-0.0185***	-0.1117
	(0.0030)	(0.0041)	(0.9755)	(0.0050)	(0.0752)	(0.0030)	(0.0041)	(0.9763)	(0.0050)	(0.0752)
Cohabiting	0.0064	0.0103*	3.0934**	0.0022	0.0727	0.0065	0.0103*	3.0923**	0.0024	0.0738

	(0.0044)	(0.0058)	(1.2719)	(0.0064)	(0.0997)	(0.0044)	(0.0058)	(1.2713)	(0.0064)	(0.0997)
Divorced	0.0054	0.0133**	2.9914**	0.0166**	0.2608**	0.0056	0.0132**	2.9785**	0.0171**	0.2622**
	(0.0043)	(0.0060)	(1.3329)	(0.0070)	(0.1030)	(0.0043)	(0.0060)	(1.3330)	(0.0070)	(0.1030)
Primary Educ	-0.0106	-0.0205	-7.8827**	0.0013	-0.1634	-0.0106	-0.0204	-7.8523**	0.0013	-0.1614
	(0.0107)	(0.0144)	(3.3068)	(0.0131)	(0.1997)	(0.0107)	(0.0144)	(3.3078)	(0.0131)	(0.1997)
Unfinished	-0.0113*	-0.0112	-6.9368***	-0.0058	-0.0030	-0.0114*	-0.0111	-6.9025***	-0.0059	-0.0014
Secondary Educ	(0.0059)	(0.0080)	(1.8083)	(0.0086)	(0.1490)	(0.0059)	(0.0080)	(1.8085)	(0.0086)	(0.1490)
Secondary Educ	-0.0056	-0.0052	-3.9589***	-0.0046	-0.0030	-0.0057	-0.0052	-3.9274***	-0.0047	-0.0013
	(0.0045)	(0.0063)	(1.4459)	(0.0074)	(0.1212)	(0.0045)	(0.0063)	(1.4465)	(0.0074)	(0.1212)
Vocational	-0.0089**	-0.0051	-3.5058***	-0.0127*	-0.1830*	-0.0089**	-0.0051	-3.4849***	-0.0128*	-0.1823*
Secondary Educ	(0.0040)	(0.0056)	(1.2836)	(0.0067)	(0.1068)	(0.0040)	(0.0056)	(1.2839)	(0.0067)	(0.1068)
Self-reported	-0.0044**	-0.0114***	-6.1824***	-0.0161***	-0.2867***	-0.0044**	-0.0114***	-6.1871***	-0.0162***	-0.2877***
Health	(0.0020)	(0.0028)	(0.6725)	(0.0030)	(0.0533)	(0.0020)	(0.0028)	(0.6722)	(0.0030)	(0.0533)
Retired	-0.0023	-0.0153***	-4.9851***	-0.0102**	-0.3882***	-0.0023	-0.0152***	-4.9681***	-0.0103**	-0.3876***
	(0.0028)	(0.0041)	(0.9264)	(0.0041)	(0.0747)	(0.0028)	(0.0041)	(0.9260)	(0.0041)	(0.0746)
Observations	120,100	120,100	120,100	101,645	101,645	120,100	120,100	120,100	101,645	101,645
R-squared	0.0019	0.0022	0.0110	0.0112	0.0074	0.0019	0.0022	0.0109	0.0113	0.0074
Number of idind	26,495	26,495	26,495	23,985	23,985	26,495	26,495	26,495	23,985	23,985

Notes: This table presents the FE results of the main models in the full sample, controlling for *num_wave* and *next_wave*. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Coefficients are presented, and robust standard errors are given in parentheses. Region and Year dummies are included but not reported. A full list of the variables used is defined in Appendix 3.1

Table 3.21 Variable additional tests in the male sample (*num_wave* and *next_wave*)

VARIABLES	Drink200	Drink150	Alcohol	Smoke	Cigarettes	Drink200	Drink150	Alcohol	Smoke	Cigarettes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
next_wave	-0.0001	0.0057	1.5576*	-0.0005	0.0437					
	(0.0034)	(0.0043)	(0.8833)	(0.0033)	(0.0665)					
num_wave						0.0018	-0.0030	-0.8740	0.0005	-0.0635
						(0.0020)	(0.0026)	(0.5516)	(0.0027)	(0.0522)
Extreme Optimism	0.0182***	0.0209***	3.8074**	0.0020	0.1200	0.0182***	0.0207***	3.7502**	0.0021	0.1191
	(0.0061)	(0.0074)	(1.5581)	(0.0056)	(0.1149)	(0.0061)	(0.0074)	(1.5583)	(0.0056)	(0.1148)
Moderate Optimism	0.0045	0.0031	0.6594	0.0007	0.0401	0.0045	0.0030	0.6357	0.0007	0.0397
	(0.0035)	(0.0045)	(0.9052)	(0.0035)	(0.0687)	(0.0035)	(0.0045)	(0.9052)	(0.0035)	(0.0687)
Extreme Pessimism	0.0001	0.0009	0.3893	0.0023	0.0513	0.0001	0.0008	0.3778	0.0023	0.0513
	(0.0038)	(0.0049)	(1.0084)	(0.0039)	(0.0789)	(0.0038)	(0.0049)	(1.0079)	(0.0039)	(0.0789)
Moderate Pessimism	0.0039	-0.0013	0.6613	0.0027	-0.0346	0.0039	-0.0015	0.5887	0.0027	-0.0372
	(0.0079)	(0.0100)	(1.9849)	(0.0069)	(0.1432)	(0.0079)	(0.0100)	(1.9842)	(0.0069)	(0.1430)
Anti-Drinking Policy	-0.1196**	-0.1644*	-61.3557***			-0.1387**	-0.1327	-52.1134**		
	(0.0559)	(0.0854)	(22.3207)			(0.0603)	(0.0895)	(23.0274)		
Anti-Smoking Policy				-0.3964***	-4.7880***				-0.4019***	-4.1132***
				(0.0863)	(1.4381)				(0.0908)	(1.5133)
Age	0.0035	0.0063	2.0773	0.0162***	0.1824*	0.0034	0.0064	2.1077	0.0161***	0.1844**
	(0.0036)	(0.0056)	(1.4693)	(0.0056)	(0.0937)	(0.0037)	(0.0056)	(1.4670)	(0.0056)	(0.0931)
Income (log)	0.0021***	0.0010	0.6829***	0.0007	0.0532***	0.0021***	0.0010	0.6801***	0.0007	0.0529***
	(0.0007)	(0.0008)	(0.1736)	(0.0007)	(0.0130)	(0.0007)	(0.0008)	(0.1735)	(0.0007)	(0.0130)
Consumption (log)	0.0019	0.0013	0.5472	-0.0036	0.0248	0.0019	0.0013	0.5598	-0.0036	0.0252
	(0.0022)	(0.0029)	(0.6356)	(0.0026)	(0.0513)	(0.0022)	(0.0029)	(0.6359)	(0.0026)	(0.0513)
Married	0.0046	0.0274**	10.5516***	0.0088	0.3672*	0.0047	0.0273**	10.5268***	0.0088	0.3632*
	(0.0085)	(0.0116)	(2.5607)	(0.0124)	(0.2090)	(0.0085)	(0.0116)	(2.5640)	(0.0124)	(0.2091)
Cohabiting	0.0260**	0.0374***	12.6150***	0.0196	0.5348**	0.0261**	0.0372***	12.5604***	0.0197	0.5283**
	(0.0105)	(0.0136)	(2.8649)	(0.0132)	(0.2345)	(0.0105)	(0.0136)	(2.8635)	(0.0132)	(0.2341)

Divorced	0.0146	0.0536***	9.8398**	0.0283	0.6852**	0.0149	0.0529***	9.6297**	0.0284	0.6700**
	(0.0151)	(0.0203)	(4.2120)	(0.0185)	(0.3329)	(0.0151)	(0.0203)	(4.2051)	(0.0185)	(0.3326)
Primary Educ	-0.0106	-0.0283	-9.9811	-0.0148	-0.6510	-0.0103	-0.0284	-10.0162	-0.0148	-0.6574
	(0.0211)	(0.0311)	(6.7321)	(0.0277)	(0.4512)	(0.0211)	(0.0312)	(6.7411)	(0.0277)	(0.4515)
Unfinished Secondary Educ	-0.0179	-0.0088	-7.3581**	-0.0126	-0.1566	-0.0179	-0.0083	-7.2028**	-0.0126	-0.1522
	(0.0115)	(0.0159)	(3.5239)	(0.0167)	(0.3119)	(0.0115)	(0.0159)	(3.5256)	(0.0167)	(0.3117)
Secondary Educ	-0.0057	0.0051	-2.6309	-0.0149	-0.1251	-0.0057	0.0055	-2.5149	-0.0149	-0.1236
	(0.0094)	(0.0136)	(3.0452)	(0.0153)	(0.2771)	(0.0094)	(0.0136)	(3.0494)	(0.0153)	(0.2770)
Vocational Secondary Educ	-0.0170*	-0.0066	-5.2591*	-0.0370***	-0.5309**	-0.0170*	-0.0064	-5.2070*	-0.0370***	-0.5310**
	(0.0090)	(0.0125)	(2.7418)	(0.0142)	(0.2528)	(0.0090)	(0.0125)	(2.7419)	(0.0142)	(0.2527)
Self-reported Health	-0.0128**	-0.0298***	-13.1746***	-0.0373***	-0.7676***	-0.0129**	-0.0298***	-13.1711***	-0.0373***	-0.7639***
	(0.0055)	(0.0075)	(1.6738)	(0.0074)	(0.1427)	(0.0055)	(0.0075)	(1.6715)	(0.0074)	(0.1426)
Retired	-0.0005	-0.0275***	-7.5829***	-0.0208**	-0.9191***	-0.0007	-0.0270***	-7.4403***	-0.0208**	-0.9108***
	(0.0070)	(0.0099)	(2.1079)	(0.0092)	(0.1767)	(0.0070)	(0.0099)	(2.1042)	(0.0092)	(0.1765)
Observations	48,197	48,197	48,197	40,622	40,622	48,197	48,197	48,197	40,622	40,622
R-squared	0.0045	0.0057	0.0166	0.0301	0.0204	0.0045	0.0057	0.0166	0.0301	0.0205
Number of idind	11,408	11,408	11,408	10,262	10,262	11,408	11,408	11,408	10,262	10,262

Notes: This table presents the FE results of the main models in the male sample, controlling for *num_wave* and *next_wave*. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Coefficients are presented, and robust standard errors are given in parentheses. Region and Year dummies are included but not reported. A full list of the variables used is defined in Appendix 3.1

Table 3.22 Variable additional tests in the female sample (*num_wave* and *next_wave*)

VARIABLES	Drink200	Drink150	Alcohol	Smoke	Cigarettes	Drink200	Drink150	Alcohol	Smoke	Cigarettes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
next_wave	-0.0018*	-0.0026*	-0.4621	-0.0029*	-0.0072					
	(0.0010)	(0.0015)	(0.3553)	(0.0017)	(0.0232)					
num_wave						0.0007	0.0007	0.2068	0.0030**	0.0302*
						(0.0005)	(0.0008)	(0.2100)	(0.0013)	(0.0165)
Extreme Optimism	-0.0023	-0.0025	-1.3204**	0.0051	0.0490	-0.0023	-0.0024	-1.3096**	0.0051	0.0489
	(0.0018)	(0.0028)	(0.6451)	(0.0032)	(0.0422)	(0.0018)	(0.0028)	(0.6453)	(0.0032)	(0.0422)
Moderate Optimism	0.0001	-0.0008	-0.2942	0.0014	0.0205	0.0001	-0.0008	-0.2911	0.0014	0.0203
	(0.0011)	(0.0016)	(0.3813)	(0.0018)	(0.0244)	(0.0011)	(0.0016)	(0.3813)	(0.0018)	(0.0244)
Extreme Pessimism	0.0003	-0.0005	0.7420*	-0.0005	0.0228	0.0003	-0.0005	0.7431*	-0.0004	0.0228
	(0.0011)	(0.0016)	(0.3917)	(0.0019)	(0.0241)	(0.0011)	(0.0016)	(0.3916)	(0.0019)	(0.0241)
Moderate Pessimism	0.0005	0.0034	0.6112	-0.0045	-0.0245	0.0006	0.0035	0.6245	-0.0044	-0.0239
	(0.0019)	(0.0032)	(0.7138)	(0.0033)	(0.0450)	(0.0019)	(0.0032)	(0.7137)	(0.0033)	(0.0450)
Anti-Drinking Policy	-0.0088	-0.0614	-49.5282***			-0.0167	-0.0704	-51.9756***		
	(0.0418)	(0.0503)	(10.8451)			(0.0423)	(0.0515)	(11.1782)		
Anti-Smoking Policy				-0.0015	0.0347				-0.0358	-0.3077
				(0.0553)	(0.5453)				(0.0568)	(0.5767)
Age	-0.0002	0.0032	2.3673***	-0.0012	0.0000	-0.0002	0.0032	2.3719***	-0.0012	-0.0002
	(0.0028)	(0.0033)	(0.7192)	(0.0036)	(0.0360)	(0.0028)	(0.0033)	(0.7197)	(0.0037)	(0.0360)
Income (log)	-0.0001	0.0003	0.3636***	0.0014***	0.0224***	-0.0001	0.0003	0.3638***	0.0014***	0.0224***
	(0.0002)	(0.0003)	(0.0769)	(0.0004)	(0.0049)	(0.0002)	(0.0003)	(0.0769)	(0.0004)	(0.0049)
Consumption (log)	0.0001	0.0018**	1.0412***	0.0028**	0.0483***	0.0001	0.0018**	1.0390***	0.0028**	0.0482***
	(0.0006)	(0.0009)	(0.2162)	(0.0011)	(0.0148)	(0.0006)	(0.0009)	(0.2161)	(0.0011)	(0.0148)
Married	-0.0029	-0.0046	-2.1938***	-0.0179***	-0.1076*	-0.0029	-0.0046	-2.1859**	-0.0178***	-0.1053*
	(0.0022)	(0.0032)	(0.8478)	(0.0049)	(0.0628)	(0.0022)	(0.0032)	(0.8487)	(0.0049)	(0.0628)
Cohabiting	-0.0021	0.0019	-0.1354	0.0046	0.0147	-0.0021	0.0019	-0.1326	0.0047	0.0164
	(0.0037)	(0.0054)	(1.2445)	(0.0075)	(0.0979)	(0.0037)	(0.0054)	(1.2446)	(0.0075)	(0.0979)

Divorced	0.0041	0.0031	1.4867	0.0202***	0.2596***	0.0041	0.0031	1.5039	0.0205***	0.2631***
	(0.0033)	(0.0048)	(1.1656)	(0.0072)	(0.0896)	(0.0033)	(0.0048)	(1.1669)	(0.0072)	(0.0896)
Primary Educ	-0.0079	-0.0095	-5.2688*	0.0093	0.1484	-0.0081	-0.0098	-5.3192*	0.0091	0.1479
	(0.0100)	(0.0100)	(2.8158)	(0.0098)	(0.1115)	(0.0100)	(0.0100)	(2.8137)	(0.0098)	(0.1115)
Unfinished Secondary Educ	-0.0030	-0.0073	-5.3191***	0.0005	0.1621	-0.0032	-0.0075	-5.3471***	0.0004	0.1628
	(0.0051)	(0.0070)	(1.6980)	(0.0087)	(0.1212)	(0.0051)	(0.0070)	(1.6982)	(0.0087)	(0.1213)
Secondary Educ	-0.0053	-0.0114**	-4.6201***	0.0043	0.1146	-0.0054	-0.0116**	-4.6511***	0.0041	0.1145
	(0.0040)	(0.0053)	(1.3194)	(0.0071)	(0.0888)	(0.0040)	(0.0053)	(1.3192)	(0.0071)	(0.0888)
Vocational Secondary Educ	-0.0027	-0.0027	-1.9682	0.0060	0.0953	-0.0028	-0.0029	-1.9982	0.0057	0.0935
	(0.0037)	(0.0051)	(1.2484)	(0.0065)	(0.0842)	(0.0037)	(0.0051)	(1.2487)	(0.0065)	(0.0842)
Self-reported Health	0.0001	-0.0017	-2.5720***	-0.0051**	-0.0391	0.0001	-0.0017	-2.5746***	-0.0052**	-0.0407
	(0.0013)	(0.0020)	(0.5353)	(0.0025)	(0.0359)	(0.0013)	(0.0020)	(0.5356)	(0.0025)	(0.0359)
Retired	-0.0034**	-0.0078***	-3.2995***	-0.0053	-0.0999**	-0.0034**	-0.0079***	-3.3089***	-0.0053	-0.0988**
	(0.0017)	(0.0028)	(0.7424)	(0.0036)	(0.0495)	(0.0017)	(0.0028)	(0.7433)	(0.0036)	(0.0496)
Observations	71,903	71,903	71,903	61,023	61,023	71,903	71,903	71,903	61,023	61,023
R-squared	0.0013	0.0012	0.0110	0.0046	0.0046	0.0013	0.0012	0.0110	0.0048	0.0047
Number of idind	15,087	15,087	15,087	13,723	13,723	15,087	15,087	15,087	13,723	13,723

Notes: This table presents the FE results of the main models in the female sample, controlling for *num_wave* and *next_wave*. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Coefficients are presented, and robust standard errors are given in parentheses. Region and Year dummies are included but not reported. A full list of the variables used is defined in Appendix 3.1

Appendix 3.1 Definitions of Variables

Dependent Variables (*UB*)	**Value**
Drink:	
Binge drinking (drink200)	Binary variable equals to 1 if an individual usually consumed more than 200 grams pure alcohol per occasion during the last month, and 0 otherwise
Problem drinking (drink150)	Binary variable equals to 1 if an individual usually consumed more than 150 grams pure alcohol per occasion during the last month, and 0 otherwise
Alcohol consumption	Numerical continuous variable relating to the amount of pure alcohol consumed per occasion during the last month
Smoke:	
Smoking	Binary variable equals to 1 if an individual is a current smoker, and 0 otherwise
Cigarettes	Numerical discrete variable relating to numbers of cigarettes consumed per day during the last month
Independent Variables	**Value**
Errors in financial expectations:	
Extreme optimism	Binary variable equals to 1 if an individual is extreme optimistic, and 0 otherwise
Moderate optimism	Binary variable equals to 1 if an individual is moderate optimistic, and 0 otherwise
Correct prediction	Binary variable equals to 1 if an individual correctly predicts the future financial situation, and 0 otherwise
Moderate pessimism	Binary variable equals to 1 if an individual is moderate pessimistic, and 0 otherwise
Extreme pessimism	Binary variable equals to 1 if an individual is extreme pessimistic, and 0 otherwise
Subjective anxiety:	
Life dissatisfaction	Binary variable equals to 1 if an individual is 'less than satisfied' or 'not at all satisfied' with life in general, 0 otherwise
Control Variables	**Value**
Anti-smoking policy	Binary variable equals to 1 from year 2013 to year 2016, referring to the anti-smoking policy implemented during this period, and 0 otherwise
Anti-drinking policy	Binary variable equals to 1 from year 2010 to year 2016, referring to the anti-drinking policy implemented during this period, and 0 otherwise
Age	Continuous variable relating to individuals' age
Male	Binary variable equals to 1 if the respondent is male, and 0 for female
Log of Income	Continuous variable relating to logarithm of total individual income in the last 3 months

Log of consumption	Continuous variable relating to logarithm of total family consumption in the last 3 months
Married	Binary variable equals to 1 if an individual is married, and 0 otherwise
Cohabiting	Binary variable equals to 1 if an individual is living with partner but not registered, and 0 otherwise
Divorced	Binary variable equals to 1 if an individual is divorced, and 0 otherwise
Primary school	Binary variable equals to 1 if an individuals' highest level of education is primary school diploma, and 0 otherwise
Unfinished secondary school	Binary variable equals to 1 if an individual drops out of secondary school, and 0 otherwise
Secondary school	Binary variable equals to 1 if an individuals' highest level of education is secondary school diploma, and 0 otherwise
Vocational secondary school	Binary variable equals to 1 if an individuals' highest level of education is vocational secondary school diploma, and 0 otherwise
Self-Reported Health	Binary variable equals to 1 if an individual reports "Very bad" or "bad" health, and 0 otherwise
Retired	Binary variable equals to 1 if an individual is currently retired, and 0 otherwise.
Region 1	Region dummy variable equals to 1 if an individual lives in Moscow/St.Petersburg, and 0 otherwise
Region 2	Region dummy variable equals to 1 if an individual lives in Northern/Northwestern, and 0 otherwise
Region 3	Region dummy variable equals to 1 if an individual lives in Central/Central Black Earth, and 0 otherwise
Region 4	Region dummy variable equals to 1 if an individual lives in Volga-Vaytski/Volga Basin, and 0 otherwise
Region 5	Region dummy variable equals to 1 if an individual lives in Urals, and 0 otherwise
Region 6	Region dummy variable equals to 1 if an individual lives in Northern Caucasus, and 0 otherwise
Region 7	Region dummy variable equals to 1 if an individual lives in Western Siberia, and 0 otherwise
Region 8	Region dummy variable equals to 1 if an individual lives in Eastern Siberia/Far East, and 0 otherwise
Log of domestic cigarettes price	Continuous variable relating to logarithm of domestic cigarettes price in the local PSU
Log of imported cigarettes price	Continuous variable relating to logarithm of imported cigarettes price within the PSU
Log of vodka price	Continuous variable relating to logarithm of vodka price within the PSU
Log of cognac price	Continuous variable relating to logarithm of cognac price within the PSU
Log of dry wine price	Continuous variable relating to logarithm of dry wine price within the PSU

Log of fortified wine price	Continuous variable relating to logarithm of fortified wine price within the PSU
Log of beer price	Continuous variable relating to logarithm of beer price within the PSU
Others:	**Value**
Financial reality	Categorical variable ranging from 1(greatly worsened), 2(slightly worsened), 3(has not changed), 4(slightly improved) to 5(greatly improved)
Financial expectation	Categorical variable ranging from 1(live much worse), 2(somewhat worse), 3(no change), 4(live somewhat better), 5(live much better)

Appendix 3.2 Quantile regressions for alcohol consumption in the full sample

Dependent variable	Alcohol consumption				
	Quantile regressions				
	Full sample				
Variables:	10th	30th	50th	70th	90th
Extreme Optimism	0.356	0.978	1.766	2.969	3.941
	(2.646)	(1.995)	(2.58)	(4.975)	(7.194)
Moderate Optimism	0.558	0.618	0.694	0.810	0.904
	(1.645)	(1.24)	(1.604)	(3.093)	(4.472)
Extreme Pessimism	-0.506	0.172	1.030	2.339	3.398
	(3.336)	(2.515)	(3.254)	(6.274)	(9.071)
Moderate Pessimism	1.264	1.169	1.049	0.865	0.717
	(1.742)	(1.313)	(1.698)	(3.275)	(4.735)
Anti-Drinking Policy	-21.070	-23.650	-26.920	-31.900	-35.920
	(51.64)	(38.94)	(50.36)	(97.11)	(140.4)
Age	3.434	3.736	4.118	4.702	5.174
	(3.423)	(2.581)	(3.338)	(6.437)	(9.307)
Income (log)	0.300	0.375*	0.470*	0.614	0.731
	(0.29)	(0.218)	(0.282)	(0.545)	(0.787)
Consumption (log)	0.109	0.325	0.598	1.015	1.352
	(1.109)	(0.836)	(1.081)	(2.085)	(3.015)
Married	2.936	2.241	1.361	0.018	-1.067
	(3.643)	(2.747)	(3.553)	(6.851)	(9.906)
Cohabiting	3.293	3.117	2.894	2.554	2.279
	(4.066)	(3.066)	(3.965)	(7.646)	(11.06)
Divorced	6.636	6.857*	7.137	7.564	7.909
	(4.954)	(3.735)	(4.831)	(9.317)	(13.47)
Primary Educ	-4.431	-5.572	-7.016	-9.220	-11.000

	(16.88)	(12.72)	(16.46)	(31.74)	(45.89)
Unfinished Secondary Educ	-7.472	-7.915*	-8.477	-9.334	-10.030
	(6.174)	(4.655)	(6.021)	(11.61)	(16.79)
Secondary Educ	-5.241	-6.054	-7.083	-8.653	-9.923
	(5.054)	(3.811)	(4.929)	(9.505)	(13.74)
Vocational Secondary Educ	-4.370	-4.803	-5.352	-6.189	-6.866
	(4.191)	(3.16)	(4.087)	(7.881)	(11.4)
Self-reported Health	-3.589	-3.576	-3.558	-3.531	-3.510
	(3.055)	(2.303)	(2.979)	(5.745)	(8.307)
Retired	-4.400	-4.632*	-4.925	-5.373	-5.735
	(3.41)	(2.571)	(3.326)	(6.413)	(9.272)
Price_vodka	0.672	0.656	0.636	0.606	0.581
	(1.7560)	(1.3240)	(1.7120)	(3.3020)	(4.7740)
Price_fortified wine	-2.601*	-2.480**	-2.328*	-2.095	-1.906
	(1.3990)	(1.0550)	(1.3640)	(2.6310)	(3.8040)
Price_table wine	1.229	0.938	0.569	0.005	-0.450
	(1.3950)	(1.0520)	(1.3600)	(2.6230)	(3.7920)
Price_beer	-0.557	-0.295	0.037	0.545	0.955
	(2.0250)	(1.5270)	(1.9750)	(3.8080)	(5.5060)
Price_cognac	0.015	0.037	0.065	0.108	0.142
	(1.1840)	(0.8930)	(1.1550)	(2.2270)	(3.2200)
Observations	62,527	62,527	62,527	62,527	62,527
Number of idind	18,745	18,745	18,745	18,745	18,745

Notes: This table reports the fixed effects Quantile regression results for alcohol consumption in the full sample. Standard errors are given in parentheses. Regional and Time dummies and PSU level of prices are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Appendix 3.3 Quantile regressions for alcohol consumption in the female sample

Dependent variable	Alcohol consumption				
	Quantile regressions				
	Female sample				
Variables:	10th	30th	50th	70th	90th
Extreme Optimism	-1.374	-1.391	-1.413	-1.449	-1.479
	(41.04)	(32.63)	(22.02)	(4.72)	(11.36)
Moderate Optimism	0.551	0.409	0.229	-0.075	-0.324
	(25.37)	(20.17)	(13.61)	(2.917)	(7.024)
Extreme Pessimism	-0.075	0.019	0.139	0.342	0.507
	(48.48)	(38.55)	(26.02)	(5.576)	(13.42)
Moderate Pessimism	1.163	1.176	1.192	1.219	1.241
	(25.86)	(20.56)	(13.88)	(2.974)	(7.16)
Anti-Drinking Policy	-21.630	-21.560	-21.460	-21.300	-21.170
	(838.8)	(666.9)	(450.1)	(96.47)	(232.2)
Age	3.247	3.265	3.288	3.326	3.357
	(55.66)	(44.25)	(29.87)	(6.401)	(15.41)
Income (log)	0.212	0.271	0.345	0.472	0.575
	(4.563)	(3.628)	(2.448)	(0.525)	(1.263)
Consumption (log)	0.795	0.802	0.811	0.826	0.838
	(16.52)	(13.13)	(8.862)	(1.899)	(4.573)
Married	-0.829	-1.391	-2.101	-3.305	-4.287
	(53.24)	(42.33)	(28.57)	(6.123)	(14.74)
Cohabiting	-2.030	-2.422	-2.917	-3.755	-4.440
	(62.16)	(49.42)	(33.36)	(7.149)	(17.21)
Divorced	2.020	1.944	1.849	1.687	1.554
	(70.35)	(55.94)	(37.75)	(8.091)	(19.48)

Primary Educ	-3.320	-4.294	-5.527	-7.614	-9.318
	(339.9)	(270.3)	(182.4)	(39.09)	(94.11)
Unfinished Secondary Educ	-3.611	-4.959	-6.664	-9.553	-11.910
	(102.9)	(81.78)	(55.19)	(11.83)	(28.48)
Secondary Educ	-2.547	-4.119	-6.109	-9.479	-12.230
	(77.7)	(61.77)	(41.69)	(8.936)	(21.51)
Vocational Secondary Educ	-1.188	-1.982	-2.985	-4.686	-6.075
	(64.640)	(51.390)	(34.68)	(7.434)	(17.900)
Self-reported Health	-0.530	-0.285	0.024	0.548	0.976
	(46.090)	(36.650)	(24.730)	(5.301)	(12.760)
Retired	-3.160	-4.144	-5.388	-7.497	-9.219
	(50.240)	(39.950)	(26.960)	(5.778)	(13.910)
Price_vodka	0.118	0.081	0.035	-0.043	-0.108
	(26.8400)	(21.3400)	(14.4000)	(3.0860)	(7.4300)
Price_fortified wine	-0.677	-0.730	-0.797	-0.910	-1.002
	(21.5600)	(17.1500)	(11.5700)	(2.4800)	(5.9700)
Price_table wine	0.252	0.152	0.026	-0.187	-0.361
	(21.1700)	(16.8300)	(11.3600)	(2.4350)	(5.8610)
Price_beer	0.54	0.355	0.121	-0.275	-0.598
	(30.7300)	(24.4300)	(16.4900)	(3.5340)	(8.5080)
Price_cognac	0.299	0.344	0.401	0.499	0.578
	(18.2600)	(14.5200)	(9.7970)	(2.1000)	(5.0550)
Observations	31,242	31,242	31,242	31,242	31,242
Number of idind	9,579	9,579	9,579	9,579	9,579

Notes: This table reports the fixed effects Quantile regression results for alcohol consumption in the female sample. Standard errors are given in parentheses. Regional and Time dummies and PSU level of prices are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.

