ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
MSc Economics & Business
Master Specialisation Financial Economics

Household discretionary saving and Locus of Control

A study examining the extent to which internal Locus of Control is a determinant of a Dutch households' discretionary saving

Author: M.M. de Jong
Student number: 412833mj
Thesis supervisor: Dr. S. van den Hauwe
Second reader: Dr. J Lemmen
Finish date: 22-3-2017

PREFACE AND ACKNOWLEDGEMENTS

This master thesis inhabits the end of nearly six and a half years study. Although this period contains its ups and downs, I am looking back at a period of fulfillment. Starting my academic career in Groningen and ending here in Rotterdam at the Erasmus University where I am honored to present my personal master dissertation.

This thesis could not be completed without some persons to whom I am very grateful. First I want to thank all the professors of the Faculty of Economics who are involved with the master Financial Economics. Especially professor Van Geen, who introduced behavioral economics to me in her seminar Nudging in cooperation with Harvard. This seminar not only brought me substance on behavioral finance but also soft skills as presenting, working in international teams and communication. Moreover, I would like to thank professor Gryglewicz for improving my skills in statistics and conducting academic research during his seminar Private Equity.

I would like to show my recognition to my thesis supervisor professor Van den Hauwe. He was involved and proactive in the thought process, something that helped me a lot. Every time I had a question or needed feedback, he provided me with on short notice, something that speeded up the writing process. His feedback was clear, to the point and very useful, this is always appreciated.

Many thanks Jozef, who helped me updating my statistics skills. With his help and knowledge on STATA my thesis really made it to the next level. Thank you for your patience and clear explanations.

By all means I am really grateful to my parents. Not only for their financial support, but moreover for the down to earth support when I was insecure about my progress and their confidence in me. It always helped me making my own decisions.

My gratitude finally goes to my boyfriend Gijs, who always supported me in times I did not see the end of the tunnel anymore. It is incredible how you manage to calm me down, and make me believe in myself again. With you by my side I was able to enjoy this period the fullest.

Rotterdam, February 2017

Marieke de Jong

ABSTRACT

Performing a 3-year cross-sectional analysis, using a unique dataset from the DNB Household survey that holds extensive Dutch household information and includes 6155 unique observations, I find that there is a positive significant effect of internal Locus of Control on a household members' discretionary savings. The results are robust after controlling for external Locus of Control, age, income, homeownership, number of members in the household, number of children in the household, degree of urbanization of the residence, whether or not someone is the most involved in the financial administration, whether or not someone is the main wage earner, and whether or not there is a partner present in the household. The findings furthermore show that age and income vary positively with discretionary savings. Pre-university- and vocational education furthermore show positive and significant effects in discretionary saving. A surprising result is the disproportionally high positive and significant effect of not having a completed education at all on discretionary savings. How these results change over time is subject to future research.

Keywords:
Locus of control, discretionary saving, behavioural finance: underlying principles, household saving, intertemporal household choice: life cycle models and saving.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

An underexplored area within economics is the influence of personality traits on saving behaviors. Locus of Control (LOC) indicates the degree to which an individual perceives a causal relationship between life experiences and its own actions. Gaining insights into this relationship may contribute to saving and consumption theories that attain high attention in macroeconomics. In this paper I test whether there is an effect of Dutch household's LOC on its discretionary savings.

Understanding consumption and saving behavior is probably one of the most prominent components in macroeconomics. Venti & Wise (1998) find that households with similar economic and demographic characteristics accumulate different amounts of wealth near retirement. They find that limits on lifetime earned financial resources or investment choices cannot explain this wealth dispersion between U.S. households. Bernheim, Skinner, Weinberg (2001) test for several factors of the life-cycle theory of saving and find no relationship between wealth accumulation and consumption growth rates. A common consumption theory is the Permanent Income Hypothesis developed by Friedman (1957), which states that current consumption depends on expected future income, i.e. permanent income. He states that permanent income is a long-term average income that depends on both physical and human capital. Dynan, Skinner & Zeldes (2004) find results that contradict Friedman's (1957) Permanent Income Hypothesis. They find a strong positive relationship between current income and saving rates across all income groups. This relationship stays positive for the whole income distribution after controlling for permanent income. Venti & Wise (1998) find roughly constant wealth-to-income ratios across income groups, which in turn supports Friedman's hypothesis. This shows that at this moment there is still finite knowledge about what is causing these differences in wealth accumulation and savings across households. From an economic perspective, whether differences in accumulated wealth between households arise from (a lack of-) available financial resources or for example self-discipline may have significant policy implications.

So far, neoclassical economic models define determinants of saving behavior such as current disposable income (Keynes, 1937), expected lifetime income (Friedman, 1957), real interest rate, subjective utility discount rate (Campbell & Mankiw, 1989), risk aversion, expected return distributions and the strength of bequest and precautionary motives (Carroll, 2001). These neoclassical determinants explain little of the variance across households that save in nearly identical circumstances, even after accounting for characteristics as income, race and education (Venti & Wise 1998, Lusardi 2003, Ameriks, Caplin & Leahy 2003, Bernheim, Skinner & Weinberg, 2001).

Common life-cycle models attribute variation in wealth and saving across households to differences in time preferences, risk tolerance, health status, exposure to uncertainty, perceived life expectancy, lifetime earnings and income replacement rates. Bernheim, Skinner & Weinberg (2001) test for the presence of these factors to explain variation in wealth and find that they are en masse incapable to do this. This lack of explanatory power emphasizes the importance of non-neoclassical determinants in explaining saving behavior.

The work of several authors is persuasive in proving that non-neoclassical influences are consequential determinants of retirement saving and related saving decisions. For example, Lusardi (2003), Ameriks, Caplin & Leahy (2003) and Behrman et al (2012), find that self-control, planning and the degree of financial literacy are positive and significantly associated with wealth accumulation. Likewise, defaults have a sizable impact on household choices such as 401(k) participation, savings rates and asset allocation (Choi et al, 2003). Laibson (1997) finds that hyperbolic discounting is associated with a low level of self-control, which creates demand for commitment opportunities like pension saving schemes.

In this paper I use a definition of LOC that is commonly shared in literature, "a generalized attitude regarding the nature of the causal relationship between one's own behavior and its consequences" (Rotter, 1966). Individuals believing life's experiences are due to external factors have an external LOC. Individuals that perceive life's experiences are due to own behavior have an internal LOC (Gatz & Karel, 1993). My focus on LOC is motivated by existing literature showing that LOC has substantial explanatory power in several economic matters, making insights on LOC economically meaningful. Heineck & Anger (2010) find that individuals with external LOC have 4 to 9% lower wages than persons with internal LOC. Moreover it predicts unemployment, life satisfaction (Becker et al, 2012) and educational achievement (Coleman & Deleire, 2003), what often is used as a proxy for permanent income. Almlund et al (2011) find that education and policy interventions that enhance LOC can lead to more successful outcomes in adulthood.

The goal of this paper is to scrutinize the explanatory power of LOC on Dutch households' discretionary saving behavior. I reach this goal by empirically examining the relationship between households' discretionary savings and LOC. As described above, an individuals' LOC has proven to be an important determinant in several behaviors and I hypothesize it similarly is an important determinant in discretionary saving. Rosenbaum (1980) finds that people with an internal LOC have more self-control. Thaler & Shefrin (1988) stress the importance of self-control as a determinant of saving rates, where a low level of self-control implies low saving rates. Based on the findings of

Rosenbaum (1980) and Thaler & Shefrin (1988) I want to test whether there is an immediate effect of LOC on savings. I hypothesize that individuals with an internal LOC are more likely to be high discretionary savers since this is associated with a higher level of self-control, and a higher level of self-control tends to increase savings rates since those individuals believe that life's outcomes are due to their own efforts. I therefore want to test whether Dutch households with a more internal LOC have higher discretionary savings.

I use data from De Nederlandsche Bank (from now on "DNB") Household Survey. My dataset contains 2011 - 2015 panel data from around 2000 representative Dutch households per year and holds detailed information about characteristics as income, wealth and psychological aspects. Total household saving exists of a part mandatory (pension) saving and a part discretionary saving (Jappelli & Modigliani, 1998). I follow Thaler & Shefrin (1988), as pension saving is done in such an automatic way it has no psychic costs. These psychic costs only occur when saving is voluntary withheld, as is the case with discretionary saving. My analysis focuses on these savings in excess of retirement savings. My research will be distinctive since to my knowledge I will be the first to link a Dutch household's discretionary savings to its LOC.