Chapter Four: How does errors in financial expectations affect individuals' borrowing decisions? - Evidence from South Africa

4.1 Introduction

In the past decades, consumer debt has grown rapidly and considerably in South Africa. Policy makers have seen debts as adrenaline doses to stimulate economic growth, and consumers have used debt as a way to redistribute future wealth and finance their current consumption. Due to greater demand and competitions in the credit markets, financial institutions have been fueling the growth of consumer debt in South Africa. According to the South African Reserve Bank (2017), the ratio of household debt to disposable income stands at 73% and has increased by over 40% since 1990. The growth rate of household debt has been increased faster than the growth rate of nominal income for over a decade. Furthermore, the interest rates in the early 2000s has halved from 14% to 7% and remained at around 9% ever since. The drop in interest rates reduces the interests payments and further stimulate the credit markets.

A wide range of new credit products have been developed and marketed to specific consumers from banks and other financial retailers, such as educational loans, store card loans, short-term loans and so on. In particular, a new type of debt has dominated the credit markets - unsecured debt. Unsecured debts (e.g., consumer and small business debts) that are not backed by assets or collateral. They becomes the fastest growing segment of South Africa's credit market and are essentially the country's own version of sub-prime loans. Unsecured debts have grown at

a 30 percent annual compounded rate since their introduction in 2007, when the National Credit Act was signed into law. Unsecured lending has become popular with banks because they are able to charge up to 31 annual interest rates, making these riskier debts far more profitable than mortgage and car debts in the low interest rate environment of the past decade. In comparison, South Africa's real GDP grew by 38% in the past decade, while unsecured lending surged by over 220%. In addition, household debts now accounts for around three-quarter of South African's disposable income (South African Reserve Bank, 2014).

Consumers utlise various types of debt to finance large-scale purchase of durable goods and services, or temporarily bridge income gaps over the business cycle. The frequent usage of consumer debt has sustained consumption growth and contributed to individuals' reluctance to save. However, there is a growing concern that consumers are at risk of accumulating an overwhelming level of debt and consequently sabotage their future consumption. In addition, a higher debt level limits consumers' ability to gain additional funds. Individuals with a high level of debt are found to be financially fragile as they have greater debt repayment burden and insufficient savings to shield them from negative shocks (financial crisis, income uncertainty) (Meniago *et al.*, 2013). Therefore, understanding the determinant of individual debt and over-indebtedness has been of much concern to policy makers in South Africa, who aim at reducing indebtedness in the country.

According to the Life-Cycle Hypothesis (LCH) (Modigliani and Brumberg, 1954), consumers maximise their intertemporal utility by choosing the optimal consumption path and the corresponding debt level subject to their preferences and their current and future financial

situations. Later studies (Carroll, 1992, 1997) stress the importance of individuals' financial expectations regarding income and consumption in determining their borrowing decisions. Empirical studies have generally documented a positive relationship between financial optimism and borrowing. Consumers with optimistic expectations regarding future income show a higher level of unsecured debt (Del Rio and Young, 2006) and secured debt (Brown *et al.*, 2005, 2008).

The role of financial expectations has attracted most of the attention in the literature, and few have investigated the consequences of making inaccurate financial expectations. To ascertain the effect of errors in financial expectations on debt, I compare individuals' expected income ranking at time period *t-1* to the actual ranking at time *t*. Individuals overestimating (underestimating) their actual income ranking are classified as displaying optimistic (pessimistic) forecast errors[20]. Grohmann *et al.* (2019) launch an experiment and find that participants with optimistic expectations are more likely to borrow higher levels of debt at time *t-1*. When the reality about the future is realised, over-optimistic participants have to adjust their consumption to correct the erroneous financial expectations. In some cases, overconfident participants are unable to pay off their debts completely and keep on borrowing.

Klove and Mehlum (2018) suggest that optimistic expectations are only problematic if the optimism is illusionary and consumers' income does not sufficiently increase to the level necessary to keep up with the repayments of principal and interests. Therefore, Hyytinen and

20 In this chapter, I define individuals with inaccurate financial expectations as being either optimistically or pessimistically biased. I use 'errors in financial expectations', 'optimistic and pessimistic biases', 'optimistic and pessimistic forecast errors' and 'illusionary optimism' interchangeably.

Putkuri (2018) indicate that only imprudent optimists who save less and borrow more are more likely to report being over-indebted and having problems with the repayments of debt.

Given the thriving of credit markets and the prevalence of borrowing decisions in South Africa, the objective of this chapter is to investigate the relationship between errors in financial expectations and borrowing decisions in South Africa. This chapter contributes to the existing literature in the following three ways. First, to the best of my knowledge, no other study has explored the consequences of errors in financial expectations on individuals' decision to borrow in South Africa. This could provide important implications for preventing people from being caught in debt traps. Next, most studies on the relationship between forecast errors and borrowing use cross-sectional or random effects models. This chapter uses fixed-effects estimates to eliminate the unobserved individual fixed traits that is prevalent among optimism literature (Hyytinen and Putkuri, 2018). Third, this chapter examines the moderating effects of low education and debt illiteracy on the relationship between forecast errors and borrowing.

The rest of this chapter is organised as follows. In section 4.2, I presents a review of theoretical and empirical studies on the relationship between errors in financial expectations and borrowing from developed and developing countries. Section 4.3 develops the hypotheses. Section 4.4 summarises the data and presents summary statistics. The baseline specifications and methodology are presented in section 4.5. Section 4.6 provides the empirical results and Section 4.7, a series of robustness tests. Section 4.8 concludes.

4.2 Literature Review

There is a large body of literature focused on the determinants of borrowing decisions in the economic and finance disciplines. This chapter first reviews the theoretical background on the determinants of debt. Next, empirical literature on the relationship between debt and optimism is presented from developed and developing countries respectively. This is followed by a review of empirical studies investigating the relationship between over-indebtedness and optimism. Last, the contribution of this chapter is presented.

4.2.1 Theoretical background

The most prominent theory on the determinants of consumption and borrowing decisions is the Life-Cycle Hypothesis (LCH) by Modigliani and Brumberg (1954), which has become the standard framework for modelling consumption and borrowing decisions. The LCH suggests that individuals choose to smooth their consumption over the life-span through borrowing at different ages. For example, young individuals borrow more to finance their current consumption since their income is relatively low compared to the average lifetime income. Middle-aged individuals usually have higher income due to being more skilled, experienced and knowledgeable, and hereby repay their loans and save for retirement. Retirees will dissave and spend down their wealth altogether.

Throughout the life-cycle, most individuals experience a rising income path, the debt to income ratio will be relatively high at the early stage of their lives and then decreases with age (Yilmazer and DeVaney, 2005). As a result, individuals maximise their utility by controlling their

consumption over time subject to their preferences and current and discounted future income and resources. The derived consumption level of individuals determines the optimal borrowing level. Thus, the LCH suggests that consumers expecting high (low) future income borrow more (less).

In support of the LCH, Cox and Jappelli (1993) use the 1983 Survey of Consumer Finances (SCF) and find that the desire for borrowing increases with age until the consumers reach the mid-thirties. The size of the gap between actual and desired debt is the largest for young individuals who suffer from liquidity constraints. Crook (2001) makes use of the 1995 SCF and suggests that the demand for borrowing decreases when individuals reach the age of 55.

Tobin (1967) notes that according to the LCH, individuals expecting higher income throughout their life should consume more than their current means at the early stage. This implies that consumers' dissaving at both ends of the life-cycle is financed by accumulated assets in middle age. Such behaviour suggests a positive relationship between expected income growth and debt, and in extreme cases, young agents could borrow more than they earn. Barro (1974) states that subject to borrowing constraints, people can borrow to smooth consumption, but they cannot consume more than their income and assets.

Over the years, several amendments have been proposed, and restrictive assumptions have been relaxed. For example by relaxing the assumption of quadratic preferences, Carroll (1992) finds that the "buffer-stock" version of LCH can better capture the divergent behaviours of consumers. His model implies that individuals withhold assets and resources to shield their consumption from unpredictable income fluctuations. He suggests that consumers can be

characterised as both “prudent” and “impatient” when facing income uncertainty. Prudent consumers have a precautionary savings motive and are reluctant to spend down their assets, while impatient consumers would borrow to finance current consumption if future income were known with certainty.

Carroll (1997) shows that under the target wealth-to-income ratio, the tension between prudence and impatience would lead to different borrowing behaviours. When wealth is below target, prudence dominates impatience and individuals will refuse to borrow and save as much as they can, while above target, impatience dominates the fear and individuals will borrow against future income. Thus, income uncertainty plays a key role in this model because when consumers become more optimistic about future income, their target buffer-stock decreases, and they increase their borrowings to smooth current consumption.

Additionally, Brown *et al.* (2005) present two versions of the life-cycle model. The first version is a two-period model with constant interest rate and time preference, perfect credit market and uncertain future income. This model predicts a positive correlation between the optimal level of the debt and the optimistic expectations of future income. The second version adds a third period to control for the possible scenarios of repayment difficulties and loan extensions. The model once again predicts that optimistic consumers would take on as much debt as possible to smooth life-time consumption. Klove and Mehlum (2018) extend the analysis and find that consumers’ willingness to borrow is positively correlated with their optimistic expectations about future income. In extreme cases, without any borrowing constraints,

optimistic consumers, under any bounded interest rate, will borrow as much as they can and eventually end up in debt traps.

4.2.2 Optimism/pessimism and borrowing

This section presents the empirical studies focussed on the relationship between financial expectations, errors in financial expectations and borrowing decisions in developed and developing counties. As only a few studies explicitly explore the role of errors in financial expectations, studies investigating the role of financial expectations in general are also included.

4.2.2.1 Optimism/pessimism and borrowing in developed countries

4.2.2.1.1 The effects of financial expectations

Del Rio and Young (2006) use the 1995 and 2000 waves of the British Household Panel Survey (BHPS) with a pooled sample of 14,369 respondents to examine the determinants of unsecured debt. They employ cross-sectional Probit and the ordinary least squares (OLS) models to estimate the debt participation and the amount of debt for participators. Positive financial expectations regarding one's future financial position are included as one of the regressor. The estimation results show that optimistic financial expectations are significantly and positively associated with the probability of debt participation and the amount of unsecured debt held in both 1995 and 2000.

Lee (2008) extends the analysis to examine the relationship between consumer sentiment and household debt decisions using the 2004 US Survey of Consumer Finances (SCF) with 4,519 observations. Two dependent variables are used: one is a dummy variable indicating whether a

household has any debt, and another is the amount of total debt measured by the sum of mortgage, installment debt, credit debt and other debts. Consumer sentiment is measured by consumers expectations regarding future income in the next year. The expectation responses are divided into four categories: positive, neutral, negative and no expectations. To control for selection bias, the author employs a traditional Heckit model and finds that only those with positive expectation about future income show a higher probability of incurring debt and a higher amount of debt. This finding is in line with the LCH as it suggests that those who are optimistic about future income and have the confidence to pay off debts are likely to smooth consumption through borrowing. However, this paper is limited by the use of only one wave of data.

Brown *et al.* (2005) use a sample of 5405 observations from the 1995 and 2000 waves of the BHPS to explore the determinants of unsecured debt with an emphasis on the role of financial expectations at both the individual and household level. They construct a financial expectations index (FEI), which is coded as 0 if an individual expects to be financially worse-off in the next 12 months, 1 if the individual expects to be the same, and 2 if the individual expects to be financially better off. Since the amount of outstanding debt is a censored variable, they employ a random-effects Tobit model at the individual level, and a cross-sectional Tobit model at the household level. The results show that the FEI, which represents individuals' expectations ranking from pessimistic to optimistic, is positively associated with the amount of individual and household debt. The study also examines how dynamic expectations affect debt formation by including 5 lags of FEI, and find that only the current and one lag of FEI are significant determinants of debt.

Additionally, the authors conduct Granger causality tests, which suggest that the direction of causality runs from financial expectations to debt. The study also examines the determinants of growth in debt and finds that optimistic expectation is associated with a higher growth rate in debt. Finally, Brown *et al.* (2005) construct various dummy variables to ascertain the accuracy of individuals financial expectations. They find that individuals who make correct or incorrect financial predictions are not significantly associated with the growth in debt. However, this study fails to examine the role of pessimistic expectations in determining unsecured debt. The authors conclude that it is financial expectations rather than the accuracy of financial expectations, which is an important determinant of debt formation at both individual and household levels.

Brown *et al.* (2008) extend the literature by employing a panel of 42,894 observations from the 1993-2001 BHPS to examine the relationship between mortgage debt (secured debt) and financial expectations at the household level. This study constructs two financial expectations variables: one is the FEI of the head of household as in Brown *et al.* (2005). Another is the household financial expectation index (HFEI), which is measured as the sum of the FEI of all household members. The authors employ a two-stages analysis: first they use a multinomial Logit (MNL) model and find that the optimistic financial expectations are positively associated with the probability of owning a house with mortgage. The second stage uses OLS to estimate the determinants of the level of mortgage. The estimated coefficients of the FEI and HFEI are both positive and significant at the 1% level indicating that households' optimism regarding their future financial situation induces higher mortgage debt.

4.2.2.1.2 The effects of errors in financial expectations

The aforementioned literature explicitly focuses on the role of financial expectations, and few studies have examined the consequences of overborrowing brought by optimistic financial expectations.

Grohmann *et al.* (2019) launch a German-based experiment to examine the relationship between errors in financial expectations and the level of borrowing and overborrowing. They employ a between-subject experiment with one control group and two treatment groups. In the treatment group, participants earn money from answering a general knowledge quiz. In the control group, participants' income is determined by random lotteries. Two treatment groups are assigned different quizzes (easy or hard), and the authors find that the easy quiz group is more optimistic about future income compared to the hard quiz and control groups. As a result, the easy (hard) quiz group is more likely to overestimate (underestimate) their actual income, and those who overestimate (underestimate) actual income compared to their expected income are classified as overconfident (underconfident) participants.

There are two rounds of quiz, and after each round, participants can enter the free market to consume with a choice to borrow interest-free debt. The study uses a linear probability model (LPM) and finds that compared to the underconfident group (hard treatment), the overconfident group (easy treatment), is related to a higher probability of debt engagement. As the participants enter the second round, they will gradually correct their erroneous income expectations. However, overconfident participants are unable to completely pay off their initial debt and keep on

borrowing. Thus, the results show that overconfident participants have a higher likelihood of overborrowing in the final stage.

Hyytinen and Putkuri (2018) explore the relationship between household optimism and over-indebtedness using a representative sample of 91,827 household observations from the 1994-2013 Finnish Income Distribution Statistics (IDS). A household is classified as optimistically (pessimistically) biased if their financial expectations in year *t-1* fall short of (better than) the realization in year *t*. As a result, two forecast error measures are constructed: the first measure, FE1, is a continuous variable coded as 1 for an extremely pessimistic forecast error, 2 for a moderately pessimistic forecast error, 3 for no forecast error, 4 for a moderately optimistic forecast error, 5 for an extremely optimistic forecast error.

The second measure, FE2, has four categories, which are pessimistic forecast error (1), no forecast error (2), prudentially optimistic forecast error (3) and nonprudentially optimistic forecast error (4). Within this measure, optimistic forecast errors are further divided into prudent optimism if a household is optimistically biased but reports their financial situation does not get worse as a result, and nonprudent optimism if a household is optimistically biased but reports their financial situation actually gets worse.

The authors find that overoptimistic (overpessimistic) households show higher (lower) debt-to-income ratios, and those who made extremely optimistic forecast error are more likely to report over-indebtedness and to struggle to pay their bills and debts. The results also show that nonprudentially optimistic household are more likely to be over-indebted and report having

troubles to make ends meet. This suggests that over-optimistic households may be financially fragile.

4.2.2.2 Optimism/pessimism and borrowing in developing countries

Only a handful of research explore the relationship between financial expectations and borrowing decisions in developing countries, of which none of them focus on the role of errors in financial expectations.

Semenova (2011) utilises data on 5,186 household observations from the 2008 Russian Longitudinal Monitoring Survey (RLMS) to examine the relationship between financial expectations and debt in Russia. She first uses a Probit model and finds that being optimistic about the future financial situation in the next year is associated with a higher probability of holding a bank loan. The study also employs the Heckman maximum likelihood method and shows that households expecting no changes in financial situation for the next year borrow 39,000-55,000 more roubles than pessimistic households. Financial optimism is not related to the amount of borrowing, which is in contrast to the previous findings (Brown *et al.*, 2005; Lee, 2008). However, this study is limited by the cross-sectional structure employed and limited information regarding household debt (the author only focuses on bank loan).

Klove and Mehlum (2018) utilise data on 84,654 observations from four waves (2008-2014) of the South Africa National Income Dynamic Survey (NIDS) to examine the relationship between optimism and borrowing. In the questionnaire, individuals are asked to rank their financial situation based on a six-step ladder where the richest are at the top, and the poorest at

the bottom. Answers are collected from individuals' ranking today and in 2 years. Optimists are defined as individuals expecting a higher ranking than today's ranking. Focusing on the poor, the authors employ a pooled OLS model and limit the sample to those ranking themselves among the bottom third. They find that individuals who expect to move up two/three/four steps respectively are associated with a 3.5/5.2/7 percentage point higher probability of holding any debt. This shows a significant gradient between optimism and debt. Restricting the sample to the bottom sixth, the authors find the positive relationship between optimism and debt withholding is still prevalent but to a lesser degree of significance. They suggest that the decreases in sample size reduces the estimation precision.

Additionally, this study investigates different sources of borrowing and different types of debt. The results show that optimists mainly borrow informal sources instead of formal sources, as the access to bank loan is difficult for low-income group in South Africa. Moreover, extreme optimists are associated with a 11 percentage point higher probability of taking up expensive loans (loans from credit card, store card and money lenders). Therefore, the study finds that optimists are associated with a higher number of loans, which reflects that borrowing from informal and expensive sources is likely to add to the debt burden for optimists and lead to debt traps.

Klove and Mehlum (2018) note that optimists taking up too much debt may not be a problem if their expectations are well-founded. Subsequently, they define people who expect to be better-off but turn out to be worse-off as illusionary optimists. They show that two thirds of optimists are actually illusionary optimists, who have already borrowed too much to repay. The

authors suggests that future research could empirically explore whether illusionary optimists may end up over-indebted.

In conclusion, the above-mentioned literature has generally established a positive relationship between financial optimism and the level of borrowing. Empirical studies first study the determinants of debt and include financial optimism as a control variable (Del Rio and Young, 2006). Studies from developed (Brown *et al.*, 2008; Lee, 2008) and developing countries (Semenova, 2011) later identify the important role of financial expectations in determining the amount of secured and unsecured debts, which is in line with the life-cycle theory suggesting that optimists will borrow to finance their consumption today (Modigliani and Brumberg, 1954; Brown *et al.*, 2005). Klove and Mehlum (2018) note that much attention should be focused on the consequences of financial optimism as debt accumulation from illusionary optimism are highly problematic. Studies from Finland (Hyytinen and Putkuri, 2018) and Germany (Grohmann *et al.*, 2019) confirm that errors in financial expectations are associated with higher debt-to-income ratios and illusionary optimists may end up over-indebted.

4.2.3 Contribution

The contributions of this paper are threefold. First, the topic builds a broad platform to discover the potential effects of errors in financial expectations on borrowing decisions. To the best of my knowledge, no study has examined the relationship between errors in financial expectations and over-indebtedness in developing countries in general, and in South Africa (SA) in particular. Over the past decades, South Africa has witnessed a rapid growth of credit markets. The household debt-to-income ratio has grew by over 40% since 1990. Consumers are getting used to

borrow to consume and thus at risk of accumulating unsustainable levels of debt. Thus, the prevalence of over-indebtedness among individuals is a growing concern for policy makers in SA (Meniago *et al.*, 2013).

To examine the determinants of over-indebtedness, the majority of previous studies (Brown *et al.*, 2005; Semenova, 2011) have documented a positive relationship between financial optimism and the level of debt. Klove and Mehlum (2018) make use of SA data and suggest that well-founded optimism can help smooth consumption over the life-cycle, while illusionary optimism is problematic as people cannot keep up with the repayment of principal and interest and may end up into a debt trap. Given the high prevalence of illusionary optimism in South Africa (Klove and Mehlum, 2018), it is important to examine the consequences of errors in financial expectations on individuals' borrowing decisions and over-indebtedness.

Second, the majority of the aforementioned literature has used cross-sectional data (Del Rio and Young, 2006; Lee, 2008) and random-effects models (Brown *et al.*, 2008), which could lead to biased estimations due to unobserved individual heterogeneity. This study uses a fixed-effects estimator to control for unobserved individual fixed traits (Hyytinen and Putkuri, 2018).

Third, this study adds to the literature by examining the extent to which low education and financial illiteracy mediate the relationship between errors in financial expectations and over-borrowing. Exploring the moderating effects helps understand whether illusionary optimism and pessimism affect the borrowing behaviour of different groups of people differently, and enables policy makers to identify the most vulnerable groups.

4.3 Hypotheses

In this section, I will demonstrate the three hypotheses related to the direct and indirect link between errors in financial expectations and borrowing decisions, which are the general hypotheses, the low education hypotheses and the debt illiteracy hypotheses.

4.3.1 General hypotheses

The determinants of consumers borrowing decisions have been one of the central issue of concern to economists, and the classic life-cycle theory provides one of the first insight suggesting that individuals will maximise their intertemporal utility based on current and future income and resources (Modigliani and Brumberg, 1954). Brown *et al.* (2005) extend the model by examining the role of financial optimism, and they predict that financial optimism is positively associated with the level of debt.

Empirical studies from the UK (Del Rio and Young, 2006; Brown *et al.*, 2005, 2008) and the US (Lee, 2008) provide evidence documenting a positive (negative) relationship between financial optimism (pessimism) and the probability of debt participation and the amount of debt. Consistent evidence has been found among Russia (Semenova, 2011) and South Africa (Klove and Mehlum, 2018).