The paper proceeds as follows: chapter 2 forms the theoretical framework that includes the theoretical foundations and closely related previous research. Chapter 3 introduces the methodology and describes my contribution to economic literature. Chapter 4 presents details of the DNB Household Survey data and the process of data preparation. Chapter 5 contains the results of the analysis. Chapter 6 shows the conclusion, limitations and recommendations for future research.

CHAPTER 2 THEORETICAL FRAMEWORK

To place my research in perspective, I now discuss the used theoretical models for describing saving and consumption behavior and describe how my research relates to these models. First I describe shortly the Life Cycle and Permanent Income hypotheses, which are theory-based consumption models and still widely used. In the next section I describe the more recent Behavioral Life Cycle Hypothesis and how my research relates to this model. In the last part I outline the closely related theoretical constructs used to form my research and the section ends with reporting my contribution to economic literature.

2.1 Life Cycle / Permanent Income Hypothesis

The Life Cycle Hypothesis (LCH) (Modigliani & Brumberg 1954) and Permanent Income Hypothesis (PIH) (Friedman, 1957) are forward looking theories of consumption. In both models consumption decisions are based on maximizing utility by choosing the optimal consumption level given their preferences and current disposable income as well as wealth accumulation now and in the future. Whether the model assumes resources to be in the form of permanent income following the PIH or total wealth following the LCH, they both are based on several simplifying assumptions. Assumptions include the price level that does not change over time, an individual does not inherit any financial resources at any point of its life cycle, interest rates are zero and the proportion of resources to consume only is determined by taste, not by the size of the resources. Lastly, theoretically the individual should consume its income at an even rate throughout his life. These assumptions make the models differ significant from real life situations, this in turn lowers reliability and practical use. As described in the introduction, research shows that these models can explain little of the wealth dispersion across households (Bernheim, Skinner & Weinberg 2001, Venti & Wise 1998, Dynan, Skinner & Zeldes 2004).

2.2 Behavioural Life Cycle Hypothesis

Shefrin & Thaler (1988) incorporate three important behavioral attitudes, self-control, mental accounting and framing, to enhance the fit with economic data and propose the Behavioral Life Cycle Hypothesis (BLC). The first, self-control, is positively related with LOC (Rosenbaum, 1980) and that is the reason why I only elaborate on this behavioral attitude of the BLC.

Self-control distinguishes ordinary choice by three elements usually excluded in economic analysis: internal conflict, temptation and willpower (Shefrin & Thaler, 1988). Some situations are more tempting than others; this leads to an internal conflict between immediate spending and long-term

benefits. Willpower represents the psychic costs of resisting this temptation. In the BLC, individuals are assumed to behave as if they have dual preferences that are inconsistent but coexist. One concerned with the long run, the planner, and the other concerned with the short run, the doer. The planner and the doer have contrasting time horizons, which describes the internal conflict of self-control. For discretionary saving some willpower is needed to overcome tempting situations in which the doer wants to maximize current utility and therefore some self-control is needed. Total doer utility (Z_t) is described by:

$$Z_t = U_t + W_t$$

Where U_t is marginal utility at time t and W_t is the planner's psychic cost of using willpower. The total doer utility is thus the sum of pleasure and pain. The authors also adopt mental accounting into the BLC that contradicts the fungibility assumption of traditional economic saving and consumption models. According to Shefrin & Thaler (1988) individuals treat wealth as non-fungible and use three mental accounts, with different degrees of temptation: 1) Current income (*I*) 2) Current assets (*A*) and 3) Future income (*F*). The temptation to spend is the highest in the current income account, where the highest amount of willpower ('pain') is needed to save. Temptation (and thus willpower and required self-control) is assumed to be lower in the *A* account, and lowest in the *F* account. Figure 1 shows the doer utility Z_t against the level of consumption C_t, given an assumed structure of mental accounts and its balances.

By assumption willpower diminishes as consumption increases. As shown in figure 1, a reduction in willpower contributes to higher doer utility. If one consumes the entire balance in account *I*, no willpower effort is needed. If one consumes a next unit, this will be financed out of the *A* account, which is by assumption less tempting then the *I* account. This implies again less needed willpower and thus higher doer utility. The authors state that consumption from the *A* account results in a disutility penalty and thus the first unit is extremely costly. Similar argument applies when the *F* account is used.