Klove and Mehlum (2018) suggest that more attention should be given to the consequences of errors in financial expectations, for instance illusionary optimists are associated with higher levels of debt and may have trouble to keep up with the repayments. This finding has been replicated in a lab experiment by Grohmann *et al.* (2019), who suggest that illusionary optimists

are more likely to be over-indebted as they are reluctant to save and possess more debts throughout the experiment. Hyytinen and Putkuri (2018) further test this theory using Finnish panel data. They find that not only individuals who expect an improvement but experience a worsening in financial situation show a higher debt-to-income ratio than individuals who make pessimistic forecast errors, but also that over-optimistic individuals are more likely to report having problems to pay-off their bills and debts. Hence, these findings motivate my general hypotheses:

H1a: Optimistic (pessimistic) forecast error is associated with a higher (lower) probability of borrowing.

H1b: Optimistic (pessimistic) forecast error is associated with a higher (lower) amount of borrowing.

H1c: Optimistic (pessimistic) forecast error is associated with a higher (lower) probability of over-indebtedness.

4.3.2 Low education hypotheses

On the one hand, empirical studies (Hyytinen and Putkuri, 2018; Grohmann *et al.*, 2019) document a positive relationship between optimistic forecast errors and borrowing. On the other hand, empirical studies (Calvet *et al.*, 2009; Lusardi and Tufano, 2015) show that people with lower education are more likely to be financially illiterate. As a result, they are more prone to make financial mistakes (loss at stock markets and taking higher shares of risky assets) and engage in high-cost borrowings without fully understanding the terms and conditions of such

loans. Taken together, I assume that overoptimistic respondents with a lower education are more likely to engage in borrowing due to lack of ability to understand debt contracts and to have a proper financial planning horizon. In the light of above evidence, the low education hypotheses are the following:

H2a: Among individuals with optimistic forecast errors, the low educated compared to the high educated, show a higher probability of borrowing

H2b: Among individuals with optimistic forecast errors, the low educated compared to the high educated, show a higher amount of borrowing

H2c: Among individuals with optimistic forecast errors, the low educated compared to the high educated, show a higher probability of over-indebtedness

4.3.3 Debt illiteracy hypotheses

Debt literacy has been more relevant in recent years than in the past as consumers are facing much more complex loan products and the potential cost of misunderstanding a loan product could be expensive. The relationship between debt illiteracy and consumers borrowing decisions has been well-documented. In particular, Lusardi and Tufano (2015) define debt literacy as the ability to apply basic knowledge about compound interests to real-life financial decisions. They find that individuals who are debt illiterate, i.e. who are unable to calculate compound interest show higher levels of debt (particularly high-cost debt). For instance, illiterate consumers are likely to pay their credit card bills late, go over the credit limit and overlook additional charges on different financial products. Disney and Gathergood (2013) find that consumers with poorer

levels of debt literacy show higher debt-to-income ratios because they are more likely acquire multiple loans in one time and have a high share of expensive loans. Taken together with previous evidence on the link between optimistic forecast errors and borrowing (Hyytinen and Putkuri, 2018), I assume that overoptimistic individuals show a high probability of borrowing and over-indebtedness due to debt illiteracy. More precisely, debt illiterate optimists who lack of the financial knowledge and ability to understand debt contracts and the associated compound interests, are more prone to engage in debt financing. On the basis of these studies, the debt illiteracy hypotheses read:

H3a: Debt illiterate optimists[21], show a higher probability of borrowing compared to their debt literate counterparts.

H3b: Debt illiterate optimists, show a higher amount of borrowing compared to their debt literate counterparts.

H3c: Debt illiterate optimists, show a higher probability of being over-indebted compared to their debt literate counterparts.

21 In here, optimists are referred to individuals with optimistic forecast errors.

4.4 Data and summary statistics

This section provides detailed information on the data used and the measurement of relevant variables. This is followed by summary statistics.

4.4.1 Data

The data is drawn from South Africa's National Income Dynamics Survey (NIDS). It is the first ever national longitudinal panel study in South Africa, initiated by the Department of Planning, Monitoring and Evaluation (DPME) and carried out by the Southern Africa Labour and Development Research Unit (SALDRU) based at the University of Cape Town's School of Economics. This survey consists of five waves (2008, 2010, 2012, 2015, 2017) with a nationally representative sample of around 28,000 individuals and 7,300 households nationwide. The survey has adopted a stratified two-stage cluster sampling design, and been repeated with a combination of individual, household, child and proxy questionnaires (Leibbrandt *et al.*, 2009). NIDS keeps track of the livelihoods of individuals and households over time. It provides comprehensive information about labour market participation and economic activity, household composition and structure, health and education, poverty and well-being, fertility and mortality, migration, human capital formation, vulnerability and social capital.

Moreover, NIDS provides the rare opportunity to examine the relationship between financial well-being and borrowing behaviour at the individual level as opposed to the household level. Klove and Mehlum (2018) suggest that individuals in developing countries are more optimistically biased than people in developed countries. People in poorer countries are also more

likely to display imprudent financial habits (e.g., no precautionary savings) and thus more vulnerable to negative shocks (financial crisis, job loss). Thus, it is interesting to explore the determinants of debt at the individual level, for a change.

This empirical analysis relies on two features of the NIDS: First, I take advantage of the biennial design of the survey and construct the errors in financial expectations variables using the mismatch of expectation and reality between two consecutive waves. Second, as opposed to most surveys focused on household debt, the structure of the NIDS allows me to make a linkage between errors in expectations and borrowing decisions at the individual level.

In the process of data cleaning, I applied the following steps: I first excluded wave 1 (2008) and wave 4 (2015) as these two waves lacked of information to construct the errors in expectation variables, and subsequently excluded wave 3 (2012) due to the lack of wealth variables. Then, I have treated non-response ('do not answer' or 'refuse to answer') as missing values. Third, based on South Africa's Basic Conditions of Employment Act (BCEA), I have excluded those who were younger than 15 as they do not have the legal right to enter labor markets. Fourth, the values of financial-related variables are inflated to the real values of March 2017 using the consumer price index from the Statistics South Africa (SSA). In addition, I have excluded the top 1% of the continuous variables used, in my analysis (i.e. income, consumption, net worth and debts) to control for the potential bias by outliers. Furthermore, Juster *et al.* (2007) suggest that sensitive questions like income and consumption have notoriously suffered from high rates of non-response. In order to secure the most complete data possible, this paper uses the imputed

variables provided by the NIDS using multiple imputation techniques to correct for missing and non-response data[22][23].

My baseline sample consists of 17,940 individuals and 23,068 observations from the 2010 and 2017 waves of the NIDS. All individuals in the same household are included in the final sample, not just the head of household. The mean number of observations per individual is 1.3, with the maximum (minimum) being 2 (1). There are 5,128 out of 17,940 individuals participated in both waves. Table 4.1 demonstrates the structure of the unbalanced panel used in this analysis. On top of the original sample members interviewed in wave 1, there are wave-on-wave replenishments to increase the sample size while maintained representativeness. The sample size increased significantly from 8,534 (wave 2) to 14,534 (wave 5) because of accumulated replenished observations from wave 3 to wave 5.

4.4.2 Measurements of dependent and independent variables

4.4.2.1 Borrowing and over-indebtedness

Probability of holding debt

In light of the increasing prevalence of debt borrowing in South Africa (South Africa Reserve Bank, 2014), I employ a wide range of debt indicators for borrowing decisions. Following Del Rio and Young (2006) and Klove and Mehlum (2018), I first make use of two debt indicator

22 Imputed variables are used for household income, household food and non-food consumption, household net worth, individual financial debt and household total debt.
23 Detailed information on how to conduct multiple imputation in the NIDS can be found in Argent (2009) and Finn *et al.* (2009).

related to the overall assessment of the debt condition of an individual: *Financial debt* and *Total debt*[24].

First, I construct a *Financial debt*, which stands for the financial debt owed by an individual. It includes personal bank loans, credit card loans, store card loans, hire purchase loans, family and friends loans, study loans from a bank, study loans from other institutions, loans from a micro-lender, loans from a 'Mashonisa'[25], loans from an employer, unpaid taxes and arrears in services and bills. Next, I construct a *Total debt*, which adds on *Financial debt* by including the vehicle debt and mortgage owed by the individual[26].

Moreover, in order to understand the mechanisms behind individuals' choices of borrowing, I also investigate different sources and types of borrowing. Due to the limited access to formal loan (in particular bank loan) in South Africa, I distinguish the types of debt between bank loans and personal loans. The focus has been drawn to personal loan as Klove and Mehlum (2018) suggest that it is more prevalence for consumers to apply for informal loan (personal loan) rather than formal loan (bank loan). *Personal debt* is the personal and informal debt owed by an individual including study loans from other institutions, store card loans, hire purchase loans, loans from family and friends, loans from a micro-lender and loan from a 'Mashonisa'.

---

24 Home ownership is very prevalent in South Africa (over 76% in our sample), thus the conventional approach to conduct research on the basis of secured debt versus unsecured debt is not suitable here. Instead, I include secured debt (e.g., mortgage) as part of the total debt to account for total indebtedness.
25'Mashonisa' stands for the unregulated/informal loan provider in South Africa (a person or a company), which mainly provides small quick cash loans for people in the local communities.
26 Other than mortgage payment (measured at the household level), every other loans used to construct the debt indicators are measured at the individual level. This suggests that *Financial debt*, *Personal debt*, *Family usage debt* and *Short-term debt* are all measured at the individual level, while *Total debt* are measured at the household level.

Furthermore, because different types of loans come with different terms and conditions, I examine how illusionary optimism/pessimism affects family usage loans and short-term loans. *Family usage debt* is the family usage loans owed by an individual, which consist of mortgages, vehicle loans, personal bank loans, study loans from a bank or an institution, store card loans and hire purchases. *Short-term debt* is the short-term loans owed by an individual, which include credit card loans, store card loans, family and friends loans and loans from a micro-lender or a 'Mashonisa'.

Finally, I examine the probability of an individual holding different types of debt. I create a dummy variable "*FD /TD /PD /FD /SD*" taking the value of 1 if an individual currently possesses "*Financial debt /Total debt /Personal debt /Family usage debt /Short-term debt* ", and 0 otherwise.

Amount of debt

In addition to exploring the probability of debt, I investigate the amount of debt owed by an individual. The continuous debt indicator "*LFD /LTD /LPD /LFD /LSD*" is constructed by taking the logarithm of the "*Financial debt /Total debt /Personal debt /Family usage debt /Short-term debt* ".

Over-indebtedness

Grohmann *et al.* (2019) suggest that over-optimistic consumers fail to recognise the consequences of their debt and may accumulate too much debt that they later regret. This is particular dangerous if they are hit by an unexpected negative shock. Over-confident consumers

may end up in an endless debt loop. Following Bryan *et al.* (2010) and Hyytinen and Putkuri (2018), I divide the amount of *Financial debt* (*Total debt*) by household income and get a financial debt-to-income (total debt-to-income) ratio. Then I create a dummy variable "*Over-indebtedness 5*" (*Over-indebtedness 3*) equal to 1 if a household's debt-to-income ratio is larger than 0.5 (0.3), and 0 otherwise. Over-indebtedness dummies are constructed for the two debt indicators: *financial debt* and *total debt*[27].

4.4.2.2 Errors in financial expectation

I use the mismatch of the following two questions to construct the errors in financial expectations variables:

Q1 Expectation (at time t-2): "On which step do you expect to be 2 years from now?"

Q2 Reality (at time t): "On which step are you today?"

Respondents are asked to answer these two questions based on a six-step scale where the highest step (6) represents the richest people in SA and the bottom step (1) represents the poorest people in SA. Formally, let E_{it-2} be the expected step of respondent *i* in time *t-2*. Let R_{it} be the current step of respondent *i* in time *t*. I use the mismatch between the expectation dated *t-2* (E_{it-2}) and reality (R_{it}) to measure optimistic and pessimistic biases. Those who expected to be in a higher step than their current step are classified as being optimistically biased. There are three degrees of optimistic bias based on how illusionary they are: '*1-step optimism*', '*2-step optimism*', '*3-step*

27 I do not create over-indebtedness dummies for *Personal debt/ Family usage debt/ Short-term debt* because the over-indebted dummies for these debts have very low percentages (ranging from 1-7 percent).

optimism'. Those who expected in a lower step than their current step are classified as being pessimistically biased ('*pessimism*'). For full details on how these categories are constructed, see Table 4.2.

4.4.2.3 Moderators

I make use of low education and debt illiteracy to investigate how these variables moderate the relationship between errors in financial expectations and borrowing. Following Hyytinen and Putkuri (2018), I define the *low education* dummy taking the value of 1 if an individual received no education or only primary education, and 0 otherwise.

As for debt illiteracy, there are two questions (*DIQ1* and *DIO2*) related to individuals' understanding of the core debt concept, namely compound interests, in the survey:

DIQ1: "Suppose you put money in the bank for two years and the bank agrees to add 15 percent per year to your account. Will the bank add more money to your account the second year than it did the first year, or will it add the same amount of money both years? - Answers: [1.1]The same; [1.2]More."

DIQ2: "Suppose you had R100 in a savings account and the bank adds 10 percent per year to the account. After five years, if you did not remove any money from the account, would you have....? - Answers: [2.1]More or less than R150; [2.2]More than R150."

I create a dummy variable *dilliterate* as a proxy for debt illiteracy questions, which equals to 1 if an individual answers "*The same"* in *DIQ1* and/or "*More than R150"* in *DIQ2*, and 0 otherwise (Lusardi and Tufano, 2015).

4.4.3 Summary statistics

Table 4.4 provides the summary statistics for all the variables used from the NIDS. On average, 38% of the sample have debt and 36% of the sample have financial debt. The proportion of personal/family usage/short-term debt is smaller at 17%/15%/14%. The average amount of total debt is R6,957 and, most notably, the average financial debt only stands at R3,010, which suggests that financial debt and non-financial debt (mortgage and vehicle finance) each contribute half to the total debt. The average amount of personal/short-term debt is R285/R301 and accounts for one tenth of the financial debt, which indicates that these debts are primarily used as transition/emergency debts because they are easier to access. As for over-borrowing, 20% (17%) of the sample report that their amount of total debt (financial debt) has exceeded 50% of their household income. 24% (22%) of respondents report their total (financial) debt-to-income ratio exceeds 30%.

Turning to errors in financial expectations indicators, I observe that over 63% of the respondents are optimistically biased, and 1-step optimists, 2-step optimists and 3-step optimists represent 30%, 21% and 12% of the sample respectively. 23.2% of the sample correctly predict their future financial situation and only 13% of the sample are pessimistically biased. This suggests that over-optimistic expectations are more prevalent among the sample than over-pessimistic expectations.

Looking at moderators, individuals with low education account for 29% of the sample. Nearlyhalf (42%) of the sample report having difficulties to fully understand compound interests.

The sample size of debt illiteracy variables is smaller than that of other variables because the debt illiteracy question only appeared in Wave 5 (2017) of the data.

As for control variables, 38% of the sample is male with an average age of 39. 33% of the sample are either married or living with a partner, while 2.5% (10%) of the sample are divorced (widowed). Over 90% of the sample have only received secondary education or less, and only 10% of the sample attended high schools or universities. Specifically, 19% and 60% of the sample attended primary and secondary schools, while 10% of the sample received no education. Average household size is 5 people per household[28]. Average household income is R7,120 and average household food and non-food consumption are R1,354 and R2,277. Household average net worth is 181,955.

In order to better examine the relationship between errors in financial expectations and the borrowing decisions, I divide the sample into five groups: *Opt1st* (1-step optimists), *Opt2st* (2-step optimists), *Opt3st* (3-step optimists), *Correct* (individuals with correct prediction) and *Pes* (pessimists). Table 4.5 presents the summary statistics of all dependent variables for these five categories. As in regressions, the *Correct* category is treated as the reference group and conclusions are drawn by comparing optimism and pessimism indicators with the *Correct* category. Thus, to gather some preliminary evidence, I use the correct prediction (*Correct*) as the reference group when performing the two sample t-tests for *Opt1st*, *Opt2st*, *Opt3st* and *Pes*.

28 I dropped the top 1% of all continuous dependent and independent variables (the amount of debt, i.e. *LTD, LFD, LPD, LFUD, LSD;* hhsize, HH income,consumption and net worth) to control for extreme outliers.

I find no significant difference when looking at the borrowing variables with regards to *Opt1st*, with one exception: The average amount of family use loans is significantly lower for 1-step optimists. Similarly, no significant difference is found in the borrowing variables between *Pes* and *Correct* (except for the average amount of personal debt). Coming to *Opt2st*, I observe the difference in the probability of holding personal/family use/short-term debt is negative between *Opt2st* and *Correct*. The differences in the probability and amount of debt are also significant and negative for *Opt3st*. The preliminary results from the optimistic and pessimistic variables generally contradict H1a-H1c. However, the two sample t-tests only provide limited information as they do not control for other co-variates and thus more sophisticated analysis is warranted.

4.5 Baseline specification and methodology

I will in turn analyze two models, to study the relationship between errors in financial expectations and borrowing behaviour. The first model will focus on the direct effects, and the second model will assess whether the relationship is moderated by low education /debt illiteracy.

4.5.1 Main specification

Following Hyytinen and Putkuri (2018) and Klove and Mehlum (2018), I investigate the relationship between errors in financial expectations and borrowing decisions by estimating the following specification:

Equation (4.1)

$$Debt_{it} = \beta_0 + \beta_1 Opt1st_{it} + \beta_2 Opt2st_{it} + \beta_3 Opt3st_{it} + \beta_4 Pes_{it}$$

$$+ \beta_5 X_{it} + v_i + v_p + v_t + \varepsilon_{it}$$

where $Debt_{it}$ represents one of the debt indicators of individual *i* at time *t*. The debt indicators are measured by either a binary variable (*TD, FD, PD, FUD, SD, TDOI5, FDOI5, TDOI3, FDOI3*) or a continuous variable (*LTD, LFD, LPD, LFUD, LSD*). The subscripts *p* and *t* index province and year respectively. *Opt1st/Opt2st/Opt3st/Pes* is 1 if an individual is a 1-step optimist/2-step optimist/3-step optimist/pessimist and 0 otherwise.

In line with H1a-H1c and with Hyytinen and Putkuri (2018) and Grohmann *et al.* (2019), the coefficients β_1, β_2, β_3 (β_4) associated with optimism (pessimism) are (is) expected to be significantly and positively (negatively) associated with individuals' borrowing decisions in

South Africa. In other words, optimistically biased individuals tend to borrow more and are at risk of accumulating unsustainable levels of debt, while pessimistically biased individuals display less debt and related problems.

X_{it} indexes demographic controls variables, which include background (age, age square, gender, rural, marital status, household size and unemployment), education (no education, primary school and secondary school) and wealth (household income, household food and non-food consumption, net worth) information. First, Del Rio and Young (2006) find an U-shaped relationship between individuals' age and borrowing decisions, which suggests that the probability of borrowing debt increases until individuals reach the mid-thirties and then decreases. Thus, to capture the life-cycle effect of age, I expect the coefficient associated with age and age square to be positive and negative respectively.

Secondly, I expect the coefficients associated with *married*, *divorce* and *widow* to be positive as married, divorced and widow people are most likely to obtain unsecured loans and mortgages due to greater needs and/or shared responsibilities, while the never-married are significantly less likely to obtain any loans (Pastrapa and Apostolopoulos, 2015).

Furthermore, I expect the coefficient associated with *urban* to be positive as urban households have easier access to a variety of loan products (Pastrapa and Apostolopoulos, 2015). Fourth, I expect the coefficient associated with *hhsize* to be positive as larger households have greater needs for consumption and these needs can be covered by debt (Swain, 2007).

Fifth, I expect the coefficients associated with *no school*, *primary* and *secondary school* to be negative as better-educated people are more confident about their future income and hence borrow more than the low-educated (Duval and Wolff, 2012). Sixth, I expect the coefficients associated with the wealth controls (*income, consumption* and *net worth*) to be positive as individuals with higher income and/or wealth are more confident regarding future financial situations and thus more likely to use debt leverage to consume (such as buying an expensive car in installment or have a larger mortgage payments). Moreover, individuals with higher consumption are also more likely use loans to cover their needs (Pastrapa and Apostolopoulos, 2015). Lastly, I expect the coefficient associated with *unemployed* to be negative as the unemployed are less likely to participate in the debt market than the employed due to facing greater uncertainty of future income (Fidrmuc *et al.*, 2013).

In equation (4.1), v_i is an individual-specific component that accounts for all time-invariant individual heterogeneity that might affect respondents' borrowing decisions. I control for this component by employing a fixed-effects estimator. $v_p$is a province-specific component that takes into account the considerable heterogeneity characterizing the South Africa territory. I control for this by including provincial dummies in all models[29]. v_t is a time-specific component that captures a time-specific or business cycle effect as well as the possible effect of changes in loan interest rates. I control for this by including year dummies in all models. ε_{it} is an idiosyncratic component. Definitions of all variables used are presented in Table 4.3.

29 There is a provincial boundary change between the 2001 Census and 2011 Census. Thus, in this analysis, the provincial dummies have been adjusted to the 2011 Census.

4.5.2 Moderating effects of low education and debt illiteracy

To further examine the moderating effects of low education and debt illiteracy on the relationship between errors in financial expectations and borrowing, I next estimate the following regression:

Equation (4.2)

$$Debt_{it} = \beta_0 + \beta_1 Error_{it} * Moderator_{it} + \beta_2 Error_{it} * (1 - Moderator_{it})$$

$$+ \beta_3 Moderator_{it} + \beta_4 X_{it} + v_i + v_p + v_t + \varepsilon_{it}$$

where $Error_{it}$ indexes the proxies of errors in financial expectations from equation (4.1), $Moderator_{it}$ is the moderator proxy of individual *i* in time *t* measured by either low education or debt illiteracy. X_{it} indexes the same vector of controls variables as in equation (4.1). The error term is also the same as in equation (4.1).

In line with the low-education hypotheses and financial illiteracy hypotheses, I expect the coefficient β_1 to be positive and larger than β_2. I estimate equation (4.2) using the fixed-effects estimator for the low education model and cross-sectional OLS for the debt illiteracy model[30].

Table 4.6 presents the correlation matrix of all variables. For optimism and pessimism variables, the correlation coefficients range from -0.145 to -0.341, which detects a weak and negative correlation between optimism and pessimism. This is only logical as individuals

---

30 Questions regarding individuals' financial literacy levels have only been introduced in Wave 5. In line with the panel estimation employed in this chapter, I treat financial literacy as a time-invariant variable and map backwards to wave 2. In the FE models (in section 4.6 and section 4.7.1), I use financial literacy as a cross-sectional variable as it will be dropped out in FE estimates. In correlated random effects model (in section 4.7.2), I map backwards and treat financial literacy as a time-invariant variable across both waves.

belonging to 1-step optimism will not be categorised into another group simultaneously. Using the rule of thumb[31], I only observe one coefficient exceeding 0.7. Expectedly, the correlation coefficient between low education and primary (secondary) education is 0.758 (-0.798), which suggests a strong correlation between education variables. Overall, I find that there are only weak correlations between the right-handed side variables, which suggests that my results are not affected by multicollinearity.

4.5.3 Estimation methodology

When estimating equation (4.1), I make use of the fixed-effects Linear probability model (LPM) for binary dependent variables, and the fixed-effects linear regression (LR) for continuous dependent variables. The main advantages of using a linear fixed-effects (FE) estimator is two-folds: First, FE estimator controls for unobserved individual heterogeneity (v_i) by removing time-invariant components in the regressions. Second, the parameter estimates can be directly interpreted as the 'mean marginal effect' of explanatory variable on the response variable. However, the linear FE model also faces problems. For instance, the fitted values obtained from the LPM are not always constrained to lie in the unit interval. In extreme cases, regressors with a large range of values will cause the LPM's predicted probabilities breach the unit interval (greater than 1 or below 0). Wooldridge (2010) addresses this issue stating that the LPM is fairly robust if the values of independent variables are near the sample means and have no extreme values.

---

31 According to Hinkle *et al.* (2003), the rule of thumb for correlation analysis is that correlation coefficient exceeding 0.7 indicates a strong correlation between variables, and thus should be treated with caution.

Another related critique is that the LPM assumes a constant marginal effect of an independent variable for all its values. However, the size of marginal effect might vary with respect to different values of a continuous independent variable. For example, household going from having zero child to one child will likely to have a higher marginal effect on household saving decision than household that goes from having one child to two (Wooldridge, 2013). However, if the independent variable is binary instead of continuous, the LPM only need to estimate one marginal effect for binary variables and thus eliminate the concern of constant marginal effect caused by continuous variables. Thus, in the empirical analysis, I take advantage of these properties of the LPM by (1) eliminating extreme values of continuous variables and (2) generating the main independent variables as binary.

4.6 Empirical results

This section first presents the regression results for the main specification. Subsequently, the estimates of equation (4.2), which includes the moderation effects are presented.

4.6.1 Main results

4.6.1.1 The probability and amount of borrowing

The probability of borrowing

Panel A of Table 4.7 presents the fixed-effects estimates for the binary borrowing models in equation (4.1). In column 1, I observe that 1-step and 2-step optimism is associated with a 3.2 and 4.7 percentage point greater probability of holding financial debt. The coefficients are significant at the 5% and 1% level respectively. This finding supports H1a. On top of financial debt, total debt includes secured debt (mortgage and car loan). I find that 2-step optimism is significantly and positively associated with the probability of holding total debt, with an estimated coefficient of 0.037. In line with H1a, in column 3, I find consistent evidence that 1-step and 2-step optimism is significantly and positively associated with the probability of holding personal debt, with estimated coefficients of 0.040 and 0.026 which are statistically significant at the 1% and 10% level respectively.

Turning to columns 4 and 5 of Panel A of Table 4.7, I find that 1-step optimism is associated with a 2.9 and 2.4 percentage point higher probability of holding family use debt and short-term debt. These coefficients are significant at the 5% level and in line with H1a. What's more, 3-step

optimism is significantly and positively associated with family use and short-term debts, with estimated coefficients of 0.028 and 0.032 which are significant at the 10 and 5 percent level respectively.