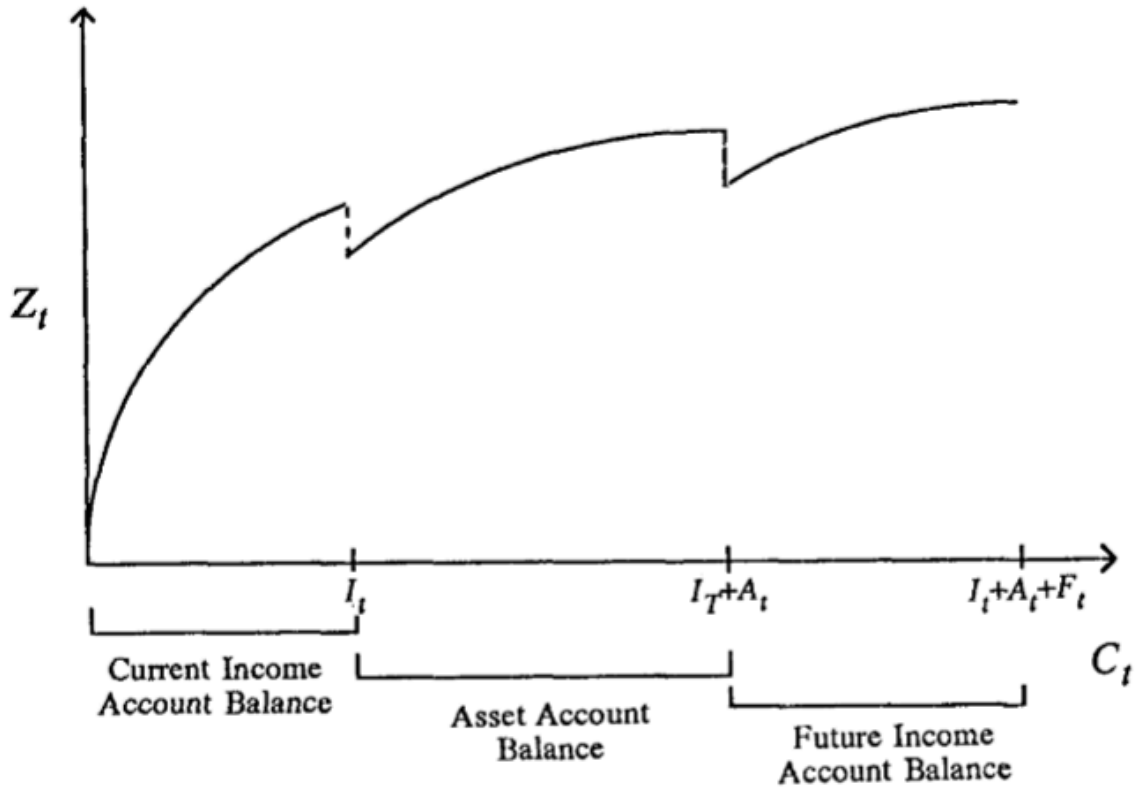

Figure 1. Doer utility utility in the Behavioural Life Cycle Theory. The figure shows the doer utility Z_t against consumption C_t for the given mental accounting structure used in the BLC. The temptation to spend is assumed to be the highest in the current income account balance; therefor the most willpower is needed to resist this temptation. This results in a low yet increasing doer utility. The temptation to spend is assumed to be lower in the Asset Account Balance and consequently again lower in the Future Income Account Balance. From Shefrin & Thaler (1988).

Internal LOC is positively related with more self-control (Rosenbaum, 1980) and this is needed for exerting willpower to resist temptation and to save for life cycle purposes (Shefrin & Thaler, 1988). Exploring the immediate relation of LOC on discretionary saving might attribute to fit the BLC to microeconomic data.

2.3 Saving behaviour and the influence of Locus of Control

Whether someone will regard a life experience as a consequence of own behavior or not will differ per individual. One of the determinants of this attitude is the degree to which an individual perceives that consequence is due to its own behavior versus the degree to which he feels the consequence is controlled by external factors and may occur regardless of his own actions (Rotter, 1966). As mentioned in the introduction, in this paper I use a definition of LOC that is commonly shared in

literature, “a generalized attitude regarding the nature of the causal relationship between one’s own behavior and its consequences” (Rotter, 1966) and individuals believing life’s experiences are due to external factors have an external LOC. Individuals that perceive life’s experiences are due to own behavior have an internal LOC (Gatz & Karel, 1993).