Across columns 1-5 of Table 4.7, in line with Grohmann *et al.*, (2019), I find consistent evidence that overoptimistic individuals (1-step and 2-step optimists) show a higher probability of borrowing (especially for financial debt and personal debt), which supports H1a. However I fail to find a significant relationship between 3-step optimism and the probability of borrowing (except for *FCU* and *SD*), which partially contradicts H1a.

The reasons why 3-step optimism is not a significant determinant of *FD, TD* and *PD* could stem from the characteristics associated with the intertemporal definition of 3-step optimism. 3-step optimists are those who made over-optimistic expectations regarding their future financial situation at time *t-2*, and subsequently reported that they have ended up in a much worse financial situation at time *t* than expected. Panel B of Table 4.11 reports the wealth by optimists. I observe that 3-step optimists have lower income and net worth than 1-step and 2-step optimists, and experience an income reduction at time *t* compared to time *t-2*. This confirms that 3-step optimists are in a worse financial situation at time *t* than 1-step and 2-step optimists.

In addition, according to Brown *et al.* (2005) and Klove and Mehlum (2018), which suggest that individuals with optimistic financial expectations show a higher probability and level of over-borrowing at time *t-2*. However, since individuals experienced a negative income shock in time *t*, due to their over-borrowing behaviour in time *t-2*, they may have no credit to take on more

debt in time *t*. Grohmann *et al.* (2019) conclude that individuals who over-borrowed in the initial periods (time *t-2*) will adjust their consumption and reduce their borrowing to make ends meet in subsequent periods (time *t*).

In order to examine whether 3-step optimists indeed over-borrowed in the previous period. In Appendix 4.1, I generate a variable *lagged opt3st*, which represents the 2nd lagged of *Opt3st*. I perform a two sample t-test for *lagged opt3st* and find that at time *t-2*, 3-step optimists indeed show a higher probability of debt holding and a higher amount of debt. Most importantly, I find that at time *t-2*, 3-step optimists have a higher prevalence of being over-indebted. Additionally, in Table 4.5, I find that 3-step optimists are on average show a lower probability of borrowing and over-indebtedness and lower levels of debt as compared to 1-step and 2-step optimists. Thus, in line with findings on the non-monotonicity found across 1/2/3-step optimism, I find that 1-step and 2-step optimists are associated with a higher probability of borrowing and a higher level of debt to smooth consumption as they are in better financial conditions (higher income and wealth, less income volatility). On the other hand, 3-step optimists may show a lower probability of borrowing and a lower level of debt in the current period to make up for their over-borrowing in the previous period (Grohmann *et al.*, 2019).

Moreover, I find that 3-step optimism is significantly and positively associated with *FUD* and *SD*, but not *FD*, *TD* and *PD*. As discussed above, 3-step optimists experience negative income shocks and end up in a much worse financial situation than 1-step and 2-step optimists, which means they have limited access to formal credits and only borrow to finance if they have to. As a result, the main source of financing available would be some debts with less requirements

and available at shorter notice. This is precisely what comprises the *FUD* and *SD*. They are mostly transitional small debts (e.g., loans from family and friends) with easy access and less terms and requirements. As for *FD*, *TD* and *PD*, these loans consist of formal loans (credit card loan and mortgage) and expensive loans (hire purchase), which have higher requirements regarding borrowers' current financial situations and assets (less accessible). Thus, 3-step optimists are prone to borrow *FUD* and *SD* as they could use these loans to cover some basic needs, and these debts could be carried more frequently in smaller amounts (Pastrapa and Apostolopoulos, 2015).

Focusing on pessimism, I observe no significant relationship between pessimism and the probability of borrowing in columns 1-5 of Panel A of Table 4.7, which partially contradicts H1a. This is in line with Hyytinen and Putkuri (2018), who find that there is an asymmetry between optimistic and pessimistic biases as pessimism is not associated with the probability of borrowing.

Looking at the control variables, first, I find that age has positive coefficient while age square has a negative coefficient. This indicates that the effect of age is positive when the respondent is young and turns to negative when old, which is in line with the LCH. In addition, I find that compared to the high educated, the low-educated show lower probability of borrowing. This is in line with Duval and Wolff, 2012 who suggest that the low educated typically have lower permanent income and wealth and fewer employment opportunities, and thus less likely to engage in borrowing.

As for wealth controls, I find individuals with higher income and wealth show a higher probability of borrowing. This is in line with Pastrapa and Apostolopoulos (2015) who suggest that wealthier individuals are more confident regarding future finance and thus more likely to use debt as a leverage to smooth current consumption. I also observe that individuals with larger non-food consumption show a higher probability of borrowing, which suggests that individuals may borrow to consume some large scale items (i.e. furniture). Lastly. I find that unemployed individuals show a lower prevalence of debt holding. This is in line with Fidrmuc *et al.* (2013), who suggest that the unemployed face greater uncertainty of future and thus less likely to borrow.

The amount of borrowing

Panel B of Table 4.7 reports the FE results for the amount of borrowing. Consistent with H1b, 1-step and 2-step optimists show a 28.7 and 36.9 percentage point higher amount of financial debt. These coefficients are significant at the 5% and 1% level respectively. Column 7 of Table 4.7 reports a significant and positive relationship between 2-step optimism and the amount of total debt, with an estimated coefficient of 0.273, which is significant at the 5 percent level.

Turning to column 8 of Table 4.7, I find consistent evidence that 1-step and 2-step optimism is not only associated with a higher probability of holding personal debt, but also with a higher level of personal debt. The corresponding coefficients are highly significant at the 1% and 5% level. They suggests that 1-step optimism and 2-step optimism is associated with a 25.3 and 21.6 percentage point higher amount of total debt, respectively. In support of H1b, 1-step optimism is associated with a 14.7 percentage point larger amount of short-term debt. Furthermore, in

contrast with H2b, I find that pessimism is not significantly associated with the amount of borrowing across columns 6-10 of Table 4.7.

Thus, the estimated relationships between all optimism/pessimism variables and the amount of borrowing are generally consistent with those found for the probability of borrowing. Specifically, I find robust evidence that 1-step and 2-step optimisms show a higher probability of borrowing and a higher level of borrowing, which supports H1a and H1b. This is in line with Grohmann *et al.* (2019), who suggest that overconfident participants show a higher probability of debt engagement. This could be that overconfident individuals are unable to pay off their debt complete and thus keep on borrowing. In line with Hyytinen and Putkuri (2018), in all my regressions, pessimism is not significantly associated with the probability and amount of borrowing.

4.6.1.2 The probability of over-indebtedness

Table 4.8 presents the FE estimates for the probability of over-indebtedness[32]. In line with H1c, I find that 1-step and 2-step optimists show a 2.5 and 3.4 percentage point higher probability of holding financial debt payments that exceed 50% of household income. These effects are significant at the 10% and 5% level respectively. In column 2, 2-step optimism is also significantly (at the 10% level) and positively associated with *FDOI3*, with an estimated coefficient of 0.028. Moving to columns 3, I find that 2-step optimism is significantly (at the 10% level) associated with the probability of holding total debt payments exceeding 50% of household

32 In addition, I make use of the financial/total debt-to-income ratio as a continuous variables and employ a fixed-effects quantile regressions as alternative specification. I generally find no significant relationship using debt to income ratio as a continuous variable. A brief discussion is provided in Appendix 4.2.

income. These findings support H1c, and are line with Hyytinen and Putkuri (2018) who suggest that over-optimistic individuals are more likely to have a higher debt-to-income ratio. They suffer from excessive debt burdens and report having troubles to keep up with the repayments of debt.

As observed in previous section, 3-step optimism is not significantly associated with a different level of over-indebtedness as they may have a hard time to clear up previous debts. Furthermore, in line with prior findings, I observe no significant relationship linking pessimism and over-indebtedness indicators in all equations. This partially contradicts H1c. As for the coefficients associated with the control variables, most of them have the same signs and significance as discussed in Section 4.6.1.1.

In the above findings, the results support the view that individuals who make optimistic forecast errors are more likely take on unsustainable levels of debt. Taken together with the previous findings, I find that 1-step and 2-step optimists are at risk of accumulating higher and potentially unsustainable levels of debt. Thus, when the next income shock strikes, they may be more financially vulnerable and dragged in debt traps.

4.6.2 Moderating results

4.6.2.1 Low education - results

Table 4.9 presents the moderating results of low education on the relationship between optimism and borrowing decisions. Panel A focuses on the results of the probability of borrowing. Looking at *Opt1st*Loweduc*, I find that 1-step optimists with low education are associated with higher probability of borrowing short-term debt, with an estimated coefficients of 0.031 and it is

significant at the 5% level. However, I do not find a significant difference between the *Opt1st*Loweduc* and *Opt1st*Higheduc* terms in *PD* and *FUD* models. This suggests that there is only a weak moderating of low education on the relationship between 1-step optimism and the probability of borrowing.

Looking at *Opt2st*Loweduc*, I find the interaction terms are significant at the 1% level with estimated coefficients of 0.101, 0.096, 0.056 and 0.027 for *FD*, *TD, PD* and *FUD* respectively. Meanwhile, *Opt2st*Higheduc* is insignificant and there is a significant difference between the two interaction terms associated with *Opt2st* in *FD*, *TD, PD*. This is consistent with H2a, which suggests that among optimists, those with lower education compared to those with higher education show a higher probability of borrowing.

Moreover, I find that the interaction terms *Opt3st*Loweduc* are associated with a 3.4 and 5.3 percentage point higher probability of holding family and short-term debts. On the other hand, the *Opt3st*Higheduc* term is insignificant across all models. In line with H2a, this finding suggests a positive moderating effect of low education on the relationship between optimism and the probability of borrowing.

Panel B of Table 4.9 reports the FE estimation results for the amount of borrowing. Consistent with H2b, I find that 1-step optimists with low education are associated with a 20.3 and 19.8 percentage point higher the amount of personal and short-term debts, while 1-step optimists with high education are associated with a 28.6 and 27.7 percentage point higher amount of holding financial and personal debts. There are no significant differences between

*Opt1st*Loweduc* and *Opt1st*Higheduc* in all columns, which partially contradicts H2b. Looking at *Opt2st*, I find consistent evidence that 2-step optimists with low education are associated with a 74.8, 70.5, 38.3 and 25.6 percentage point higher amount of financial, total, personal and short-term debts, which are all significant at the 1% level.

To sum up, I find low education generally has a positive moderating effect on the relationship between optimism and the probability and amount of borrowing. This is in line with Calvet *et al.* (2009) who suggest that low-educated optimists are more prone to make financial mistakes such as taking up high-cost debts (i.e. short-term debt) due to lack of financial knowledge. In columns 1 and 6 of Table 4.9, I also document that the interaction terms associated with higher education (*Opt1st*Higheduc* and *Opt2st*Higheduc*) are significant while those with lower education (*Opt1st*Loweduc* and *Opt2st*Loweduc*) are insignificant. This contradicts H2a and H2b. However, this contradictory finding is not consistent across all models, and thus should be treated with caution.

Panel C of Table 4.9 reports the estimation results for the probability of over-borrowing. Across columns 11-14, I find consistent evidence that 2-step optimists with low education show a higher probability of over-borrowing, with estimated coefficients of 0.051, 0.050, 0.043 and 0.044 for *FDOI5*, *FDOI3*, *TDOI5* and *TDOI3* respectively. This supports H2c and suggests that low-educated optimists are more likely to engage in over-borrowing.

4.6.2.2 Debt illiteracy - results

Table 4.10 presents the moderating results for debt illiteracy. In columns 1-5, I examine the moderating effects of debt illiteracy on the relationship between optimism and the probability of borrowing. In line with H3a, I find that 1-step optimists with debt illiteracy are associated with a 4.8 percentage point higher probability of holding financial debt. The relevant coefficient is significant at the 5% level. On the contrary, debt literate optimists are not significantly associated with the probability of borrowing. Consistent with Lusardi and Tufano (2015), compared to literate optimists, debt illiterate optimists are more likely to misunderstand the cost of loan products and fail to make financially prudent choices. Furthermore, I find consistent evidence that 1-step optimists with debt illiteracy are significantly and positively associated with *TD*, *PD*, *FUD* and *SD*, with estimated coefficients of 0.045, 0.030, 0.019 and 0.020 respectively.

Turning to 2-step optimists, the interaction terms *Opt2st*Illiterate* for *FD* and *TD* are significant at the 1% level with estimated coefficients of 0.036 and 0.035, while the interaction terms *Opt2st*Literate* are insignificant across columns 1-5 of Table 4.10. This is in line with H3a suggesting that debt illiteracy has a positive moderating effect on the relationship between optimism and the probability of borrowing. In line with previous findings, I find that 3-step optimists with debt illiteracy are significantly and positively associated with the probability of holding family debt. The estimated coefficient for *FUD* is 0.023, and it is significant at the 10% level.

In columns 6-10 of Table 4.10, I find supporting evidence that debt illiterate optimists are associated with higher levels of debts. Consistent with H3b, the interaction terms

*Opt1st*Illiterate* are significant at the 1% level with estimated coefficients of 0.394, 0.371 and 0.157 for *LFD*, *LTD* and *LPD* respectively. I also observe that 2-step optimists with debt illiteracy are associated with a 27.9, 28.8 and 15.0 percentage point increased in the amount of financial, total and personal debts. These coefficients are significant at the 1% and 5% levels. As for *Opt3st*, I only observe a negative relationship between *Opt3st*Literate* and the amount of total debt with an estimated coefficient of -0.285. Generally speaking, in line with previous findings of *Opt3st*, I find no other significant relationship between literate 3-step optimists and the amount of borrowing.

Turning to the probability of over-borrowing, I find that 1-step illiterate optimists show a 2.8 and 3.4 (2.5 and 3.2) percentage point higher probability of having financial (total) debt payments that exceed 50% and 30% of household income respectively. Supporting H3c, I find the interaction terms *Opt2st*Illiterate* are significant at the 10% level with coefficients of 0.020 and 0.028 (0.020 and 0.024) for *FDOI5* and *FDOI3* (*TDOI5* and *TDOI3*). This is in line with Lusardi and Tufano (2015), who show that debt illiterate optimists are more likely to engage in over-indebtedness due to limited understanding of debt contracts and interest compounding and misjudgement of individuals' debt positions.

4.7 Robustness checks

In this section, I investigate the robustness of the aforementioned empirical findings in two ways. First, I employ alternative indicators in defining the optimistic forecast errors, developed by Hyytinen and Putkuri (2018). Next, I make use of the correlated random effects by Wooldridge (2019) to control for the fitted values issues of the LPM and take into account the censorship of the continuous debt variables.

4.7.1 Alternative optimism biases indicators

Following Hyytinen and Putkuri (2018) and Klove and Mehlum (2018), I adopt alternative optimism indicators by dividing optimism into two sub-categories: prudent and imprudent optimism. I define prudent optimism as a dummy equal to 1 if an individual makes an optimistic forecast error but remains in a good financial position ($E_{it-2} > R_{it}$ and R_{it}>2), and 0 otherwise. Imprudent optimism is 1 if an individual makes an optimistic forecast error but ends up in a bad financial position ($E_{it-2} > R_{it}$ and R_{it}<=2), and 0 otherwise.

The previous measures of optimism have only considered how future expectations and the related errors affect individuals' borrowing decisions. However, the LCH suggests that individuals borrowing decisions depend on not only future income (expectation) but also current income (current financial position; Modigliani and Brumberg, 1954). Thus, this justifies why it is important to consider: (1) the effect of future income (expectation and expectation errors), (2) the effect of current income (current financial position) and (3) the net effect of (1) and (2) on individuals' borrowing decisions.

Table 4.11 presents the summary statistics and distribution for optimistic forecast errors. Panel A reports the summary statistics for prudent and imprudent optimism. I can see that 21.4% of the respondents are prudent optimists, while over 42% of the sample are imprudent optimists. This is in line with Klove and Mehlum (2018), who suggest that optimism is more prevalent among the poor[33] (imprudent optimists).

However, there is an ongoing debate about whether the perceived subjective ranking of the individuals accurately represents the actual ranks of the individuals based on reported income or wealth. Panel B of Table 4.11 shows the household income, wealth and the average changes in household income for the different optimism and pessimism categories. The average change in household income is measured by the biennial differences in household income. In Panel B, I observe an overall decreasing pattern going from pessimistic forecast errors to optimistic forecast errors. This diminishing pattern is in line with the assumptions that optimists who experience negative income shocks are most likely to be in lower income and wealth positions, while pessimists experiencing unexpected good outcomes end up with higher income and wealth.

Specifically, the average change in household income decreases as I go from columns 1 to 5 of Panel B of Table 4.11, and turns negative for 2-step, 3-step and imprudent optimists. This indicates that individuals who subjectively report making optimistic forecast errors, do experience negative income shock in real-life. Thus, the above findings suggest that individuals' subjective answers are well-based on actual changes happened in their life.

---

33 Klove and Mehlum (2018) define the poor as individuals ranking themselves among the bottom third on the income ladder.

Panel C of Table 4.11 reports the cross-tabulation for the two measures of errors in financial expectations. I can see that out of prudent optimists, 62% are 1-step optimistically biased. Around 60% of the imprudent optimists are 2-step or 3-step optimistically biased. However, around 55% of the 1-step optimists are also classified as imprudent optimists, while 27% (18%) of the 2-step (3-step) optimists are prudent optimists. This suggests that the two measures each represent different features of errors in financial expectations and thus are complementary.

Table 4.12 presents the regressions results for prudent and imprudent optimism. Panel A displays the main results by the alternative optimism indicators. Consistent with H1a, I find consistent evidence showing that prudent optimists display a higher probability of holding *FD*, *TD*, *PD*, *FUD* and *SD* with estimated coefficients of 0.059, 0.042, 0.037, 0.033 and 0.025. These are significant at the 5% level. Imprudent optimists show a 2.9 and 1.8 percentage point higher probability of holding personal and family debts, which are significant at the 5% and 10% level respectively. This is in line with Hyytinen and Putkuri (2018), who suggest that compared to those with correct prediction, individuals who experience negative shocks but remain in good financial situation (prudent optimists) are more likely to engage in various types of borrowing. Individuals who experience negative shocks and end up in worse financial situation (imprudent optimists) are more likely to engage in informal loans with less requirements, higher cost and available at short notice.

In line with H1b, prudent optimism is positively and significantly associated with the amount of borrowing, while imprudent optimism is associated with a 15.7 percentage point higher amount of personal debt. Turning to over-indebtedness, I observe that prudent optimism is

associated with a 3.2 (3.4) percentage point higher probability of holding financial debt payments that exceed 50% (30%) of household income, while a significant and positive relationship between imprudent optimism and *FDOI5* has also been documented. Thus, H1c is supported.

Panel B of Table 4.12 displays the moderating results of low education for prudent and imprudent optimism. I observe that both the *POpt*Loweduc* and *POpt*Higheduc* terms are significant and positive, and the differences between these terms are not significant. This indicates that both high and low educated prudent optimists are associated with a higher probability of borrowing. Additionally, in contrast to H2, I find that prudent optimists with high education are associated with a higher probability of borrowing (*FD*, *TD, PD* and *FUD*), a greater level of debt (*LFD*, *LTD* and *LPD*) and a higher probability of over-borrowing (*FDOI5* and *FDOI3*). This could be explained by the LCH, which suggests that individuals' decisions to borrow are based on current and future income. By definition, prudent optimists remain in good financial situation after facing negative income shocks, and higher education can lead to over-confidence regarding future income (Duval and Wolff, 2012). Thus, prudent optimists with higher education are positively associated with borrowing decisions.

In line with hypotheses H2, I find that imprudent optimists with low education are more likely to borrow, while imprudent optimists with high education have no significant association with borrowing decisions. This supports the previous discussion about the high-educated prudent optimists. High-educated prudent optimists only engage in more debts because they are currently in good financial position with good prospects. However, imprudent optimists suffer from

negative income shocks and thus only the low-educated of them will turn to borrowing due to lack of financial sensitive and knowledge.

Panel C of Table 4.12 reports the moderating results of debt illiteracy. I find that among prudent optimists, the debt illiterate and literate both show a higher probability and a higher amount of borrowing. There is no significant difference between illiterate and literate prudent optimists. This result can be explained by the fact that prudent optimists are in good financial position and thus more likely to borrow to finance their current consumption. The illiterate optimists may engage in more debts due to lack of understandings of contract and interest rates, while the literate optimists may borrow more debts to finance current consumption after carefully calculations and considerations.

However, I find that the illiterate imprudent optimists display higher probabilities of borrowing and of being over-indebted and a higher amount of borrowing, while the literate imprudent optimists are significantly and negatively associated with borrowing decisions. In support of H3, the differences between illiterate and literate imprudent optimists are significant. This implies that for imprudent optimists, the debt illiterate engage in more debts and are more likely to be over-indebted than the literate due to misunderstanding of debt contracts and compound interests (Lusardi and Tufano, 2015).

4.7.2 Correlated random effects

I further explore the robustness of my findings by employing a new set of estimation tools. As discussed in Section 4.5.3, there are two drawbacks of the fixed-effects estimates for main

analysis. First, the FE LPM estimates of binary dependent variables may breach the unit interval and cause over-fitting of the predicted dependent variables. Second, the levels of individual debt are generally treated as a censored variable because they have no negative values (Brown *et al.*, 2005, 2008). Using linear estimates of censored variables could lead to downward biased estimates. Thus, nonlinear models such as Probit and Tobit could be suitable alternatives.

However, incorporating the fixed-effects into Probit or Tobit model causes the incidental parameters problem and lead to biased estimates (Wooldridge, 2010). Wooldridge (2019) suggests that instead of the traditional fixed-effects model, a correlated random effects (CRE) approach should be used for nonlinear models such as Probit and Tobit models. Following Mundlak (1978) and Wooldridge (2019), I estimate equation (4.3) for CRE Probit models and equation (4.4) for CRE Tobit models:

Equation (4.3)

$$\Pr\left(Debt_{it} = 1\right) = \Phi(\beta_0 + \beta_1 X_{it} + \beta_2 C_i + \beta_3 \overline{X}_i + e_i + v_p + v_t + \varepsilon_{it})$$

Equation (4.4)

$$\ln\left(Debt_{it}\right) = \beta_0 + \beta_1 X_{it} + \beta_2 C_i + \beta_3 \overline{X}_i + e_i + v_p + v_t + \varepsilon_{it}$$

$$\ln\left(Debt_{it}\right) = \begin{cases} 0 & if \ \ \ln\left(Debt^*_{it}\right) \leq 0 \\ \ln\left(Debt^*_{it}\right) & if \ \ \ln\left(Debt^*_{it}\right) > 0 \end{cases}$$

where $\Phi(.)$ denotes the standard normal cumulative distribution function (cdf), X_{it} indexes the time-varying independent variables, $\overline{X}_i$ represents the mean of time-varying independent variables, C_i indexes the time-invariant independent variables, e_i is the individual-specific

component that is uncorrelated with the independent and control variables and subscripts are the same as quoted in equation (4.1). The same set of independent and control variables are utilised as in equation (4.1) and (4.2).

Table 4.13 presents the marginal effects of key independent variables for the Probit and Tobit models. Panel A of Table 4.13 reports the main results for the CRE Probit and Tobit models. Consistent with H1a, I find that 1-step optimism is associated with a 2.9, 3.9, 2.3 and 2.0 percentage point higher probability of holding financial, personal, family and short-term debts, and these are significant at the 10% level or beyond. I also observe similar evidence showing that 2-step optimism is associated with a 5.5, 4.4 and 2.8 percentage point higher probability of holding financial, total and personal debts, while 3-step optimism is associated with a 2.5 and 2.9 percentage point higher probability of holding family and short-term debts. These findings are in line with the previous findings using the FE LPM estimates, which suggests that the main results of this paper are robust to linear and non-linear models.

Columns 6-10 of Panel A of Table 4.13 reports the CRE Tobit results. I find consistent evidence that 1-step and 2-step optimists show a 21.1 and 39.0 (33.3 and 34.4) percentage point higher probability of holding financial (personal) debts, which are significant at the 10% and 1% (1%) respectively. Furthermore, I find that 2-step optimism is significantly and positively associated with a higher probability of holding financial (total) debt exceeding 50% and 30% of household income, with estimated coefficients of 0.035 and 0.030 (0.031 and 0.029). The results obtained from the CRE Probit and Tobit models are similar to those obtained from FE estimates,

which suggests that the positive relationship between optimism and borrowing are significant and fairly robust to the use of different methodologies and specifications.

Panel B of Table 4.13 presents the moderating results of low education for CRE Probit and Tobit models. I find that 1-step optimists with low education show a 4.0 and 3.8 percentage point higher probability of holding personal and family debts. 2-step optimists with low education show a higher probability of holding financial, total and personal debts. Furthermore, I find that 2-step optimists with low education consistently display greater levels of borrowing and greater probabilities of over-borrowing. On the other hand, the association between optimists with high education and the probability and amount of borrowing are generally insignificant (except for *PD* and *LPD*). This is in line with H2 and suggests that the positive moderating effect of low education dominates the effect of high education. In line with Calvet *et al.* (2009), I find that compared to the high-educated, the low-educated optimists are prone to make financial mistakes such as borrowing more debts and associated with a higher probability of being over-indebted.

Panel C of Table 4.13 presents the moderating results of debt illiteracy for CRE Probit and Tobit models. Consistent with H3, I find that 1-step and 2-step optimists with debt illiteracy show greater levels of debt and higher probabilities of holding debt and of over-indebtedness. In line with the previous findings, optimists with debt literacy are generally not significantly associated with borrowing. This is in line with Lusardi and Tufano (2015), who suggest that optimists are more likely to engage in borrowing due to lack of understanding of debt contracts and compound interests.

The above findings provide strong support of the positive relationship between optimism and borrowing as well as the positive moderating effects of low education and debt illiteracy. Thus, my previous empirical findings are found to be consistent and robust to the use of both linear and nonlinear models.

4.8 Conclusion

Using the 2010 and 2017 waves of the National Income Dynamics Survey (NIDS), this paper investigate the relationship between errors in financial expectations and individuals' borrowing decisions. To provide an in-depth analysis of the determinants of debt, I make use of three proxies for the borrowing decisions: the probability of borrowing, the amount of borrowing and the probability of being over-indebted. The main concern of this paper is the role of errors in financial expectations, and thus I employ two sets of indicators for errors in financial expectations, which are the degree of optimism and pessimism (Klove and Mehlum, 2018), and prudent and imprudent optimism (Hyytinen and Putkuri, 2018). This study aims to answer the following questions: the extent to which optimistic and pessimistic forecast errors affect individuals borrowing decisions, and whether low education or debt illiteracy play a moderating role in the relationship between errors in financial expectations and borrowing.

The empirical findings suggest that there is a clear asymmetry: optimists are associated with higher levels of debt and over-indebtedness, while pessimism does not affect borrowing decisions. More specifically, I find that 1-step and 2-step optimists show higher levels of debt and a higher probability of borrowing and over-indebtedness. This result is robust to using different types of borrowing (financial or total debts), as well as different source of borrowing (personal, family or short-term debts).

I further extend the research on the relationship between optimistic forecast errors and borrowing by examining the moderating roles of low education and debt illiteracy. I observe that

the higher the degree of optimism, the more significant the moderating effect of low education. For instance, I find that 1-step optimists with low education are not significantly associated with borrowing decisions. As I move along the degree of optimism, I find that 2-step and 3-step optimists with low education are significantly and positively associated with higher probabilities of borrowing and amounts of debts. This suggests that the higher the degree of optimism, the more influential the moderating effect of low education in determining individuals' borrowing decisions.

Furthermore, I focus explicitly on the moderating effect of debt illiteracy, and find consistent evidence that debt illiterate optimists are prone to borrow more and may accumulate overwhelming level of debts. Meanwhile, debt literate optimists are associated with lower levels of debts. These findings provide some plausible explanations for the relationship between optimism and borrowing, which suggests that optimists are more likely to borrow if they have low education and/or are debt illiterate. In these cases, they lack of understanding of the debt contracts and compound interests and miscalculate the cost of loan products, and eventually fail to take precautionary moves and swamp in debt traps (Lusardi and Tufano, 2015).

Given the thriving of credit markets and the prevalence of borrowing behaviour in South Africa, policies aimed at reducing the indebtedness levels are of particular interest to policy makers in SA. This paper provides two main policy implications: First, extreme optimists (2-step and imprudent optimists) are associated with more debts due to lack of understanding of debt contracts and judgement of debt positions. Thus, policy makers in South Africa can increase debt literacy by introducing financial education programmes. Financial education programmes can

take the form of economic classes in the schools and interactive workshops in corporate sector and communities. Next, policy makers could also consider providing financial advisers for unsophisticated households and communities to reduce the cost of misunderstanding of debt contracts and products.

There are several limitations in this work. To begin with, this study lost some observations in the original sample due to the intertemporal structure of the key variables of interest (errors in financial expectations). This could affect the representativeness of the empirical results. Second, this study does not investigate the casual relationship between errors in financial expectations and borrowing. Future research could explore the potential casual mechanisms linking errors in financial expectations and borrowing. Next, since borrowing decisions are closely related to saving and consumption decisions, future research could explore the relationship between errors in financial expectations and saving and consumption.

Table 4.1 Structure of the unbalanced panel

Year	Frequency	Percent	Cum(%)
2010	8,534	36.99	36.99
2017	14,534	63.01	100
Total	23,068	100	

Notes: This table presents the structure of the unbalanced panel used in this chapter.

Table 4.2 Errors in financial expectations categories

	Expectation at time *t-2*	Reality at time *t*	Value	Category	ABBRE	Percent(%)
Optimism	$Optimism_{it} = E_{it-2} - R_{it}$		3-5	3-step Optimism	*Opt3*	11.36
			2	2-step Optimism	*Opt2*	20.28
			1	1-step Optimism	*Opt1*	29.27
No change	$Correct_{it} = E_{it-2} - R_{it}$		0	Correct expectation	*Correct*	23.65
Pessimism	$Pessimism_{it} = E_{it-2} - R_{it}$		(-1)(-4)	Pessimism	*Pes*	15.44

Notes: Columns 1-6 presents the classification of errors in financial expectations. The final column reports the distribution of the errors categories. ABBRE stands for abbreviation. Definitions of all variables are shown in Table 4.3.

Table 4.3 Description of variables

Variable	Type	Description	Expected Direction of Coefficient:
Dependent Variables:			
Debt			
Total debt (*TD*)	Binary	=1 if total debt of the household is larger than 0 =0 otherwise	N/A
Financial debt (*FD*)	Binary	=1 if financial debt of the individual is larger than 0 =0 otherwise	N/A
Personal debt (*PD*)	Binary	=1 if personal debt of the individual is larger than 0 =0 otherwise	N/A
Family usage debt (*FUD*)	Binary	=1 if family usage debt of the household is larger than 0 =0 otherwise	N/A
Short-term debt (*SD*)	Binary	=1 if short-term debt of the individual is larger than 0 =0 otherwise	N/A
Log of Total debt (*LTD*)	Continuous	Logarithm of total debt by a household	N/A
Log of Financial debt (*LFD*)	Continuous	Logarithm of financial debt by an individual	N/A
Log of Personal debt (*LPD*)	Continuous	Logarithm of personal debt by an individual	N/A
Log of Family usage debt (*LFUD*)	Continuous	Logarithm of family usage debt by an individual	N/A
Log of Short-term debt	Continuous	Logarithm of short-term debt by an individual	N/A

(*LSD*)			
Over-indebtedness			
Total debt over-indebtedness 5 (*TDOI5*)	Binary	=1 if ratio of total debt to household income of the household is larger than 50% =0 otherwise	N/A
Total debt over-indebtedness 3 (*TDOI3*)	Binary	=1 if ratio of total debt to household income of the household is larger than 30% =0 otherwise	N/A
Financial debt over-indebtedness 5 (*FDOI5*)	Binary	=1 if ratio of financial debt to household income of the household is larger than 50% =0 otherwise	N/A
Financial debt over-indebtedness 3 (*FDOI3*)	Binary	=1 if ratio of financial debt to household income of the household is larger than 30% =0 otherwise	N/A
Independent Variables:			
Errors in financial expectations			
1-step Optimism (*Opt1st*)	Binary	=1 if an individual is 1 step optimistic biased =0 otherwise	Positive (+)
2-step Optimism (*Opt2st*)	Binary	=1 if an individual is 2 step optimistic biased =0 otherwise	Positive (+)
3-step Optimism (*Opt3st*)	Binary	=1 if an individual is 3/4/5 step optimistic biased =0 otherwise	Positive (+)
Correct expectation (*Correct*)	Binary	=1 if an individual correctly predicts future financial situation =0 otherwise	N/A
Pessimism (*Pes*)	Binary	=1 if an individual is pessimistic biased	Negative (-)

		=0 otherwise	
Prudent Optimism (*POpt*)	Binary	=1 if an individual is optimistic biased but remains in a good financial position =0 otherwise	Positive (+)
Imprudent Optimism (*IOpt*)	Binary	=1 if an individual is optimistic biased but ends up in a bad financial position =0 otherwise	Positive (+)
Moderators			
Low education (*Loweduc*)	Binary	=1 if an individual received no education or only primary education =0 otherwise	N/A
Debt illiteracy (*DIlliterate*)	Binary	=1 if an individual answers the compound interest question incorrectly =0 otherwise	N/A
Control Variables			
Age	Continuous	Individuals' age	Positive (+)
Age square	Continuous	Square of an individuals' age	Negative (-)
Male	Binary	=1 if an individual is male =0 otherwise	Positive (+)
Urban	Binary	=1 if an individual is living in urban area, and =0 otherwise	Positive (+)
Household size	Continuous	Household size	Positive (+)
Log of household income	Continuous	Logarithm of total household income in the last months	Positive (+)
Log of food consumption	Continuous	Logarithm of household food consumption in the last months	Positive (+)
Log of non-food	Continuous	Logarithm of household non-food consumption in the last months	Positive (+)

consumption			
Log of net worth	Continuous	Logarithm of the value of household net worth	Positive (+)
Married	Binary	=1 if an individual is married or cohabiting with a partner =0 otherwise	Positive (+)
Divorce	Binary	=1 if an individual is divorced =0 otherwise	Positive (+)
Widow	Binary	=1 if an individual is widowed =0 otherwise	Positive (+)
No school	Binary	=1 if an individuals receive no education =0 otherwise	Negative (-)
Primary school	Binary	=1 if an individuals' highest level of education is primary school diploma =0 otherwise	Negative (-)
Secondary school	Binary	=1 if an individuals' highest level of education is secondary school diploma =0 otherwise	Negative (-)
Unemployed	Binary	=1 if an individuals is unemployed =0 otherwise	Negative (-)
Western Cape	Binary	=1 if an individual lives in Western Cape =0 otherwise	N/A
Eastern Cape	Binary	=1 if an individual lives in Eastern Cape =0 otherwise	N/A
Northern Cape	Binary	=1 if an individual lives in Northern Cape =0 otherwise	N/A
Free State	Binary	=1 if an individual lives in Free State	N/A

		=0 otherwise	
KwaZulu-Natal	Binary	=1 if an individual lives in KwaZulu-Natal =0 otherwise	N/A
North West	Binary	=1 if an individual lives in North West =0 otherwise	N/A
Gautent	Binary	=1 if an individual lives in Gautent =0 otherwise	N/A
Mpumalanga	Binary	=1 if an individual lives in Mpumalanga =0 otherwise	N/A
Limpopo	Binary	=1 if an individual lives in Limpopo =0 otherwise	N/A
Notes: This table presents the definitions of all variables used.			

Table 4.4 Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Dependent variables					
The probability of holding debt:					
TD	22,747	0.381	0.486	0	1
FD	22,597	0.366	0.482	0	1
PD	23,044	0.177	0.382	0	1
FUD	23,045	0.159	0.365	0	1
SD	23,044	0.147	0.354	0	1
The amount of debt:					
LTD	22,747	6957	24727	0	308872
LFD	22,597	3010	8082	0	86338
LPD	23,044	285	1059	0	14716
LFUD	23,045	984	6798	0	236197
LSD	23,044	301	1213	0	20774
The probability of being over-indebted:					
TDOI5	22,590	0.202	0.402	0	1
FDOI5	22,447	0.176	0.381	0	1
TDOI3	22,590	0.248	0.432	0	1
FDOI3	22,447	0.225	0.417	0	1
Independent variables					
Errors in financial expectations:					
Opt1st	23,045	0.300	0.458	0	1
Opt2st	23,045	0.213	0.409	0	1
Opt3st	23,045	0.124	0.329	0	1
Correct	23,068	0.232	0.423	0	1
Pes	23,045	0.130	0.337	0	1
Moderators:					
Low education	23,045	0.298	0.458	0	1
Debt Illiterate	14,534	0.416	0.492	0	1
Control variables:					
Male	23,045	0.386	0.487	0	1
Age	23,045	39.574	17.346	15	110
Age square	23,045	1867.003	1600.174	225	12100
Married	23,045	0.333	0.471	0	1

Divorce	23,045	0.025	0.155	0	1
Widow	23,045	0.100	0.300	0	1
HHsize	23,045	5.330	3.348	1	22
Urban	23,045	0.500	0.500	0	1
No school	23,045	0.102	0.303	0	1
Primary	23,045	0.196	0.397	0	1
Secondary	23,045	0.599	0.490	0	1
HH Income	23,045	7120	7292	548	61256
Food exp	23,045	1354	847	200	6480
Non-food exp	23,045	2277	3505	35	35538
Net worth	23,045	181955	316887	0	2640758
Unemployed	23,045	0.134	0.341	0	1

Notes: This table presents the summary statistics of relevant variables. *TD*/*FD*/*PD*/*FUD*/*SD* stands for the probability of holding total debt/financial debt/personal debt/family usage debt/short-term debt. *LTD*/*LFD*/*LPD*/*LFUD*/*LSD* denotes the amount of total debt/financial debt/personal debt/family usage debt/short-term debt. *TDOI5*/*TDOI3* stands for the probability of holding total debt payments that exceed 50%/30%. *FDOI5*/*FDOI3* indexes the probability of holding financial debt payments that exceed 50%/30%. The number of observation for *debt illiterate* is lower because the debt illiteracy question is only available in the 2017 wave of the NIDS. The amount of debt (*LTD/LFD/LPD/LFUD/LSD*), *HH income, consumption* and *net worth* are in log form in all the regressions, and inflated into the real value of March 2017 using the CPI from the SSA. The top 1% of *LTD, LFD, LPD, LFUD, LSD, HH income, consumption, net worth* and *hhsize* are also dropped to avoid biased estimates due to extreme outliers. Definitions of all variables are shown in Table 4.3.

Table 4.5 Comparison between Optimism and Pessimism groups

Variable	1-step optimism *(Opt1st)*		2-step optimism *(Opt2st)*		3-step optimism *(Opt3st)*		*Correct*	Pessimism*(Pes)*	
	Mean	Diff1	Mean	Diff2	Mean	Diff3	Mean	Mean	Diff4
TD	0.392	0.004	0.374	-0.014	0.348	-0.040***	0.388	0.382	-0.006
FD	0.380	0.008	0.360	-0.012	0.332	-0.040***	0.372	0.363	-0.009
PD	0.183	0.002	0.165	-0.016**	0.165	-0.016*	0.181	0.187	0.006
FUD	0.164	0.001	0.148	-0.015**	0.153	-0.010	0.163	0.160	-0.003
SD	0.152	0.001	0.134	-0.017**	0.142	-0.008	0.151	0.152	0.002
LTD	6861.579	-393.938	6965.457	-290.061	6197.754	-1057.763*	7255.517	7303.389	47.871
LFD	3062.963	-44.175	3006.521	-100.617	2764.077	-343.061*	3107.138	2936.661	-170.477
LPD	299.049	31.150	279.575	11.677	262.002	-5.897	267.899	308.346	40.447*
LFUD	882.678	-220.622*	1150.287	46.986	764.040	-339.261**	1103.300	936.695	-166.606
LSD	313.779	18.513	292.155	-3.111	271.711	-23.555	295.266	324.918	29.652
TDOI5	0.203	-0.002	0.205	-0.001	0.190	-0.015*	0.205	0.201	-0.005
FDOI5	0.179	0.001	0.178	-0.001	0.163	-0.015*	0.178	0.171	-0.007
TDOI3	0.252	0.001	0.252	0.001	0.232	-0.019*	0.251	0.242	-0.009
FDOI3	0.231	0.004	0.229	0.003	0.208	-0.018*	0.226	0.215	-0.011
N=	6,922		4,910		2,849		5,371	3,016	

Note: Diff1/Diff2/Diff3/Diff4 reports differences in the equality of means between those who are *Opt1st/Opt2st/Opt3st/Pes* and those who made correct prediction. Statistical significance is indicated by * at the 10%, ** at the 5%, *** at the 1% levels. *TD*/*FD*/*PD*/*FUD*/*SD* stands for the probability of holding total debt/financial debt/personal debt/family usage debt/short-term debt. *LTD*/*LFD*/*LPD*/*LFUD*/*LSD* denotes the amount of total debt/financial debt/personal debt/family usage debt/short-term debt. *TDOI5*/*TDOI3* stands for the probability of holding total debt payments that exceed 50%/30%. *FDOI5*/*FDOI3* indexes the probability of holding financial debt payments that exceed 50%/30%. Definitions of all variables are shown in Table 4.3.

Table 4.6 Correlation matrix

	Opt1st	*Opt2st*	*Opt3st*	*Pes*	*Loweduc*	*DIlliterate*	*Male*	*Age*	*Married*	*Divorce*	*Widow*
Opt1st	1.000										
Opt2st	-0.341	1.000									
Opt3st	-0.246	-0.195	1.000								
Pes	-0.254	-0.201	-0.145	1.000							
Loweduc	0.012	-0.005	-0.052	0.006	1.000						
DIlliterate	-0.008	-0.021	-0.018	0.044	-0.105	1.0000					
Male	0.009	-0.010	-0.001	-0.002	-0.074	0.032	1.000				
Age	0.012	-0.017	-0.082	0.019	0.559	-0.056	-0.117	1.000			
Married	0.009	-0.013	-0.033	0.007	0.156	0.022	0.041	0.316	1.000		
Divorce	0.005	-0.009	0.000	0.004	0.013	0.003	-0.025	0.103	-0.112	1.000	
Widow	-0.003	-0.002	-0.036	0.009	0.296	-0.051	-0.190	0.469	-0.236	-0.053	1.000
HHsize	-0.002	0.016	0.018	-0.015	0.044	-0.036	-0.099	-0.096	-0.054	-0.084	-0.039
Urban	-0.011	-0.009	0.024	-0.001	-0.169	-0.008	0.031	-0.043	0.053	0.051	-0.061
No school	0.013	0.008	-0.038	0.003	0.516	-0.082	-0.066	0.432	0.064	-0.004	0.269
Primary	0.004	-0.011	-0.031	0.004	**0.758**	-0.061	-0.036	0.315	0.131	0.018	0.136
Secondary	-0.010	-0.002	0.049	-0.002	**-0.798**	0.060	0.079	-0.481	-0.152	-0.031	-0.233
HH Income	-0.003	-0.013	-0.028	0.022	-0.176	0.056	0.042	0.003	0.118	0.007	-0.056
Food exp	0.006	-0.010	-0.023	0.012	-0.135	0.022	-0.007	-0.002	0.117	-0.004	-0.048
Non-food exp	-0.012	-0.018	-0.030	0.036	-0.181	0.065	0.026	-0.016	0.127	0.028	-0.056
Net worth	-0.010	-0.010	-0.016	0.028	-0.148	0.047	0.036	0.067	0.112	0.027	-0.015

Unemployed	0.004	0.017	0.025	-0.015	-0.116	-0.004	-0.016	-0.184	-0.075	-0.021	-0.095

Notes: This table presents the correlation matrix of all independent variables. The correlation coefficient is on bold if the coefficient is larger than 0.7. Definitions of all variables are shown in Table 4.3.

Table 4.6 Correlation matrix (continued)

	HHsize	*Urban*	*No school*	*Primary*	*Secondary*	*HH Income*	*Food*	*Non-food*	*Net worth*	*Unemployed*
HHsize	1.000									
Urban	-0.212	1.000								
No school	0.039	-0.176	1.000							
Primary	0.021	-0.061	-0.167	1.000						
Secondary	0.018	0.094	-0.412	-0.605	1.000					
HH Income	0.134	0.170	-0.104	-0.124	-0.025	1.000				
Food exp	0.243	0.108	-0.088	-0.088	0.027	0.495	1.000			
Non-food exp	-0.003	0.204	-0.112	-0.123	0.000	0.542	0.451	1.000		
Net worth	-0.017	0.119	-0.088	-0.103	0.003	0.460	0.337	0.426	1.000	
Unemployed	0.071	-0.007	-0.089	-0.065	0.114	-0.077	-0.026	-0.045	-0.038	1.000

Notes: This table presents the correlation matrix of all independent variables. The correlation coefficient is on bold if the coefficient is larger than 0.7. Definitions of all variables are shown in Table 4.3.

Table 4.7 FE results for the probability and amount of borrowing

	Panel A: the probability of borrowing					Panel B: the amount of borrowing				
Variables	*FD*	*TD*	*PD*	*FUD*	*SD*	*LFD*	*LTD*	*LPD*	*LFUD*	*LSD*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Opt1st	0.032**	0.022	0.040***	0.029**	0.024**	0.287**	0.167	0.253***	0.145	0.147*
	(0.015)	(0.015)	(0.013)	(0.012)	(0.012)	(0.123)	(0.128)	(0.088)	(0.093)	(0.083)
Opt2st	0.047***	0.037**	0.026*	0.015	0.008	0.369***	0.273**	0.216**	0.078	0.062
	(0.017)	(0.017)	(0.015)	(0.013)	(0.013)	(0.136)	(0.139)	(0.096)	(0.101)	(0.092)
Opt3st	-0.017	-0.024	0.023	0.028*	0.032**	-0.118	-0.204	0.085	0.044	0.08
	(0.02)	(0.02)	(0.018)	(0.015)	(0.016)	(0.165)	(0.167)	(0.112)	(0.117)	(0.105)
Pes	0.009	0.007	0.019	0.017	0.016	0.064	0.031	0.11	0.075	0.118
	(0.02)	(0.02)	(0.017)	(0.015)	(0.015)	(0.164)	(0.166)	(0.112)	(0.115)	(0.105)
Age	0.015	0.014	0.030**	0.029**	0.027**	0.158	0.12	0.190**	0.236**	0.192**
	(0.016)	(0.016)	(0.014)	(0.012)	(0.012)	(0.132)	(0.134)	(0.09)	(0.094)	(0.085)
Age square	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.006	0.009	0.002	0.016	-0.008	0.124	0.183	0.147	0.203	0.051
	(0.024)	(0.024)	(0.022)	(0.019)	(0.019)	(0.195)	(0.196)	(0.141)	(0.141)	(0.133)
Divorce	0.058	0.069	0.026	0.078*	0.038	0.529	0.758*	0.297	0.429	0.296

	(0.052)	(0.049)	(0.046)	(0.042)	(0.041)	(0.428)	(0.404)	(0.319)	(0.332)	(0.309)
Widow	-0.021	-0.026	0.04	0.011	0.017	-0.041	-0.063	0.193	0.073	0.071
	(0.035)	(0.035)	(0.03)	(0.023)	(0.024)	(0.278)	(0.284)	(0.187)	(0.18)	(0.163)
HHsize	0.00611**	0.005	-0.007***	-0.006***	-0.002	0.046**	0.029	-0.048***	-0.034**	-0.029**
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.022)	(0.022)	(0.015)	(0.015)	(0.014)
Urban	0.007	0.022	-0.032	-0.006	-0.039	0.053	0.184	-0.115	-0.06	-0.232
	(0.031)	(0.031)	(0.026)	(0.024)	(0.024)	(0.258)	(0.268)	(0.174)	(0.188)	(0.173)
No school	-0.212**	-0.195**	-0.056	-0.054	-0.025	-1.717**	-1.584**	-0.124	-0.037	-0.124
	(0.098)	(0.099)	(0.071)	(0.05)	(0.052)	(0.772)	(0.788)	(0.402)	(0.386)	(0.327)
Primary	-0.107*	-0.109**	-0.072	-0.06	-0.026	-0.936**	-0.953**	-0.278	-0.248	-0.158
	(0.055)	(0.054)	(0.047)	(0.044)	(0.043)	(0.44)	(0.451)	(0.305)	(0.342)	(0.297)
Secondary	-0.044	-0.042	-0.070**	-0.085***	-0.060**	-0.468*	-0.483*	-0.437**	-0.404*	-0.363*
	(0.033)	(0.032)	(0.029)	(0.028)	(0.029)	(0.266)	(0.269)	(0.199)	(0.224)	(0.212)
HH Income	0.076***	0.078***	0.022**	0.028***	0.014*	0.686***	0.738***	0.131**	0.186***	0.154***
	(0.01)	(0.01)	(0.009)	(0.008)	(0.008)	(0.084)	(0.085)	(0.057)	(0.058)	(0.054)
Food exp	0.007	0.014	0.002	0.016	0.007	0.085	0.168	-0.014	0.071	-0.009
	(0.013)	(0.013)	(0.011)	(0.01)	(0.01)	(0.106)	(0.109)	(0.071)	(0.075)	(0.07)
Non-food exp	0.046***	0.049***	0.035***	0.030***	0.029***	0.424***	0.490***	0.176***	0.219***	0.174***
	(0.007)	(0.007)	(0.006)	(0.005)	(0.005)	(0.058)	(0.059)	(0.039)	(0.041)	(0.038)
Net worth	0.011***	0.012***	0.001	-0.003**	-0.001	0.076***	0.072***	0.005	-0.023*	-0.006

	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.017)	(0.018)	(0.013)	(0.013)	(0.012)
Unemployed	-0.029*	-0.029*	-0.040***	-0.030**	-0.039***	-0.291**	-0.307**	-0.254**	-0.225**	-0.219**
	(0.018)	(0.018)	(0.015)	(0.013)	(0.013)	(0.139)	(0.141)	(0.099)	(0.096)	(0.093)
N=	22,626	22,776	23,067	23,068	23,067	22626	22776	23067	23068	23067
R-squared	0.282	0.285	0.157	0.173	0.143	0.279	0.285	0.114	0.117	0.105
Number of ID	17,659	17,729	17,939	17,940	17,939	17659	17729	17939	17940	17939

Notes: This table presents the FE estimates of the probability and the amount of borrowing. Provincial and Year dummies are included but not reported. Coefficients are presented, and robust standard errors are given in parentheses. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. *TD*/*FD*/*PD*/*FUD*/*SD* stands for the probability of holding total debt/financial debt/personal debt/family usage debt/short-term debt. *LTD*/*LFD*/*LPD*/*LFUD*/*LSD* denotes the amount of total debt/financial debt/personal debt/family usage debt/short-term debt. Definitions of all variables are shown in Table 4.3.