An important aspect to discuss about LOC is the tension in previous literature about potential endogeneity. Some research points out that LOC may not be strictly exogenous and may instead respond to individuals’ life experiences. Research focuses on the possibility that LOC is partly determined by educational, health or labor outcomes (Cobb-Clark & Schurer 2013). For example, Goldsmith, Veum and Darity (1996) find that people who are more often shocked by unexpected life experiences are more likely to adjust their beliefs about how much they control their lives. Seligman (1975) notes that the more uncontrollable an event is regarded to be, the more it would lead to a feeling of less control over life. In the more recent study, Cobb-Clark & Schurer (2013) find however that an individuals’ LOC essentially does not change over their timespan of four years and nine negative life events. They find a small tendency of positive life events such as experiencing a promotion or other significant improvements in their finances to provide directions for more internal LOC. Overall, Cobb-Clark & Schurer (2013) do not find life experiences that have a significant predictive power in explaining changes in LOC. Suited to this latest work of Cobb-Clark & Schurer (2013) and for reasons of restrained research time, I assume LOC to be strict exogenous.

An individuals’ LOC has proven to be a determinant of its learning process and attributes to a persons’ self-control (Rosenbaum 1980). Rosenbaum (1980) designs a Self Control Schedule and finds individuals that score high on this self-control scale have an internal LOC, hold fewer irrational beliefs and are better able to control harmful stimuli. The author reports moreover that persons that score high on this scale also have a higher perceived self-efficacy. Self-efficacy contributes to the conviction that one can successfully execute behaviors to produce a desired outcome, which is related to saving an important aspect. The Dutch government as well as non-governmental organizations as Nibud target self-efficacy and finding out how a person’s LOC attributes to saving behavior and implicitly to self-control and self-efficacy would be of interest for designing policy interventions.

Laibson, Repetto & Tobacman (1998) investigate how self-control affects saving for retirement. Economists tend to overestimate the sophistication in decision-making and assume that intentions and actions are aligned. The authors conclude that this is incorrect and that failures arise in problems involving delayed gratification. These findings imply that self-control is costly, i.e. willpower is needed, and thus rules are needed to resist temptation of immediate gratification. In the research of

these authors, high level of self-control is positively related to pension saving, as this helps to control the delayed gratification decisions.

Abay, Berhane & Assefa (2016) study individuals' discounting behavior in Ethiopia and find that LOC predicts hyperbolic preferences. They conduct a Randomized Control Trial (RCT) with 2400 households where respondents were asked to choose between X amount today and X+100 after a month and find that " … [farmers] with an external LOC are more likely to discount future payoffs heavily if the alternative payoff is immediate". Thus in case of external LOC, a lack of commitment instruments may lead to low saving rates (Laibson, 1997). Abay, Berhane & Assefa (2016) find this behavioral explanation for improving saving rates in developing countries and find that LOC affects saving behavior, where internal LOC is associated with high savers since they show lower hyperbolic preferences. This evidence from rural Ethiopia is a foundation for my paper to expect this saving behavior in Western economic circumstances too.

Kimball & Shumway (2009) identify preliminary correlations between whether a survey respondent is a high or low retirement saver and some individual attitudes using the Surveys of Consumers with 325 selected U.S. adults. They find that individuals having an external LOC had a significantly lower propensity to save than people with an internal LOC. The authors use statements to measure LOC such as "No one can predict the future, so trying to save doesn't do much good" and "Many of the things that keep me from saving more money are out of my control". These preliminary findings for LOC being a determinant in retirement saving of U.S. households forms a direction for my paper to test whether this LOC is a determinant of Dutch discretionary saving.

CHAPTER 3 METHODOLOGY

In this section I describe my contribution to economic literature and research approach. The section starts by stating my contribution and introducing my empirical model. Section 3.2 constructs my dependent and independent variables. Additionally, based on existing literature I will determine relevant control variables in my research and explain in section 3.3 how I deal with the expected standard errors. All statistical analyses are performed using STATA.