Table 4.8 FE results for the probability of over-indebtedness

Variables	*FDOI5*	*FDOI3*	*TDOI5*	*TDOI3*
	(1)	(2)	(3)	(4)
Opt1st	0.025*	0.020	0.015	0.008
	(0.013)	(0.014)	(0.013)	(0.014)
Opt2st	0.034**	0.028*	0.025*	0.022
	(0.014)	(0.015)	(0.015)	(0.016)
Opt3st	0.008	-0.004	0.000	-0.011
	(0.017)	(0.018)	(0.017)	(0.018)
Pes	0.012	0.001	0.008	-0.002
	(0.017)	(0.018)	(0.017)	(0.018)
Age	0.007	0.018	0.003	0.016
	(0.013)	(0.015)	(0.014)	(0.015)
Age square	-0.0001***	-0.0001***	-0.0001***	-0.0001***
	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.029	0.042*	0.044**	0.055**
	(0.021)	(0.022)	(0.021)	(0.022)
Divorce	0.024	0.079	0.077*	0.109**
	(0.047)	(0.050)	(0.044)	(0.048)
Widow	0.042	0.050	0.049*	0.052
	(0.028)	(0.031)	(0.029)	(0.032)
hhsize	-0.001	0.003	0.000	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Urban	0.006	0.005	0.041	0.031
	(0.028)	(0.029)	(0.028)	(0.029)
No school	-0.111	-0.092	-0.126	-0.107
	(0.071)	(0.080)	(0.079)	(0.086)
Primary	-0.091**	-0.062	-0.097**	-0.070

	(0.046)	(0.050)	(0.048)	(0.051)
Secondary	-0.039	-0.017	-0.045	-0.027
	(0.029)	(0.030)	(0.030)	(0.03)
Food exp	0.026**	0.025**	0.029***	0.027**
	(0.011)	(0.011)	(0.011)	(0.012)
Non-food exp	0.034***	0.041***	0.042***	0.047***
	(0.006)	(0.006)	(0.006)	(0.006)
Net worth	0.007***	0.007***	0.005***	0.006***
	(0.001)	(0.001)	(0.002)	(0.002)
Unemployed	-0.031**	-0.039**	-0.034**	-0.040**
	(0.015)	(0.016)	(0.015)	(0.016)
N=	22,476	22,476	22,619	22,619
R-squared	0.106	0.144	0.121	0.155
Number of ID	17,588	17,588	17,633	17,633

Notes: This table presents the FE estimates of the probability of over-indebtedness. Provincial and Year dummies are included but not reported. Coefficients are presented, and robust standard errors are given in parentheses. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. *TDOI5*/*TDOI3* stands for the probability of holding total debt payments that exceed 50%/30%. *FDOI5*/*FDOI3* indexes the probability of holding financial debt payments that exceed 50%/30%. Definitions of all variables are shown in Table 4.3.

Table 4.9 Moderating results for low education

Panel A: the probability of borrowing					
Variables	*FD*	*TD*	*PD*	*FUD*	*SD*
	(1)	(2)	(3)	(4)	(5)
Opt1st Loweduc*	0.029	0.015	0.041**	0.033**	0.031**
	(0.022)	(0.023)	(0.018)	(0.014)	(0.014)
Opt1st Higheduc*	0.032*	0.025	0.038**	0.025*	0.019
	(0.019)	(0.019)	(0.017)	(0.016)	(0.016)
Diff1:	0.900	0.723	0.886	0.688	0.507
Opt2st Loweduc*	0.101***	0.096***	0.056***	0.027*	0.025
	(0.025)	(0.025)	(0.021)	(0.014)	(0.017)
Opt2st Higheduc*	0.019	0.007	0.010	0.008	-0.001
	(0.021)	(0.021)	(0.018)	(0.017)	(0.017)
Diff2:	0.008	0.004	0.072	0.348	0.231
Opt3st Loweduc*	-0.028	-0.043	0.046	0.034*	0.053**
	(0.034)	(0.034)	(0.029)	(0.019)	(0.022)
Opt3st Higheduc*	-0.016	-0.019	0.013	0.024	0.023
	(0.024)	(0.024)	(0.022)	(0.019)	(0.020)
Diff3:	0.768	0.541	0.346	0.669	0.267
N=	22,626	22,776	23,067	23,068	23,067
R-squared	0.283	0.286	0.158	0.173	0.143
Number of ID	17,659	17,729	17,939	17,940	17,939
Panel B: the amount of borrowing					
Variables	*LFD*	*LTD*	*LPD*	*LFUD*	*LSD*
	(6)	(7)	(8)	(9)	(10)
Opt1st Loweduc*	0.271	0.114	0.203*	0.152	0.198**
	(0.175)	(0.182)	(0.108)	(0.101)	(0.090)
Opt1st Higheduc*	0.286*	0.188	0.277**	0.139	0.114
	(0.156)	(0.160)	(0.115)	(0.125)	(0.112)
Diff1:	0.946	0.744	0.617	0.931	0.531
*Opt2st**	0.748***	0.705***	0.383***	0.171	0.256***

	(0.192)	(0.199)	(0.118)	(0.104)	(0.095)
Opt2st Higheduc*	0.168	0.053	0.135	0.031	-0.037
	(0.173)	(0.175)	(0.125)	(0.135)	(0.124)
Diff2:	0.018	0.009	0.125	0.978	0.045
Opt3st Loweduc*	-0.144	-0.327	0.049	0.110	0.129
	(0.275)	(0.279)	(0.153)	(0.132)	(0.119)
Opt3st Higheduc*	-0.129	-0.178	0.092	0.014	0.047
	(0.197)	(0.200)	(0.141)	(0.150)	(0.136)
Diff3:	0.964	0.654	0.827	0.612	0.631
N=	22,626	22,776	23,067	23,068	23,067
R-squared	0.280	0.286	0.114	0.118	0.105
Number of ID	17,659	17,729	17,939	17,940	17,939

Panel C: the probability of over-indebtedness

Variables	*FDOI5*	*FDOI3*	*TDOI5*	*TDOI3*
	(11)	(12)	(13)	(14)
Opt1st Loweduc*	0.029	0.025	0.013	0.007
	(0.018)	(0.019)	(0.019)	(0.020)
Opt1st Higheduc*	0.023	0.016	0.017	0.009
	(0.017)	(0.018)	(0.017)	(0.018)
Diff1:	0.788	0.732	0.870	0.954
Opt2st Loweduc*	0.051***	0.050**	0.043**	0.044**
	(0.019)	(0.021)	(0.020)	(0.021)
Opt2st Higheduc*	0.025	0.016	0.015	0.011
	(0.019)	(0.020)	(0.019)	(0.020)
Diff2:	0.293	0.203	0.266	0.229
Opt3st Loweduc*	-0.027	-0.036	-0.039	-0.048
	(0.028)	(0.031)	(0.029)	(0.031)
Opt3st Higheduc*	0.021	0.007	0.014	0.003
	(0.020)	(0.022)	(0.021)	(0.022)
Diff3:	0.156	0.232	0.123	0.170
N=	22,476	22,476	22,619	22,619
R-squared	0.107	0.144	0.122	0.156

Number of ID	17,588	17,588	17,633	17,633

Notes: This table presents the FE estimates of the moderating effect of low education. Coefficients are presented, and robust standard errors are given in parentheses. *Diff1/2/3* is the Wald test performed to test whether *Opt1st/2st/3st*Loweduc* is significant different from *Opt1st/2st/3st*Higheduc*. P-value for the Wald-test is presented. Other control variables include all regressors defined and listed in Table 4.3, and the estimated coefficients are presented in Appendix 4.3. Provincial and Year dummies are included but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. *TD*/*FD*/*PD*/*FUD*/*SD* stands for the probability of holding total debt/financial debt/personal debt/family usage debt/short-term debt. *LTD*/*LFD*/*LPD*/*LFUD*/*LSD* denotes the amount of total debt/financial debt/personal debt/family usage debt/short-term debt. *TDOI5*/*TDOI3* stands for the probability of holding total debt payments that exceed 50%/30%. *FDOI5*/*FDOI3* indexes the probability of holding financial debt payments that exceed 50%/30%. Definitions of all variables are shown in Table 4.3.

Table 4.10 Moderating results for debt illiteracy

Variables	*FD*	*TD*	*PD*	*FUD*	*SD*	*LFD*	*LTD*	*LPD*	*LFUD*	*LSD*	*FDOI5*	*FDOI3*	*TDOI5*	*TDOI3*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
*Opt1st**	0.048***	0.045***	0.030***	0.019**	0.020**	0.394***	0.371***	0.157**	0.078	0.091	0.028***	0.034***	0.025**	0.032***
Illiterate	(0.012)	(0.012)	(0.010)	(0.009)	(0.009)	(0.095)	(0.096)	(0.066)	(0.070)	(0.062)	(0.011)	(0.011)	(0.011)	(0.011)
*Opt1st**	-0.015	-0.013	-0.009	0.004	-0.004	-0.116	-0.085	0.061	0.033	0.036	-0.014	-0.007	-0.009	-0.005
Literate	(0.013)	(0.013)	(0.011)	(0.011)	(0.010)	(0.106)	(0.108)	(0.074)	(0.080)	(0.071)	(0.012)	(0.013)	(0.012)	(0.013)
Diff1:	0.000	0.000	0.001	0.149	0.027	0.000	0.000	0.223	0.593	0.467	0.000	0.001	0.004	0.005
*Opt2st**	0.036***	0.035***	0.016	0.011	0.002	0.279***	0.288***	0.150**	0.046	0.068	0.020*	0.028**	0.020*	0.024*
Illiterate	(0.014)	(0.013)	(0.012)	(0.010)	(0.010)	(0.108)	(0.109)	(0.074)	(0.077)	(0.069)	(0.012)	(0.013)	(0.012)	(0.013)
*Opt2st**	-0.011	-0.013	-0.021	-0.006	-0.008	0.011	-0.021	-0.011	-0.035	0.031	0.022	0.020	0.021	0.022
Literate	(0.015)	(0.015)	(0.013)	(0.012)	(0.012)	(0.125)	(0.125)	(0.086)	(0.093)	(0.084)	(0.014)	(0.015)	(0.014)	(0.015)
Diff2:	0.005	0.003	0.13	0.207	0.458	0.054	0.026	0.092	0.424	0.690	0.922	0.608	0.945	0.877
*Opt3st**	0.005	0.006	0.005	0.023*	0.014	0.056	0.100	0.048	0.095	0.085	0.024	0.015	0.024	0.016
Illiterate	(0.017)	(0.016)	(0.014)	(0.013)	(0.013)	(0.133)	(0.134)	(0.090)	(0.097)	(0.087)	(0.015)	(0.016)	(0.015)	(0.016)
*Opt3st**	-0.025	-0.033	-0.004	0.010	0.009	-0.180	-0.285*	0.087	0.099	0.077	-0.023	-0.016	-0.029*	-0.029
Literate	(0.020)	(0.020)	(0.017)	(0.016)	(0.016)	(0.160)	(0.162)	(0.111)	(0.121)	(0.107)	(0.017)	(0.018)	(0.017)	(0.018)
Diff3:	0.207	0.092	0.669	0.447	0.822	0.207	0.041	0.761	0.978	0.945	0.021	0.142	0.008	0.037

N=	14,147	14,331	14,533	14,534	14,533	14,147	14,331	14,533	14,534	14,533	14,046	14,046	14,235	14,235
R-squared	0.171	0.192	0.106	0.203	0.145	0.216	0.264	0.093	0.147	0.120	0.086	0.099	0.136	0.140

Notes: This table presents the FE estimates of the moderating effect of debt illiteracy. Coefficients are presented, and robust standard errors are given in parentheses. *Diff1/2/3* is the Wald test performed to test whether *Opt1st/2st/3st*Illiterate* is significant different from *Opt1st/2st/3st*Literate*. P-value for the Wald-test is presented. Provincial and Year dummies are included but not reported. Other control variables include all regressors defined and listed in Table 4.3, and the estimated coefficients are presented in Appendix 4.4. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. *TD*/*FD*/*PD*/*FUD*/*SD* stands for the probability of holding total debt/financial debt/personal debt/family usage debt/short-term debt. *LTD*/*LFD*/*LPD*/*LFUD*/*LSD* denotes the amount of total debt/financial debt/personal debt/family usage debt/short-term debt. *TDOI5*/*TDOI3* stands for the probability of holding total debt payments that exceed 50%/30%. *FDOI5*/*FDOI3* indexes the probability of holding financial debt payments that exceed 50%/30%. Definitions of all variables are shown in Table 4.3.

Table 4.11 Summary statistics and distribution for optimistic forecast errors

Panel A: Summary statistics for prudent and imprudent optimism					
Variables	Obs	Mean	Std. Dev.	Min	Max
POpt	23,068	0.214	0.410	0	1
IOpt	23,068	0.423	0.494	0	1

Panel B: Wealth by optimism and pessimism indicators					
Variables	*Pes*	*Correct*	*Opt1st*	*Opt2st*	*Opt3st*
	(1)	(2)	(3)	(4)	(5)
HH Income	7,549	7,382	7,089	6,937	6,579
Net worth	204,097	187,771	177,159	175886	168,444
Change in income	658	579	324	-120	-600
N=	3,016	5,371	6,922	4,910	2,849
Variables	*POpt*	*IOpt*			
	(6)	(7)			
HH Income	9,996	5,395			
Net worth	287,531	118,238			
Change in income	311	-162			
N=	4,926	9,755			

Panel C: Cross-tabulation between different measures of forecast errors					
Variables	*Pes*	*Correct*	*POpt*	*IOpt*	Total
Pes	3,016	0	0	0	3,016
	13.07%				13.06%
Correct	0	5,371	0	0	5,371
		23.05%			23.05%
Opt1st	0	0	3,074	3,848	6,922
			13.33%	16.68%	30.01%
Opt2st	0	0	1,341	3,569	4,910
			5.81%	15.47%	21.28%
Opt3st	0	0	511	2,338	2,849
			2.22%	10.14%	12.35%

Total	3,016	5,371	4,926	9,755	23,068
	13.07%	23.05%	21.35%	42.23%	100%

Notes: This table presents the summary statistics and distribution for optimistic forecast errors indicators. Definitions of all variables are shown in Table 4.3.

Table 4.12 Results for alternative optimism measures

Panel A: Main results														
Variables	*FD*	*TD*	*PD*	*FUD*	*SD*	*LFD*	*LTD*	*LPD*	*LFUD*	*LSD*	*FDOI5*	*FDOI3*	*TDOI5*	*TDOI3*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
POpt	0.059***	0.042**	0.037**	0.033**	0.025*	0.490***	0.315**	0.306***	0.199*	0.169	0.032**	0.034**	0.024	0.026
	(0.018)	(0.018)	(0.016)	(0.014)	(0.014)	(0.145)	(0.150)	(0.106)	(0.116)	(0.104)	(0.015)	(0.017)	(0.016)	(0.017)
IOpt	0.011	0.006	0.029**	0.018*	0.017	0.106	0.038	0.157*	0.053	0.073	0.021*	0.009	0.011	0.001
	(0.015)	(0.015)	(0.013)	(0.011)	(0.011)	(0.118)	(0.120)	(0.081)	(0.084)	(0.076)	(0.012)	(0.013)	(0.012)	(0.013)
Pes	0.007	0.006	0.019	0.017	0.016	0.047	0.020	0.105	0.072	0.117	0.012	0.000	0.008	-0.003
	(0.020)	(0.020)	(0.017)	(0.015)	(0.015)	(0.164)	(0.166)	(0.112)	(0.115)	(0.105)	(0.017)	(0.018)	(0.017)	(0.018)
N=	22,626	22,776	23,067	23,068	23,067	22,626	22,776	23,067	23,068	23,067	22,476	22,476	22,619	22,619
R-squared	0.282	0.284	0.157	0.173	0.142	0.279	0.284	0.114	0.118	0.105	0.106	0.144	0.121	0.155
Panel B: Moderating results - low education														
Variables	*FD*	*TD*	*PD*	*FUD*	*SD*	*LFD*	*LTD*	*LPD*	*LFUD*	*LSD*	*FDOI5*	*FDOI3*	*TDOI5*	*TDOI3*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
*POpt**	0.074**	0.052*	0.031	0.014	0.021	0.482**	0.249	0.179	0.015	0.083	0.022	0.022	0.006	0.008
Loweduc	(0.031)	(0.031)	(0.025)	(0.021)	(0.021)	(0.246)	(0.256)	(0.164)	(0.168)	(0.141)	(0.025)	(0.028)	(0.026)	(0.028)
*POpt**	0.050**	0.035*	0.036*	0.036**	0.024	0.456***	0.302*	0.327***	0.230	0.167	0.033*	0.036*	0.029	0.030
Higheduc	(0.021)	(0.021)	(0.019)	(0.017)	(0.017)	(0.171)	(0.177)	(0.126)	(0.140)	(0.127)	(0.018)	(0.020)	(0.019)	(0.020)

*IOpt**	0.034*	0.027	0.051***	0.035***	0.035***	0.322**	0.230	0.249***	0.181**	0.234***	0.028*	0.024	0.017	0.011
Loweduc	(0.021)	(0.021)	(0.017)	(0.012)	(0.013)	(0.161)	(0.166)	(0.096)	(0.086)	(0.078)	(0.016)	(0.018)	(0.017)	(0.018)
*IOpt**	-0.005	-0.008	0.016	0.008	0.005	-0.039	-0.090	0.097	-0.031	-0.031	0.016	0.000	0.007	-0.006
Higheduc	(0.019)	(0.018)	(0.016)	(0.015)	(0.015)	(0.151)	(0.152)	(0.108)	(0.116)	(0.105)	(0.016)	(0.017)	(0.016)	(0.017)
Diff1:	0.495	0.644	0.873	0.398	0.929	0.929	0.858	0.459	0.306	0.643	0.695	0.664	0.462	0.513
Diff2:	0.130	0.182	0.093	0.102	0.092	0.077	0.125	0.248	0.099	0.024	0.538	0.291	0.656	0.460
N=	22,626	22,776	23,067	23,068	23,067	22,626	22,776	23,067	23,068	23,067	22,476	22,476	22,619	22,619
R-squared	0.282	0.284	0.158	0.173	0.143	0.279	0.285	0.114	0.118	0.106	0.106	0.144	0.121	0.155

Panel C: Moderating results - debt illiteracy

Variables	*FD*	*TD*	*PD*	*FUD*	*SD*	*LFD*	*LTD*	*LPD*	*LFUD*	*LSD*	*FDOI5*	*FDOI3*	*TDOI5*	*TDOI3*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
*POpt**	0.039***	0.039***	0.027**	0.040***	0.031***	0.343***	0.356***	0.198**	0.153*	0.173**	0.034***	0.037***	0.038***	0.038***
Illiterate	(0.014)	(0.014)	(0.013)	(0.012)	(0.012)	(0.114)	(0.114)	(0.083)	(0.090)	(0.081)	(0.013)	(0.014)	(0.013)	(0.014)
*POpt**	0.035**	0.034**	0.014	0.031**	0.020	0.324***	0.326***	0.285***	0.271***	0.260***	0.014	0.032**	0.022*	0.037***
Literate	(0.015)	(0.014)	(0.013)	(0.012)	(0.012)	(0.120)	(0.121)	(0.089)	(0.097)	(0.087)	(0.014)	(0.015)	(0.014)	(0.014)
*IOpt**	0.034***	0.032***	0.016*	0.006	0.004	0.262***	0.255***	0.097*	0.024	0.033	0.019**	0.024**	0.016*	0.020**
Illiterate	(0.011)	(0.011)	(0.009)	(0.008)	(0.008)	(0.087)	(0.088)	(0.058)	(0.061)	(0.054)	(0.009)	(0.010)	(0.010)	(0.010)
*IOpt**	-0.048***	-0.050***	-0.029***	-0.019**	-0.018*	-0.358***	-0.384***	-0.125*	-0.150**	-0.110*	-0.016	-0.021*	-0.020*	-0.025**
Literate	(0.013)	(0.012)	(0.011)	(0.009)	(0.010)	(0.099)	(0.101)	(0.065)	(0.069)	(0.061)	(0.011)	(0.012)	(0.011)	(0.012)

Diff1:	0.831	0.758	0.377	0.561	0.434	0.889	0.828	0.405	0.300	0.394	0.194	0.732	0.317	0.926
Diff2:	0.000	0.000	0.000	0.005	0.013	0.000	0.000	0.000	0.007	0.014	0.000	0.000	0.000	0.000
N=	14,147	14,331	14,533	14,534	14,533	14,147	14,331	14,533	14,534	14,533	14,046	14,046	14,235	14,235
R-squared	0.173	0.193	0.107	0.204	0.145	0.217	0.266	0.094	0.148	0.122	0.085	0.100	0.136	0.140

Notes: This table presents the FE estimates for prudent and imprudent optimism. Coefficients are presented, and robust standard errors are given in parentheses. Diff1(2) is the Wald test performed to test whether *POpt*Loweduc* (*Illiterate*) is significant different from *IOpt*Higheduc* (*Literate*). P-value for the Wald-test is presented. The same set of control variables (include Provincial and Year dummies) are employed as in Table 4.7, but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. *TD*/*FD*/*PD*/*FUD*/*SD* stands for the probability of holding total debt/financial debt/personal debt/family usage debt/short-term debt. *LTD*/*LFD*/*LPD*/*LFUD*/*LSD* denotes the amount of total debt/financial debt/personal debt/family usage debt/short-term debt. *TDOI5*/*TDOI3* stands for the probability of holding total debt payments that exceed 50%/30%. *FDOI5*/*FDOI3* indexes the probability of holding financial debt payments that exceed 50%/30%. Definitions of all variables are shown in Table 4.3.

Table 4.13 Results for correlated random effects

Panel A: Main results														
Variables	*FD*	*TD*	*PD*	*FUD*	*SD*	*LFD*	*LTD*	*LPD*	*LFUD*	*LSD*	*FDOI5*	*FDOI3*	*TDOI5*	*TDOI3*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	CRE Probit					CRE Tobit					CRE Probit			
Opt1st	0.029*	0.020	0.039***	0.023**	0.020*	0.211*	0.127	0.333***	0.159	0.210*	0.021	0.016	0.012	0.008
	(0.016)	(0.016)	(0.013)	(0.012)	(0.012)	(0.117)	(0.118)	(0.120)	(0.123)	(0.121)	(0.014)	(0.015)	(0.014)	(0.015)
Opt2st	0.050***	0.044**	0.028*	0.014	0.007	0.390***	0.333**	0.344***	0.097	0.110	0.035**	0.030*	0.031**	0.029*
	(0.018)	(0.018)	(0.015)	(0.013)	(0.013)	(0.129)	(0.131)	(0.131)	(0.135)	(0.132)	(0.015)	(0.017)	(0.015)	(0.017)
Opt3st	-0.013	-0.019	0.022	0.025*	0.029*	-0.073	-0.137	0.118	0.057	0.103	0.012	0.001	0.004	-0.004
	(0.021)	(0.021)	(0.017)	(0.015)	(0.015)	(0.150)	(0.152)	(0.153)	(0.157)	(0.154)	(0.018)	(0.019)	(0.018)	(0.019)
Pes	0.007	0.002	0.014	0.009	0.008	0.048	-0.004	0.106	0.037	0.107	0.008	-0.005	0.001	-0.008
	(0.020)	(0.020)	(0.017)	(0.015)	(0.015)	(0.147)	(0.149)	(0.150)	(0.154)	(0.151)	(0.017)	(0.019)	(0.018)	(0.019)
N=	22,620	22,770	23,067	23,068	23,067	22,620	22,770	23,067	23,068	23,067	22,470	22,470	22,613	22,613
N2=						8,272	8,669	2,954	2,871	2,520				
N3=						14,348	14,101	20,113	20,197	20,547				

Panel B: Moderating results - low education														
Variables	*FD*	*TD*	*PD*	*FUD*	*SD*	*LFD*	*LTD*	*LPD*	*LFUD*	*LSD*	*FDOI5*	*FDOI3*	*TDOI5*	*TDOI3*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
			CRE Probit					CRE Tobit					CRE Probit	
*Opt1st**	0.031	0.018	0.040*	0.038*	0.038	0.872	0.408	1.831	1.645	2.816*	0.030	0.030	0.011	0.013
Loweduc	(0.024)	(0.024)	(0.020)	(0.022)	(0.023)	(0.593)	(0.576)	(1.222)	(1.465)	(1.521)	(0.020)	(0.023)	(0.019)	(0.022)
*Opt1st**	0.026	0.020	0.038**	0.019	0.014	0.508	0.326	1.828**	0.719	0.844	0.016	0.008	0.011	0.005
Higheduc	(0.020)	(0.019)	(0.017)	(0.013)	(0.013)	(0.414)	(0.401)	(0.751)	(0.773)	(0.795)	(0.017)	(0.018)	(0.017)	(0.018)
*Opt2st**	0.116***	0.111***	0.062**	0.031	0.032	2.870***	2.659***	4.286***	1.903	3.846**	0.062**	0.069**	0.052**	0.061**
Loweduc	(0.030)	(0.029)	(0.026)	(0.025)	(0.026)	(0.682)	(0.666)	(1.392)	(1.705)	(1.737)	(0.027)	(0.029)	(0.026)	(0.028)
*Opt2st**	0.020	0.013	0.012	0.009	-0.001	0.514	0.319	1.200	0.280	-0.057	0.024	0.013	0.022	0.017
Higheduc	(0.022)	(0.021)	(0.018)	(0.015)	(0.014)	(0.453)	(0.438)	(0.820)	(0.845)	(0.869)	(0.019)	(0.020)	(0.019)	(0.020)
*Opt3st**	-0.011	-0.026	0.043	0.031	0.068*	-0.165	-0.669	-0.169	0.057	0.666	-0.021	-0.027	-0.027	-0.032
Loweduc	(0.034)	(0.033)	(0.031)	(0.032)	(0.037)	(0.855)	(0.832)	(1.767)	(2.116)	(2.167)	(0.024)	(0.029)	(0.024)	(0.029)
*Opt3st**	-0.015	-0.018	0.012	0.024	0.020	-0.291	-0.351	0.779	0.350	0.517	0.025	0.010	0.018	0.007
Higheduc	(0.024)	(0.024)	(0.021)	(0.018)	(0.018)	(0.518)	(0.503)	(0.942)	(0.976)	(1.004)	(0.023)	(0.023)	(0.022)	(0.023)
N=	22,620	22,770	23,067	23,068	23,067	22,620	22,770	23,067	23,068	23,067	22,470	22,470	22,613	22,613
N2=						8,272	8,669	2,954	2,871	2,520				
N3=						14,348	14,101	20,113	20,197	20,547				

Panel C: Moderating results - debt illiteracy														
Variables	*FD*	*TD*	*PD*	*FUD*	*SD*	*LFD*	*LTD*	*LPD*	*LFUD*	*LSD*	*FDOI5*	*FDOI3*	*TDOI5*	*TDOI3*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
			CRE Probit					CRE Tobit					CRE Probit	
*Opt1st**	0.038**	0.029	0.048***	0.040**	0.031**	0.884	0.580	2.099**	1.465	1.700*	0.024	0.017	0.018	0.013
Illiterate	(0.019)	(0.019)	(0.017)	(0.015)	(0.015)	(0.468)	(0.455)	(0.840)	(0.890)	(0.920)	(0.016)	(0.017)	(0.017)	(0.017)
*Opt1st**	0.013	0.005	0.026	0.003	0.005	0.348	0.116	1.475	0.185	0.651	0.012	0.009	0.001	-0.001
Literate	(0.022)	(0.022)	(0.020)	(0.016)	(0.016)	(0.538)	(0.522)	(0.945)	(0.990)	(1.022)	(0.019)	(0.021)	(0.018)	(0.020)
*Opt2st**	0.059***	0.062***	0.032*	0.014	0.008	1.405***	1.435***	1.868**	0.220	0.395	0.027	0.028	0.031	0.033*
Illiterate	(0.021)	(0.021)	(0.019)	(0.016)	(0.016)	(0.511)	(0.497)	(0.910)	(0.978)	(1.003)	(0.018)	(0.019)	(0.019)	(0.020)
*Opt2st**	0.034	0.013	0.022	0.017	0.009	1.005*	0.386	1.944*	1.114	1.133	0.045*	0.030	0.028	0.022
Literate	(0.026)	(0.025)	(0.023)	(0.019)	(0.019)	(0.609)	(0.589)	(1.069)	(1.106)	(1.147)	(0.023)	(0.024)	(0.022)	(0.024)
*Opt3st**	-0.010	-0.012	0.026	0.014	0.026	-0.168	-0.265	0.267	-0.466	-0.419	0.022	0.006	0.021	0.008
Illiterate	(0.024)	(0.024)	(0.022)	(0.019)	(0.020)	(0.600)	(0.582)	(1.080)	(1.149)	(1.177)	(0.021)	(0.022)	(0.022)	(0.022)
*Opt3st**	-0.009	-0.021	0.020	0.045	0.039	-0.173	-0.472	1.179	1.505	2.173	-0.004	-0.003	-0.018	-0.017
Literate	(0.029)	(0.028)	(0.026)	(0.024)	(0.025)	(0.713)	(0.694)	(1.247)	(1.313)	(1.370)	(0.024)	(0.027)	(0.023)	(0.026)
N=	10,091	10,168	10,256	10,256	10,256	10,091	10,168	10,256	10,256	10,256	10,009	10,009	10,106	10,106
N2=						3,313	3,486	1,326	1,284	1,136				
N3=						6,778	6,682	8,930	8,972	9,120				

Notes: Panel A, B and C report the correlated random effects Probit and Tobit models for main results, low education and debt illiteracy respectively. N represents the overall observations, N2 (N3) represents the uncensored (left-censored) observations. Marginal effects are presented, and standard errors are given in parentheses. The same set of control variables (include Provincial and Year dummies) are employed as in Table 4.7, but not reported. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. *TD*/*FD*/*PD*/*FUD*/*SD* stands for the probability of holding total debt/financial debt/personal debt/family usage debt/short-term debt. *LTD*/*LFD*/*LPD*/*LFUD*/*LSD* denotes the amount of total debt/financial debt/personal debt/family usage debt/short-term debt. *TDOI5*/*TDOI3* stands for the probability of holding total debt payments that exceed 50%/30%. *FDOI5*/*FDOI3* indexes the probability of holding financial debt payments that exceed 50%/30%. Definitions of all variables are shown in Table 4.3.

Appendix 4.1 Borrowing decisions by 3-step optimists at time t-2

Variables	*lagged opt3st (2nd lag)*		Diff
	1	0	
TD	0.623	0.528	0.095***
FD	0.606	0.511	0.096***
PD	0.228	0.204	0.024***
FUD	0.183	0.18	0.003
SD	0.175	0.167	0.007
LTD	5.371	4.528	0.842***
LFD	5.035	4.222	0.814***
LPD	1.146	1.019	0.127***
LFUD	1.12	1.083	0.037
LSD	0.952	0.897	0.055
TDOI5	0.379	0.31	0.069***
FDOI5	0.34	0.275	0.065***
TDOI3	0.447	0.374	0.073***
FDOI3	0.417	0.343	0.074***
N=	3187	22763	

Notes: This table presents the two sample t-test between 3-step optimists and non 3-step optimists for borrowing decisions made at time t-2. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels.*TD*/*FD*/*PD*/*FUD*/*SD* stands for the probability of holding total debt/financial debt/personal debt/family usage debt/short-term debt. *LTD*/*LFD*/*LPD*/*LFUD*/*LSD* denotes the amount of total debt/financial debt/personal debt/family usage debt/short-term debt. *TDOI5*/*TDOI3* stands for the probability of holding total debt payments that exceed 50%/30%. *FDOI5*/*FDOI3* indexes the probability of holding financial debt payments that exceed 50%/30%. Definitions of all variables are shown in Table 4.3.

Appendix 4.2 Quantile regressions for debt-to-income ratio

I have performed the FE quantile regressions (QR) for financial debt-to-income ratio and total debt-to-income ratio. Specifically, I run separate regressions for the 10^{th}, 30^{th}, 50^{th}, 70^{th}, 90^{th} percentiles of the distribution of debt-to-income ratio. Columns 1-5 of Appendix 4.2a present the FE QR estimates for financial debt-to-income ratio. I can see that 1-step and 2-step optimists show a 0.0437 unit higher debt-to-income ratio at the 70^{th} and 90^{th} percentiles. This finding supports the construction of my dummy debt-to-income ratio indicator, implying that the significant relationship between optimism and financial debt-to-income ratio is observed at the higher end of the distribution.

Columns 6-10 of Appendix 4.2a report the FE QR estimates for total debt-to-income ratio. I find no significant relationship between optimism and total debt-to-income ratio in all equations. To sum up, I only find significant relationship between optimism and debt-to-income ratio in 2 out of 10 equations. This suggests that using the entire distribution of the debt-to-income ratio generally produces insufficient results with lesser explanatory power than using the dummy indicators.

Appendix 4.2a FE Quantile Regressions results for the debt-to-income ratio

	Financial debt-to-income ratio					Total debt-to-income ratio				
Variables	10th	30th	50th	70th	90th	10th	30th	50th	70th	90th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Opt1st	0.0437	0.0437	0.0437	0.0437**	0.0437**	-0.0031	-0.0031	-0.0031	-0.0031	-0.0031
	(0.0365)	(0.0365)	(0.0365)	(0.0214)	(0.0214)	(0.0711)	(0.0711)	(0.0419)	(0.0419)	(0.0419)
Opt2st	0.0410	0.0410	0.0410	0.0410*	0.0410*	0.0575	0.0575	0.0575	0.0575	0.0575
	(0.0417)	(0.0417)	(0.0417)	(0.0245)	(0.0245)	(0.0741)	(0.0741)	(0.0437)	(0.0437)	(0.0437)
Opt3	-0.0115	-0.0115	-0.0115	-0.0115	-0.0115	-0.0541	-0.0541	-0.0541	-0.0541	-0.0541
	(0.0502)	(0.0502)	(0.0502)	(0.0295)	(0.0295)	(0.0912)	(0.0912)	(0.0538)	(0.0538)	(0.0538)
Pes	0.0423	0.0423	0.0423	0.0423	0.0423	0.0172	0.0172	0.0172	0.0172	0.0172
	(0.0452)	(0.0452)	(0.0452)	(0.0265)	(0.0265)	(0.0891)	(0.0891)	(0.0526)	(0.0526)	(0.0526)
Age	-0.0072	-0.0072	-0.0072	-0.0072	-0.0072	-0.0499	-0.0499	-0.0499	-0.0499	-0.0499
	(0.0382)	(0.0382)	(0.0382)	(0.0224)	(0.0224)	(0.0693)	(0.0693)	(0.0409)	(0.0409)	(0.0409)
Age square	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002	-0.0002	-0.0002**	-0.0002**	-0.0002**
	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Married	0.0506	0.0506	0.0506	0.0506	0.0506	0.1350	0.1350	0.1350**	0.1350**	0.1350**

	(0.0588)	(0.0588)	(0.0588)	(0.0345)	(0.0345)	(0.1100)	(0.1100)	(0.0649)	(0.0649)	(0.0649)
Divorce	0.1770	0.1770	0.1770	0.1770*	0.1770*	0.3430*	0.3430*	0.3430***	0.3430***	0.3430***
	(0.1550)	(0.1550)	(0.1550)	(0.0908)	(0.0908)	(0.2050)	(0.2050)	(0.1210)	(0.1210)	(0.1210)
Widow	0.0513	0.0513	0.0513	0.0513	0.0513	-0.0072	-0.0072	-0.0072	-0.0072	-0.0072
	(0.0844)	(0.0844)	(0.0844)	(0.0496)	(0.0496)	(0.1400)	(0.1400)	(0.0825)	(0.0825)	(0.0825)
hhsize	-0.0017	-0.0017	-0.0017	-0.0017	-0.0017	-0.0044	-0.0044	-0.0044	-0.0044	-0.0044
	(0.0055)	(0.0055)	(0.0055)	(0.0032)	(0.0032)	(0.0107)	(0.0107)	(0.0063)	(0.0063)	(0.0063)
Urban	-0.0106	-0.0106	-0.0106	-0.0106	-0.0106	0.125	0.125	0.125	0.125	0.125
	(0.0753)	(0.0753)	(0.0753)	(0.0442)	(0.0442)	(0.1400)	(0.1400)	(0.0824)	(0.0824)	(0.0824)
No school	-0.1	-0.1	-0.1	-0.1	-0.1	-0.702	-0.702	-0.702*	-0.702*	-0.702*
	(0.1530)	(0.1530)	(0.1530)	(0.0900)	(0.0900)	(0.6190)	(0.6190)	(0.3650)	(0.3650)	(0.3650)
Primary	-0.0942	-0.0942	-0.0942	-0.0942	-0.0942	-0.283	-0.283	-0.283*	-0.283*	-0.283*
	(0.1240)	(0.1240)	(0.1240)	(0.0727)	(0.0727)	(0.2730)	(0.2730)	(0.1610)	(0.1610)	(0.1610)
Secondary	-0.025	-0.025	-0.025	-0.025	-0.025	-0.123	-0.123	-0.123	-0.123	-0.123
	(0.0747)	(0.0747)	(0.0747)	(0.0439)	(0.0439)	(0.1550)	(0.1550)	(0.0915)	(0.0915)	(0.0915)
Food exp	0.00628	0.00628	0.00628	0.00628	0.00628	-0.00104	-0.00104	-0.00104	-0.00104	-0.00104
	(0.0321)	(0.0321)	(0.0321)	(0.0189)	(0.0189)	(0.0587)	(0.0587)	(0.0347)	(0.0347)	(0.0347)
Non-food exp	0.0537***	0.0537***	0.0537***	0.0537***	0.0537***	0.166***	0.166***	0.166***	0.166***	0.166***

	(0.0173)	(0.0173)	(0.0173)	(0.0101)	(0.0101)	(0.0361)	(0.0361)	(0.0213)	(0.0213)	(0.0213)
Net worth	0.0150***	0.0150***	0.0150***	0.0150***	0.0150***	-0.0057	-0.0057	-0.0057	-0.0057	-0.0057
	(0.0051)	(0.0051)	(0.0051)	(0.0030)	(0.0030)	(0.0089)	(0.0089)	(0.0053)	(0.0053)	(0.0053)
Unemployed	-0.0393	-0.0393	-0.0393	-0.0393	-0.0393	-0.0668	-0.0668	-0.0668	-0.0668	-0.0668
	(0.0414)	(0.0414)	(0.0414)	(0.0243)	(0.0243)	(0.0741)	(0.0741)	(0.0437)	(0.0437)	(0.0437)
Observations	22,470	22,470	22,470	22,470	22,470	22,613	22,613	22,613	22,613	22,613

Notes: This table presents the FE quantile regression results for the debt-to-income ratio. Provincial and Year dummies are included but not reported. Coefficients are presented, and robust standard errors are given in parentheses. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. Definitions of all variables are shown in Table 4.3.

Appendix 4.3 Coefficients of other control variables - low education models

Variables	*FD*	*TD*	*PD*	*FUD*	*SD*	*LFD*	*LTD*	*LPD*	*LFUD*	*LSD*	*FDOI5*	*FDOI3*	*TDOI5*	*TDOI3*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Pes	0.008	0.006	0.019	0.017	0.015	0.056	0.023	0.109	0.075	0.116	0.012	0.000	0.008	-0.002
	(0.020)	(0.020)	(0.017)	(0.015)	(0.015)	(0.164)	(0.166)	(0.112)	(0.115)	(0.105)	(0.017)	(0.018)	(0.017)	(0.018)
Loweduc	-0.135**	-0.134**	-0.086*	-0.067	-0.038	-1.143**	-1.138**	-0.302	-0.277	-0.259	-0.096**	-0.0713	-0.100**	-0.076
	(0.056)	(0.056)	(0.048)	(0.044)	(0.0438)	(0.455)	(0.464)	(0.312)	(0.346)	(0.303)	(0.047)	(0.051)	(0.049)	(0.052)
Age	0.014	0.013	0.028**	0.028**	0.026**	0.152	0.114	0.186**	0.232**	0.188**	0.007	0.018	0.003	0.016
	(0.016)	(0.016)	(0.014)	(0.012)	(0.012)	(0.132)	(0.134)	(0.090)	(0.094)	(0.085)	(0.013)	(0.015)	(0.014)	(0.015)
Age square	-0.0001***	-0.0002***	-0.0003***	-0.0003***	-0.0003***	-0.0018***	-0.0017***	-0.0019***	-0.0022***	-0.0020***	-0.0001***	-0.0001***	-0.0001***	-0.0001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	0.007	0.010	0.002	0.015	-0.008	0.129	0.186	0.147	0.202	0.050	0.029	0.042*	0.044**	0.055**
	(0.024)	(0.024)	(0.022)	(0.019)	(0.019)	(0.195)	(0.196)	(0.141)	(0.141)	(0.133)	(0.021)	(0.022)	(0.021)	(0.022)
Divorce	0.055	0.066	0.025	0.077*	0.037	0.505	0.737*	0.296	0.429	0.289	0.024	0.078	0.076*	0.108**
	(0.051)	(0.048)	(0.046)	(0.042)	(0.041)	(0.427)	(0.402)	(0.319)	(0.332)	(0.308)	(0.047)	(0.050)	(0.044)	(0.048)
Widow	-0.023	-0.028	0.039	0.011	0.017	-0.054	-0.080	0.186	0.071	0.065	0.040	0.049	0.047*	0.050
	(0.035)	(0.035)	(0.030)	(0.023)	(0.025)	(0.279)	(0.284)	(0.187)	(0.180)	(0.163)	(0.028)	(0.031)	(0.029)	(0.032)
HHsize	0.006**	0.004*	-0.006***	-0.005***	-0.002	0.047**	0.031	-0.047***	-0.033**	-0.028**	-0.001	0.003	0.000	0.003
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.022)	(0.022)	(0.015)	(0.015)	(0.014)	(0.002)	(0.002)	(0.002)	(0.002)
Urban	0.007	0.021	-0.032	-0.005	-0.039	0.049	0.178	-0.117	-0.060	-0.232	0.005	0.005	0.040	0.030

	(0.031)	(0.032)	(0.026)	(0.024)	(0.024)	(0.258)	(0.268)	(0.174)	(0.188)	(0.173)	(0.028)	(0.029)	(0.028)	(0.029)
Secondary	-0.044	-0.043	-0.070**	-0.084***	-0.060**	-0.474*	-0.493*	-0.439**	-0.404*	-0.365*	-0.040	-0.017	-0.045	-0.027
	(0.033)	(0.032)	(0.029)	(0.028)	(0.029)	(0.266)	(0.269)	(0.199)	(0.224)	(0.212)	(0.029)	(0.030)	(0.030)	(0.030)
HH Income	0.075***	0.077***	0.021**	0.028***	0.014*	0.686***	0.737***	0.131**	0.186***	0.155***				
	(0.010)	(0.010)	(0.009)	(0.008)	(0.008)	(0.084)	(0.085)	(0.057)	(0.058)	(0.054)				
Food exp	0.007	0.014	0.002	0.016	0.007	0.087	0.169	-0.014	0.071	-0.009	0.025**	0.024**	0.028***	0.026**
	(0.013)	(0.013)	(0.011)	(0.010)	(0.010)	(0.106)	(0.109)	(0.071)	(0.075)	(0.069)	(0.011)	(0.012)	(0.011)	(0.012)
Non-food exp	0.045***	0.048***	0.035***	0.030***	0.028***	0.422***	0.488***	0.175***	0.219***	0.173***	0.034***	0.041***	0.041***	0.046***
	(0.007)	(0.007)	(0.006)	(0.005)	(0.005)	(0.058)	(0.059)	(0.039)	(0.041)	(0.038)	(0.006)	(0.006)	(0.006)	(0.006)
Net worth	0.011***	0.012***	0.001	-0.003**	-0.001	0.076***	0.072***	0.005	-0.022*	-0.006	0.007***	0.006***	0.004***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.017)	(0.018)	(0.013)	(0.013)	(0.012)	(0.002)	(0.002)	(0.002)	(0.002)
Unemployed	-0.029	-0.029	-0.040***	-0.030**	-0.039***	-0.289**	-0.305**	-0.254**	-0.225**	-0.219**	-0.031**	-0.038**	-0.034**	-0.040**
	(0.018)	(0.018)	(0.015)	(0.013)	(0.013)	(0.139)	(0.141)	(0.099)	(0.096)	(0.093)	(0.015)	(0.016)	(0.015)	(0.016)
N=	22,626	22,776	23,067	23,068	23,067	22,626	22,776	23,067	23,068	23,067	22,476	22,476	22,619	22,619
R-squared	0.283	0.286	0.158	0.173	0.143	0.280	0.286	0.114	0.118	0.105	0.107	0.144	0.122	0.156
Number of ID	17,659	17,729	17,939	17,940	17,939	17,659	17,729	17,939	17,940	17,939	17,588	17,588	17,633	17,633

Notes: This table is a complementary table for Table 4.9. This table presents the estimated coefficients of other control variables in the low education models. Coefficients are presented, and robust standard errors are given in parentheses. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. *TD*/*FD*/*PD*/*FUD*/*SD* stands for the probability of holding total debt/financial debt/personal debt/family usage debt/short-term debt. *LTD*/*LFD*/*LPD*/*LFUD*/*LSD* denotes the amount of total debt/financial debt/personal debt/family usage debt/short-term debt. *TDOI5*/*TDOI3* stands for the probability of holding total debt payments that exceed 50%/30%. *FDOI5*/*FDOI3* indexes the probability of holding financial debt payments that exceed 50%/30%. Definitions of all variables are shown in Table 4.3.

Appendix 4.4 Coefficients of other control variables - debt illiteracy models

Variables	*FD*	*TD*	*PD*	*FUD*	*SD*	*LFD*	*LTD*	*LPD*	*LFUD*	*LSD*	*FDOI5*	*FDOI3*	*TDOI5*	*TDOI3*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Pes	-0.012	-0.011	0.012	-0.001	0.001	-0.100	-0.099	0.133*	0.010	0.065	-0.001	-0.009	-0.003	-0.011
	(0.013)	(0.013)	(0.011)	(0.010)	(0.010)	(0.105)	(0.106)	(0.073)	(0.077)	(0.070)	(0.011)	(0.012)	(0.012)	(0.012)
Male	0.008	0.009	-0.060***	-0.022***	-0.052***	0.151**	0.174***	-0.435***	-0.167***	-0.386***	0.026***	0.022***	0.032***	0.027***
	(0.008)	(0.008)	(0.007)	(0.006)	(0.006)	(0.066)	(0.067)	(0.044)	(0.048)	(0.042)	(0.007)	(0.008)	(0.007)	(0.008)
Age	0.006***	0.006***	0.020***	0.020***	0.0179***	0.066***	0.069***	0.107***	0.136***	0.102***	0.007***	0.007***	0.008***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.010)	(0.010)	(0.006)	(0.006)	(0.006)	(0.001)	(0.001)	(0.001)	(0.001)
Age square	-0.0001***	-0.0001***	-0.0002***	-0.0002***	-0.0001***	-0.0008***	-0.0008***	-0.001***	-0.0014***	-0.0010***	-0.0001***	-0.0001***	-0.0001***	-0.0001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Married	-0.001	0.008	-0.008	0.016*	-0.012	0.000	0.105	-0.024	0.087	-0.033	-0.004	-0.002	0.007	0.010
	(0.011)	(0.010)	(0.010)	(0.009)	(0.009)	(0.085)	(0.085)	(0.064)	(0.068)	(0.061)	(0.009)	(0.010)	(0.010)	(0.010)
Divorce	0.024	0.032	0.022	0.046**	0.024	0.167	0.286	0.049	0.259	0.138	-0.002	0.029	0.013	0.042*
	(0.026)	(0.025)	(0.025)	(0.023)	(0.024)	(0.211)	(0.208)	(0.167)	(0.184)	(0.166)	(0.024)	(0.026)	(0.024)	(0.025)
Widow	0.011	0.014	0.005	0.021*	-0.002	0.116	0.178	0.053	0.131	-0.012	0.016	0.019	0.026*	0.026*
	(0.017)	(0.017)	(0.015)	(0.013)	(0.014)	(0.133)	(0.135)	(0.094)	(0.096)	(0.086)	(0.014)	(0.016)	(0.015)	(0.016)
hhsize	0.015***	0.012***	-0.003***	-0.009***	-0.006***	0.105***	0.064***	-0.031***	-0.058***	-0.045***	0.003***	0.005***	0.001	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.011)	(0.011)	(0.007)	(0.008)	(0.007)	(0.001)	(0.001)	(0.001)	(0.001)
Urban	0.061***	0.063***	0.024***	0.034***	0.031***	0.564***	0.574***	0.187***	0.241***	0.199***	0.062***	0.071***	0.065***	0.076***
	(0.010)	(0.010)	(0.009)	(0.008)	(0.008)	(0.083)	(0.083)	(0.057)	(0.060)	(0.054)	(0.009)	(0.010)	(0.009)	(0.010)
No school	-0.124***	-0.119***	-0.220***	-0.315***	-0.279***	-1.160***	-1.180***	-1.346***	-1.942***	-1.579***	-0.087***	-0.098***	-0.112***	-0.119***

	(0.021)	(0.021)	(0.018)	(0.016)	(0.016)	(0.169)	(0.171)	(0.114)	(0.119)	(0.108)	(0.018)	(0.020)	(0.018)	(0.020)
Primary	-0.075***	-0.079***	-0.187***	-0.309***	-0.250***	-0.842***	-0.978***	-1.265***	-1.877***	-1.538***	-0.092***	-	-0.133***	-0.117***
	(0.017)	(0.016)	(0.016)	(0.015)	(0.015)	(0.140)	(0.141)	(0.104)	(0.114)	(0.102)	(0.016)	(0.017)	(0.016)	(0.016)
Secondary	-0.042***	-0.045***	-0.136***	-0.206***	-0.178***	-0.513***	-0.635***	-0.828***	-1.217***	-1.005***	-0.047***	-0.047***	-0.080***	-0.074***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.114)	(0.114)	(0.094)	(0.104)	(0.094)	(0.013)	(0.014)	(0.013)	(0.013)
HH Income	0.092***	0.102***	0.025***	0.064***	0.039***	0.972***	1.178***	0.215***	0.393***	0.281***				
	(0.007)	(0.007)	(0.006)	(0.005)	(0.005)	(0.055)	(0.056)	(0.037)	(0.040)	(0.036)				
Food exp	-0.032***	-0.031***	-0.018**	-0.012	-0.015**	-0.223***	-0.201***	-0.144***	-0.091*	-0.105**	0.001	0.000	0.0136*	0.006
	(0.009)	(0.009)	(0.008)	(0.007)	(0.007)	(0.071)	(0.072)	(0.050)	(0.053)	(0.047)	(0.007)	(0.008)	(0.008)	(0.008)
Non-food	0.069***	0.076***	0.045***	0.051***	0.047***	0.661***	0.822***	0.243***	0.296***	0.255***	0.063***	0.076***	0.090***	0.099***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.040)	(0.040)	(0.027)	(0.028)	(0.026)	(0.004)	(0.004)	(0.004)	(0.004)
Net worth	0.002	0.003	-0.009***	-0.003	-0.004**	0.029	0.045**	-0.032**	-0.023	-0.016	0.001	0.002	0.004**	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.020)	(0.020)	(0.015)	(0.016)	(0.015)	(0.002)	(0.002)	(0.002)	(0.002)
Unemployed	0.012	0.015	-0.026***	-0.046***	-0.027***	0.016	0.058	-0.179***	-0.280***	-0.159***	-0.003	-0.008	-0.004	-0.008
	(0.012)	(0.012)	(0.010)	(0.009)	(0.009)	(0.094)	(0.095)	(0.061)	(0.062)	(0.058)	(0.010)	(0.011)	(0.010)	(0.011)
N=	14,147	14,331	14,533	14,534	14,533	14,147	14,331	14,533	14,534	14,533	14,046	14,046	14,235	14,235
R-squared	0.171	0.192	0.106	0.203	0.145	0.216	0.264	0.093	0.147	0.120	0.086	0.099	0.136	0.140

Notes: This table is a complementary table for Table 4.10. This table presents the estimated coefficients of other control variables in the debt illiteracy models. Coefficients are presented, and robust standard errors are given in parentheses. Statistical significance is indicated by * at the 10%, ** by 5%, and by *** at the 1% levels. *TD*/*FD*/*PD*/*FUD*/*SD* stands for the probability of holding total debt/financial debt/personal debt/family usage debt/short-term debt. *LTD*/*LFD*/*LPD*/*LFUD*/*LSD* denotes the amount of total debt/financial debt/personal debt/family usage debt/short-term debt. *TDOI5*/*TDOI3* stands for the probability of holding total debt payments that exceed 50%/30%. *FDOI5*/*FDOI3* indexes the probability of holding financial debt payments that exceed 50%/30%. Definitions of all variables are shown in Table 4.3.