3.1 My contribution and empirical model

This research is to my knowledge the first to link LOC directly to a Dutch households' discretionary savings rate. This newly formed direct link contains two new elements. The first is the focus on the Dutch population with new data of the DNB Household Survey. My sample contains 2015, 2013 and 2011 data collected in yearly waves, which is more recent then existing research has ever performed. The data contains detailed information on household finances and personal preferences. Second, my analysis will concentrate on savings in excess of retirement saving. As Jappelli & Modigliani (1998) formulate, total household saving exists of a part mandatory (pension) saving and a part active 'discretionary' saving. Following Thaler & Shefrin (1981), pension saving is done in such an automatic way that it has no psychic costs so no willpower is needed. These psychic costs only occur when saving is voluntarily withheld, as is the case with discretionary saving. So far, only preliminary findings, since their work is not published, of LOC on retirement savings are presented by Kimball & Shumway (2009). My research furthermore uses a new identification strategy by regressing a households' discretionary saving over its degree of LOC, conditioning for income, education, age and homeownership where the measure for internal and external LOC is formed by factor analysis.

Cross-sectional analysis of the relationship between the savings and LOC will be performed for three years using the following OLS regression:

$$s_i = \alpha + \beta_1 il_i + \beta_2 el_i + \beta_3 y_i + \beta_4 g_i + \beta_5 e_i + \beta_6 h_i + \sum_{i=7}^{k} \beta_i z_i + \varepsilon_i$$

Where:

s_i is the measure for discretionary savings displayed in 2015 euros;

il_i is the measure of internal locus of control constructed by factor analysis;

el_i is the measure of external locus of control constructed by factor analysis;

y_i is the measure for current disposable income displayed in 2015 euros;
g_i is the measure for age, calculated at the moment of filling in the survey;
e_i is the measure for degree of completed education;
h_i is a dummy for homeownership, 1 indicating that someone owns a house, 0 indicating rental housing;
ε_i is the error term that is included to account for measurement errors.

Both measures for LOC are unitless cardinal measures. I expect that discretionary savings will vary positively with the degree of internal LOC. Additional covariates ($\sum_{i=7}^{k} \beta_i z_i$) are added to collect information on the relation between household characteristics and discretionary savings. Extra covariates are the number of household members *(aantalhh),* number of children in the household *(aantalki),* degree of urbanization *(sted)* where 1 indicates very high degree of urbanization and 5 very low degree of urbanization, the composition of the household *(woonvorm),* a dummy for whether or not someone is the most involved in the financial administration *(account),* a dummy for whether or not someone is the main wage earner *(kostwin),* and a dummy for whether or not they have a partner present in the household *(partner).*

3.2 Dependent and independent variables

I use discretionary savings as a measure of saving behavior only focusing on the aspects of the extent to which one is a high or low saver. This excludes the asset allocation and the influence of risk aversion in my analysis. Between brackets I show the name of the variables used in my dataset.

My dependent variable will be the discretionary savings. In constructing this measure, I follow the approach of Dynan, Skinner & Zeldes (2004) who measure savings as the net worth in year *t.* My measure of net worth excludes values of life insurances and the value of defined-contribution pension plans. This focuses on the savings that are not consumed and actively put aside in a given period: discretionary saving. Due to a lack of data I cannot eliminate unrealized capital gains or losses on primary homes. Discretionary savings for household *i* is defined as:

$$S_{it} = A_t$$

Where A_t is the net worth that is defined as the sum of financial assets less debt at the end of year *t*. Financial assets include bank accounts (*bet141, bet142, bet143, bet 144 & spa131, spa132, spa133, spa134*), total market value of investments trough mutual funds (*bel101, bel102, bel103*), all shares

(*aan601, aan602, aan603*), bonds (*obl3*), put-options (*opt4*) and call-options (*opt8*). Debt includes remaining debt in private loans (*per301, door301, fin201, ps101*), loans from family (*fam101*), outstanding credit card debt (*cred2*), study loans (*stud101*) and any other type of loans not mentioned earlier (*and201*). Mortgage is excluded from this measure since no data is available on current market value of primary residence.

If one does not know or did not enter any exact value for the mentioned variables, one is asked to indicate into which category the mentioned variable is situated. DNB handles these values as adding the average of the range to the calculations. Since DNB is an appreciated organization, I choose to follow their approach and generate discretionary savings by adding the average of the category indicated. This is applicable for the following asset variables: *bet151, bet152, spa141, spa142, bel121, bel122, aan801, aan802, obl5, opt4a, opt8a* and for the following debt variables: *per401, doo401, fin301, ps201, fam201, stu201, cred3, and301*. An overview of these categories can be found in figure A in the appendix. In the highest category the lower bound value is used in calculations.