Chapter Five: Conclusion

5.1 Summary of findings and policy implication

This thesis comprises three empirical studies on errors in financial expectations using micro-level survey data from Russia and South Africa. More precisely, I examine the role of past income changes and job insecurity in determining errors in financial expectations in Russia, the association between errors in financial expectations and unhealthy behaviours in Russia, and whether financial forecast errors affect individuals' borrowing decisions in South Africa. This thesis adds to the existing literature by providing new empirical evidence and associated policy implications on several heated topics.

In Chapter Two, using the 2002-2016 waves of the Russian Longitudinal Monitoring Survey, I find that individuals are more likely to make optimistic forecast errors following a positive change in their financial situation, while after a negative change in financial situation, they are more likely to be over-pessimistic regarding their future finances. This is in line with the over-extrapolation of past experience (Massenot and Pettinicchi, 2019). I also find that individuals who have a more stable job prospect are more likely to over-estimate their future financial situation. Meanwhile, those who feel insecure are more likely to show pessimistic forecast errors. These findings are robust to controlling for endogeneity and to using different proxies for forecast errors.

Given that over-optimistic individuals are more likely to display imprudent financial behaviours (less saving, more borrowing), they may run the risk of over-exposing themselves to adverse economic outcomes (being over-indebted, unemployment, financial crisis). Thus, the empirical findings in this chapter provide the following policy implications: First, policy makers can reduce individuals' forecast errors by implementing financial education schemes and classes in schools, community and even corporations to teach individuals how to develop prudent financial planning and expectations regardless of their past experience. Second, policies aimed at increasing the availability and accessibility of job and skill training programmes can help reduce workers insecurity in case of job loss. Third, policies targeted at increasing the resources dedicated to providing social unemployment benefits and insurances can relieve worker from the stress related to finding a new job in the case of sudden lay off.

In Chapter Three, using data on 120,000 observations from the 2001-2016 waves of the Russian Longitudinal Monitoring Survey, I find that extreme optimists show a significantly higher probability of binge drinking and higher levels of alcohol consumption in the male sample, while they show a higher probability of smoking and higher levels of cigarette consumption in the female sample. Additionally, I find that psychological ill-being is the channel linking extreme optimism to smoking and drinking. This suggests that Russian men and women have different coping strategies in response to psychological stress. Male optimists are more likely to translate stress into drinking, while female optimists are more prone to use smoking to relieve psychological stress. The empirical findings are robust to using both static and dynamic models.

As for policy implications arisen from this chapter, policies aimed at reducing extreme optimism such as financial education programmes can provide a new perspective for Russia's anti-smoking and anti-drinking programmes. Next, considering that optimists may use alcohol and cigarettes to combat psychological ill-being, more interventions and workshops on how to control and address individuals' negative emotions and moods are needed.

In Chapter Four, I make use of the National Income Dynamics Survey to examine the association between errors in financial expectations and borrowing in South Africa. I find that respondents who make optimistic forecast errors show higher levels of debt and a higher probability of being over-indebted, while pessimism is not significantly associated with borrowing decisions. I further examine the moderating role of low education and debt illiteracy and find that the low educated or debt illiterate optimists are more likely to borrow and to become over-indebted. This suggests that individuals are more likely to borrow due to a lack of understanding of debt contracts and a miscalculation of the cost of loan products (Lusardi and Tufano, 2015). Thus, policy makers could pay more attention on how to increase the transparency of loan products, as well as the debt literacy levels of consumers. Next, policies targeted at providing financial advice for unsophisticated households and communities could help consumers reduce the cost of misunderstanding of debt contracts and products.

5.2 Future research

Regarding Chapter Two, Dickerson and Green (2012) suggest that job insecurity consists of various other components (e.g., risk of job loss and skill level of the job). Thus future research

could investigate how different proxies of job, or more generally economic, insecurity affect individuals' forecast errors in alternative ways. Next, individuals may build their expectations based on past forecast errors, so future research could examine the determinants of expectation errors in a dynamic context.

As for Chapter Three, my work could be biased by the self-reported survey data. Future research could use experimental studies to correct for self-reported bias and further examine the casual mechanisms linking expectation errors and unhealthy behaviours.

As for Chapter Four, there is a lack of evidence exploring the casual mechanisms linking expectation errors and borrowing in both developed and developing countries. Future research could explore the potential casual relationship between individuals' expectations (errors) and borrowing. Secondly, given that borrowing decisions and other financial behaviours (saving, consumption) are closely related, future research could explore the relationship between expectation errors and saving, consumption and portfolio choice in South Africa and/or other developing countries.

Reference

Abbey, A., Smith, M.J. & Scott, R.O. 1993. The relationship between reasons for drinking alcohol and alcohol consumption: An interactional approach. *Addictive Behaviors*, 18(6), 659–670.

Arellano, M. & Bover, O. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1), 29-51.

Argent, J. 2009. *Household income: Report on NIDS wave 1*. Available at: http://www.nids.uct.ac.za/publications/technical-papers/110-nids-technical-paper-no3/file[Accessed: 2nd Oct 2020].

Armor, D.A. & Taylor, S.E. 1997. *Mindset, prediction and performance: The causes and consequences of optimism*. University of California, Los Angeles.

Armor, D.A. & Taylor, S.E. 1998. Situated optimism: specific outcome expectancies and self-regulation. *Advances in Experimental Psychology*, 30, 309-379.

Balasuriya, J., Muradoglu, Y.G. & Ayton, P. 2010. *Is it better to be optimistic? - Financial optimism and well-being*. Available at SSRN: http://ssrn.com/abstract=2307065 [Accessed: 2nd Oct 2020].

Barro, R.J. 1974. Are government bonds net wealth. *Journal of Political Economy*, 82(6), 1095-117.

Bertoni, M. 2015. Hungry today, unhappy tomorrow? Childhood hunger and subjective well-being later in life. *Journal of Health Economics*, 40, 40-53.

Bissonnette, L. & Van Soest, A. 2015. Heterogeneity in consumers' income and pension expectations. *Journal of Pension, Economics & Finance*, 14(4), 439-465.

Blundell, R. & Bond, S. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115-143.

Blundell, R., Bond, S. & Windmeijer, F. 2001. *Estimation in dynamic panel data models:*

improving on the performance of the standard GMM estimator. In Badi H. Baltagi, Thomas B. Fomby, R. Carter Hill (ed.) Nonstationary Panels, Panel Cointegration, and Dynamic Panels (Advances in Econometrics, Volume 15), Emerald Group Publishing Limited, pp.53-91.

Brody, A.L. 2006. Functional brain imaging of tobacco use and dependence. *Journal of Psychiatric Research*, 40(5), 404-418.

Brown, S., Garino, G. & Taylor, K. *et al.* 2005. Debt and financial expectation: An individual- and household-level analysis. *Economic Inquiry*, 43 (1), 100-120.

Brown, S., Garino, G. & Taylor, K. 2008. Mortgages and financial expectations: A household-level analysis. *Economic Inquiry*, 74 (3), 857-878.

Brown, S. & Taylor, K. 2006. Financial expectations, consumption and saving: A microeconomic analysis. *Fiscal Studies*, 27 (3), 313-338.

Brunnermeier, M.K. & Parker, J.A. 2005. Optimal expectations. *The American Economic Review*, 95(4), 1092-1119.

Bryan, M., Taylor, M. & Veliziotis, M. 2010. *Over-indebtedness in Great Britain: An analysis using the wealth and assets survey and household annual debtors survey*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/31897/11-747-over-indebtedness-in-great-britain-analysis.pdf [Accessed: 2nd Oct 2020].

Buis, M.L. 2010. Interpretation of interactions in non-linear models. *The Stata Journal*, 10(2), 305-308.

Calvet, L.E., Campbell, J.Y. & Sodini, P. 2009. Measuring the financial sophistication of households. *American Economic Review*, 99(2), 393-399.

Carroll, C.D. 1992. The Buffer-Stock Theory of saving: Some macroeconomic Evidence. *Brookings Papers on Economic Activity*, 1992 (2), 61-156.

Carroll, C.D. 1997. Buffer-stock saving and the life-cycle/permanent income Hypothesis. *Quarterly Journal of Economics*, 112(1), 1-55.

Carver, C.S., Scheier, M.F. & Segerstrom, S.C. 2010. Optimism. *Clinical Psychological Review*, 30(7), 879-889.

Chamberlain, G. 1980. Analysis of covariance with qualitative data. *Review of Economic Studies*, 47, 225-238.

Cocco, J.F., Gomes, F. & Lopes, P. 2017. *Evidence on the links between household finances, expectations and emotional well-being*. Available at: http://faculty.london.edu/fgomes/cgl.pdf [Accessed: 2nd Oct 2020].

Cocco, J.F., Gomes, F. & Lopes, P. 2019. *Evidence on expectations of household finances*. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3362495 [Accessed: 2nd Oct 2020].

Cockerham, W.C. 2000. Health lifestyles in Russia. *Social Science & Medicine*, 51(9), 1313-1324.

Cockerham, W.C., Hinote, B.P. & Abbott, P. 2006. Psychological distress, gender and health lifestyles in Belarus, Kazakhstan, Russia and Ukraine. *Social Science & Medicine*, 63(9), 2381-2394.

Conger, J. 1956. Reinforcement theory and the dynamics of alcoholism. *Quarterly Journal of Studies on Alcohol*, 17, 296-305.

Cox, D. & Jappelli, T. 1993. The effect of borrowing constraints on consumer liabilities. *Journal of Money, Credit & Banking*, 25(2), 197-203.

Crook, J. 2001. The demand for household debt in the USA: evidence from the 1995 survey of consumer finance. *Applied Financial Economics*, 11(1), 83-91.

Das, M. & Donkers, B. 1999. How certain are Dutch households about future income? An empirical analysis. *The Review of Income & Wealth*, 45(3), 325-338.

Das, M. & van Soest, A. 1997. Expected and realized income changes: evidence from the dutch socio-economic panel. *Journal of Economic Behavior & Organization,* 32 (1), 137-154.

Das, M. & van Soest, A. 1999. A panel data model for subjective information on household income growth. *Journal of Economic Behavior & Organization,* 40 (4), 409-426.

Davidson, K. & Prkachin, K. 1997. Optimism and unrealistic optimism have an interacting impact on health-promoting behaviour and knowledge changes. *Personality & Social Psychology Bulletin*, 23(6), 617-625.

Del Rio, A. & Young, G. 2006. The Determinants of Unsecured Borrowing: Evidence from the BHPS. *Applied Financial Economics*, 16 (15), 1119-1144.

Dewberry,C., Ing, M. & James, S. *et al.* 1990. Anxiety and unrealistic optimism. *Journal of Social Psychology*, 130(2), 151-156.

Dickerson, A. & Green, F. 2012. Fears and realisations of employment insecurity. *Labour Economics*, 19(2), 198-210.

Dillard, A.J., Midboe, A.M. & Klein, W.M.P. 2009. The dark side of optimism: Unrealistic optimism about problems with alcohol predicts subsequent negative event experiences. *Personality & Social Psychology Bulletin*, 35, 1540-1550.

Disney, R. & Gathergood, J. 2013. Financial literacy and consumer credit portfolios. *Journal of Banking & Finance*, 37(7), 2246-2254.

Dominitz, J. 1998. Earnings expectations, revisions and realizations. *The Review of Economics & Statistics*, 80(3), 374-388.

Duval, L & Wolff, F.C. 2012. Longitudinal evidence on financial expectations in Albania - Do remittance matter? *Economics of Transition*, 20(1),137-161.

Fidrmuc, J., Hake, M., & Stix, H. 2013. Households' foreign currency borrowing in Central and Eastern Europe. *Journal of Banking & Finance*, 37(6), 1880-1897.

Finn, A., Franklin, S. & Keswell, M. *et al*. 2009. *Expenditure: Report on NIDS wave 1*. Available at: http://www.nids.uct.ac.za/publications/technical-papers/111-nids-technical-paper-no4/file [Accessed: 2nd Oct 2020].

Floyd, D.L., Prentice-Dunn, S. & Rogers, R.W. 2000. A meta-analysis of research on protection motivation theory. *Journal of Applied Social Psychology*, 30(2), 407–429.

Geers, A.L., Wellman, J.A. & Fowler, S.L. 2013. Comparative and dispositional optimism as separate and interactive predictors. *Psychology & Health*, 28(1), 30-48.

Giltay, E.J., Geleijnse, J.M. & Zitman, F.G. *et al.* 2007. Lifestyle and dietary correlates of dispositional optimism in men: the Zutphen Elderly Study. *Journal of Psychosomatic Research*, 63(5), 483-490.

Grohmann, A., Menkhoff, L. & Merkle, C. *et al*. 2019. *Earn more tomorrow: Overconfident income expectations and consumer indebtedness*. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3372805 [Accessed: 2nd Oct 2020].

Hansen, L. P. 1982. Large sample properties of GMM estimators. *Econometrica*, 50, 1029-1054.

Harris, M., Loundes, J. & Webster, E. 2002. Determinants of Household Saving in Australia. *The Economic Record*, 78 (241), 207-223.

Herzfeld, T., Huffman, S. & Rizov, M. 2014. The dynamics of food, alcohol and cigarettes consumption in Russia during transition. *Economics & Human Biology*, 13, 128-143.

Hinkle, D.E., Wiersma, W. & Jurs, S.G. 2003. *Applied Statistics for the Behavioral Sciences.* 5th ed. Boston: Houghton Mifflin.

Hyytinen, A. & Putkuri, H. 2018. Household optimism and overborrowing. *Journal of Money, Credit & Banking*, 50(1), 55-76.

Ikard, F.F., Green, D. & Horn, D. 1969. A scale to differentiate between types of smoking as related to management of affect. *International Journal of the Addictions*, 4, 649-659.

Iwanaga, M., Yokoyama, H. & Seiwa, H. 2004. Coping availability and stress reduction for optimistic and pessimistic individuals. *Personality & Individual Differences*, 36(1), 11-22.

Jacobsen, B., Lee, J.B. & Marquering, W. *et al*. 2014. Gender differences in optimism and asset allocation. *Journal of Economic Behavior & Organization*, 107, 630-651.

Jefferson, A., Bortolotti, L. & Kuzmanovic, B. 2017. What is unrealistic optimism? *Consciousness & Cognition*, 50, 3-11.

Jiang, H. & Livingston, M. 2015. The dynamic effects of changes in prices and affordability on alcohol consumption: An impulse response analysis. *Alcohol & Alcoholism*, 50(6), 631-638.

Jones, A.M., Koolman, X. & Rice, N. 2006. Health-related non-response in the BHPS and ECHP: Using inverse probability weighted estimators in nonlinear models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169 (3), 543-569.

Juster, F.T., Cao, H. & Couper, M.P. 2007. *Enhancing the quality of data on the measurement of income and wealth*. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1095815 [Accessed: 2nd Oct 2020].

Kelloniemi, H., Ek, E. & Laitinen, J. 2005. Optimism, dietary habits, body mass index and smoking among young Finnish adults. *Appetite*, 45(2), 169-176.

Khodarahimi, S. 2015. The role of marital status in emotional intelligence, happiness, optimism and hope. Journal of Comparative Family Studies, 46(3), 351-371.

Kleibergen, F. & Paap, R. 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133 (1), 97–126.

Klove, E. & Mehlum, H. 2018. Positive illusions and the temptation to borrow. *Oxford Economic Papers*, 71 (3), 623-644.

Kossova, T., Kossova, E. & Sheluntcova, M. 2018. Anti-smoking policy in Russia: Relevant factors and program planning. *Evaluation and Program Planning*, 69, 43-52.

Kozyreva, P., Kosolapov, M. & Popkin, B.M. 2016. Data Resource Profile: The Russia Longitudinal Monitoring Survey - Higher School of Economics (RLMS-HSE) Phase II: Monitoring the Economic and Health Situation in Russia, 1994-2013. *International Journal of Epidemiology*, 45(2), 395-401.

Lee, J. 2008. Psychological aspects of household debt decision: The use of the Heckman's procedure. *International Journal of Human Ecology*, 9 (1), 81-95.

Leibbrandt, M., Woolard, I. & de Villiers, L. 2009. *Methodology: Report on NIDS wave 1*. Available at: http://www.nids.uct.ac.za/publications/technical-papers/108-nids-technical-paper-no1/file [Accessed: 2nd Oct 2020].

Leonard, K.E. & Blane, H.T. 1987. *Psychological theories of drinking and alcoholism*. New York: Guilford Press.

Linz, S. & Semykina, A. 2010. Perceptions of economic insecurity: Evidence from Russia. *Economic Systems*, 34(4), 357-385.

Lusardi, A. & Tufano, P. 2015. Debt literacy, financial experiences and overindebtedness. *Journal of Pension Economics & Finance*, 14(4), 332-368.

Machado, J.A. & Silva, S. 2019. Quantiles via moments. *Journal of Econometrics*, 213(1),145-173.

Maeshiro, A. 1996. Teaching regressions with a lagged dependent variable and autocorrelated disturbances. *Journal of Economic Education*, 27(1), 72-84.

Maeshiro, A. 1999. A lagged dependent variable, autocorrelated disturbances, and unit root tests - peculiar OLS bias properties - a pedagogical note. *Applied Economics*, 31(3), 381-396.

Mandal, B., Ayyagari, P. & Gallo, W.T. 2011. Job loss and depression: The role of Subjective expectations. *Social Science & Medicine*, 72(4), 576-583.

Massenot, B. & Pettinicchi, Y. 2019. Can households see into the future? Survey evidence from the Netherlands. *Journal of Economic Behavior & Organization*, 164, 77-90.

Medvec, V.H., Madey, S.F. & Gilovich, T. 1995 When less is more: Counterfactual thinking and satisfaction among Olympic medalists. *Journal of Personality and Social Psychology*, 69, 603-610.

Meniago, C., Mukuddem-Petersen, J. & Petersen, M.A. *et al*. 2013. What cause household debt to increase in South Africa? *Economic Modelling*, 33, 482-492.

Modigliani, F. & Brumberg, R. 1954. Utility analysis and the consumption function: An Interpretation of cross-section data. In Kurihara, K.K. (ed.) *Post-Keynesian Economics*, New Brunswick: Rutgers University Press, 388–436.

Mundlak, Y. 1978. On the pooling of time series and cross section data. *Econometrica,* 46, 69-85.

Neyman, J. & Scott, E.L. 1948. Consistent estimates based on partially consistent observations. *Econometrica*, 16(1), 1-32.

Nijman, T. & Verbeek, M. 1992. Nonresponse in panel data: The impact on estimates of a life cycle consumption function. *Journal of Applied Econometrics*, 7(3), 243-257.

Niketiah-Amponsah, E., Afful-Mensah, G. & Ampaw, S. 2018. Determinants of cigarettes smoking and smoking intensity among adult males in Ghana. *BMC Public Health*, 18(1), 941.

Pastrapa, E. & Apostolopoulos, C. 2015. Estimating determinants of borrowing: Evidence from Greece. *Journal of Family & Economic Issues*, 36(2), 210-223.

Pearlin, L.I. & Radabaugh, C.W. 1976. Economic strains and the coping functions of alcohol. *American Journal of Sociology*, 82(3), 652-663.

Pogosov, I.A. & Sokolovskaya, E.A. 2016. Economic crises and unemployment in Russia. *Studies on Russian Economic Development*, 27(6), 669-675.

Puri, M. & Robinson, D.T. 2007. Optimism and Economic Choice. *Journal of Financial Economics*, 86(1), 71-99.

Quirmbach, D. & Gerry, C.J. 2016. Gender, education and Russia's tobacco epidemic: A life-course approach. *Social Science & Medicine*, 160, 54-66.

Radaev, V. 2015. Impact of a new alcohol policy on homemade alcohol consumption and sales in Russia. *Alcohol & Alcoholism*, 50(3), 365-372.

Roodman, D. 2009. How to do xtabond2: An introduction to Difference and System GMM in Stata. *The Stata Journal*, 9(1), 86-136.

Ruhm, C.J. & Black, W.E. 2002. Does drinking really decrease in bad times? *Journal of Health Economics*, 21(4), 659-678.

Sanderson, E. & Windmeijer, F. 2016. A weak instrument F-test in linear IV models with multiple endogenous variables. *Journal of Econometrics*, 190(2), 212-221.

Scheier, M.F. & Carver, C.S. 1985. Optimism, coping and health assessment and implications of generalized outcome expectancies. *Health Psychology*, 4(3), 219-247.

Scheier, M.F. & Carver, C.S. 1992. Effects of optimism on psychological and physical well-being: theoretical overview and empirical update. *Cognitive Therapy & Research*, 16(2), 201-228.

Scheier, M.F. & Carver, C.S. 2018. Dispositional optimism and physical health: A long look back, a quick look forward. *American Psychological Association*, 73(9), 1082-1094.

Semenova, M. 2011. Save or Borrow - What Determines Russian Households' Financial Strategies? *BOFIT Discussion Papers,* 28, 1-23.

Shepperd, J.A., Klein, W.M. & Waters, E.A., Weinstein, N.D. 2013. Taking stock of unrealistic optimism. *Perspectives on Psychological Science*, 8(4), 395-411.

Shepperd, J.A. & McNulty, J.K. 2002. The affective consequences of expected and unexpected outcomes. *Psychological Science*, 13(1), 85-88.

Shepperd, J.A., Ouellette, J.A. & Fernandez, J.K. 1996. Abandoning unrealistic optimism: Performance estimates and the temporal proximity of self-relevant feedback. *Journal of Personality & Social Psychology*, 70, 844-855.

Shepperd, J.A., Pogge, G. & Howell, J.L. 2017. Assessing the consequences of unrealistic optimism: challenges and recommendations. *Consciousness & Cognition*, 50, 69-78.

South African Reserve Bank (2014) *Quarterly Economic Review (December)*. Available at: https://www.resbank.co.za/Lists/News%20and%20Publications/Attachments/6532/01Full%20Quarterly%20Bulletin%20%E2%80%93%20December%202014.pdf [Accessed: 2nd Oct 2020].

South African Reserve Bank (2017) *Quarterly Economic Review (March)*. Available at: https://www.resbank.co.za/Lists/News%20and%20Publications/Attachments/7718/01Full%20Quarterly%20Bulletin%20%E2%80%93%20March%202017.pdf [Accessed: 2nd Oct 2020].

Staudigel, M. 2016. A soft pillow for hard times? Economic insecurity, food intake and body weight in Russia. *Journal of Health Economics*, 50, 198-212.

Steptoe, A., Wright, C. & Kunz-Ebrecht, S.R. *et al.* 2006 Dispositional optimism and health behaviour in community-dwelling older people: Associations with healthy ageing. *British Journal of Health Psychology*, 11, 71-84.

Stuckler, D. & Basu, S. 2013. *The body economic: Why austerity kills*. New York: Basic Books.

Swain, R. 2007. The demand and supply of credit for households. *Applied Economics*, 39(21), 2681–2692.

Teng, A., Atkinson, J. & Disney, G. *et al.* 2017. Changing smoking -mortality association over time and across social groups: National census-mortality cohort studies from 1981 to 2011. *Science Report*, 7(1): 11465.

Tobin, J. 1967. Life-cycle saving and balanced growth. In Fellner, W. *et al.* (ed.) *Ten Economic Essays in the Tradition of Irving Fisher*, London and New York: John Wiley, 231-56.

United Nation. 2019. *World economic situation prospects*. Available at: https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/WESP2019_BOOK-ANNEX-en.pdf [Accessed: 2nd Oct 2020].

Van Gundy, K., Schieman, S. & Kelley, M.S. *et al.* 2005. Gender role orientations and alcohol use among Moscow and Toronto adults. *Social Science & Medicine*, 61(11), 2317-2330.

Weinstein, N.D. 1980. Unrealistic optimism about future life events. *Journal of Personality & Social Psychology*, 39(5), 806-820.

Williams, R. 2012. Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal*, 12(2), 308-331.

Wooldridge, J.M. 2002. Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal*, 1(2), 117-139.

Wooldridge, J.M. 2010. *Econometric analysis of cross section and panel data.* 2nd ed. Cambridge, Massachusetts: The MIT Press.

Wooldridge, J.M. 2013. *Introductory econometrics: A modern approach*. 5th ed. South-Western: Cengage Learning.

Wooldridge, J.M. 2019. Correlated random effects models with unbalanced panels. *Journal of Econometrics*, 211(1), 137-150.

World Health Organization. 2016. *Global adult tobacco survey: Executive summary 2016.* Available at: https://www.who.int/tobacco/surveillance/survey/gats/gats-2016-rus-exeuctive-summary-En.pdf?ua=1 [Accessed: 2nd Oct 2020].

Wray, T.B., Dvorak, R.D. & Hsia, J.F. *et al.* 2013. Optimism and pessimism as predictors of alcohol use trajectories in adolescence. *Journal of Child & Adolescent Substance Abuse*, 22, 58-68.

Xu, X. 2013. The business cycle and health behaviors. *Social Science & Medicine*, 77, 126-136.

Yakovlev, E. 2018. Demand for alcohol consumption in Russia and its implication for mortality. *American Economic Journal: Applied Economics*, 10(1), 106-149.

Yeltsin, B. 2002. *Midnight diaries*. New York: Public Affairs.

Yilmazer, T. & DeVaney, S.A. 2005. Household debt over the life cycle. *Financial Services Reviews*, 14(4), 285-305.

Zaridze, D., Lewington, S. & Boroda, A. *et al.* 2014. Alcohol and mortality in Russia: prospective observational study of 151 000 adults. *Lancet*, 383, 1465-1473.