To form my main independent variable of interest, I follow the construct that, according to DNB, measures LOC. This includes thirteen variables of which each respondent marks every statement to the extent the statement holds true on a scale from one to seven. The variables included are locus1 t/m locus13 and have labels such as "It is chiefly a matter of fate whether I become rich or poor" and "Whether or not I become wealthy depends mostly on my ability". An overview of variables included and their labels are described in table 1.

Table 1. Overview of Locus of Control variables and variable labels

variable	*variable label*
locus01	Saving and careful investing is a key factor in becoming rich (Internal)
locus02	Whether or not I get to become wealthy depends mostly on my ability (Internal)
locus03	In the long run, people who take very goodcare of their finances stay wealthy (Internal)
locus04	If I become poor, it's usually my own fault (Internal)
locus05	I am usually able to protect my personal interests (Internal)
locus06	When I get what I want, it's usually because I worked hard for it (Internal)
locus07	My life is determined by my own actions (Internal)
locus08	There is little one can do to prevent poverty (External)

locus09	Becoming rich has nothing to do with luck (Internal)
locus10	Regarding money, there isn't much you can do for yourself when you are poor (External)
locus11	It's not always wise for me to save because many things turn out to be a matter of luck (External)
locus12	It is chiefly a matter of fate whether I become rich or poor (External)
locus13	Only those who inherit or win money can possible become rich (External)

Table 1. Variables and corresponding labels measuring LOC. Internal or external between brackets indicates the direction to which the labels are formulated. Each correspondent marks every statement to the extent the statement holds true on a scale from one to seven. Used from the DNB Household Survey.

As an individuals' degree of internal or external LOC cannot be asked for explicitly, I expect there to be underlying factors in the thirteen separate variables that measure LOC. A table of coefficients will be constructed to show correlation within the variable set. I expect there to be two underlying factors, as some variable labels are directed towards internal LOC and other labels are directed externally. A factor analysis will be performed and factors that satisfy the Kaiser-Criterion with Eigenvalues higher than one will be retained. I will perform an additional check whether my data is suitable for factor analysis, performing a Kaiser-Meyer-Olkin measure for adequacy. This measure provides an indication of how much variance each factor can explain separate variables.

3.3 Control variables and errors

I include several control variables to get a more unbiased estimation of $ß_1$. Dynan, Skinner & Zeldes (2004) find a strong positive association between savings rates and current income, and a weaker but still significant higher marginal propensity to save for higher-income households than for low-income households. They find that savings rates increase across the entire income distribution. To control for this discrepancy I include disposable income (y) as a control variable, which is the total net labor income of the household member over year t, measured by variable *in18* and *in20*. If one does not know or did not enter any exact value in variable *in18*, one is asked to indicate into which category the yearly income is situated. An overview of these categories and values used to transform the variable are discussed in figures 2 and 3 in the results section. The log of income is taken to create more normally distributed errors.

Jappelli & Modigliani (1998) provide an age-saving profile that strongly contradicts the Life-Cycle Hypothesis. This is supported by their micro-level data and shows that savings rates differ per age. Following their findings, I include age (*g)* as a control variable in my regression, measured by the variable *age*. This variable measures the age of individuals measured at the time of filling in the questionnaire. Jappelli & Modigliani (1998) also indicate that education and homeownership are positively related with saving. Kain & Quigley (1962) state that savings vary positively with homeownership due to an increase in value of favorable tax shields and savings for annual maintenance expenditures. The formation of tax shields is also applicable in the Netherlands[1] on the condition that the household is repaying the mortgage. Optimally using my available data, I include a dummy for homeownership (*h*) (*woning*) and condition for completed education (*e)* (*oplmet*).

Additional covariates are added to collect information on the relation between household characteristics and discretionary savings. Extra covariates are the number of household members *(aantalhh),* number of children in the household *(aantalki),* degree of urbanization *(sted),* composition of the household *(woonvorm),* a dummy for whether or not someone is the most involved in the financial administration *(account),* a dummy for whether or not someone is the main wage earner *(kostwin),* and a dummy for whether or not they have a partner present in the household *(partner).*

The ε_i is assumed to be the independent and identically distributed error. However, in practice the errors might turn out to be heteroscedastic. This will be tested using the White test. The advantage of the White test is that it does not require any specific assumptions about the form of heteroscedasticity. Although it is true that this advantage comes at the cost of bit lower power compared to other tests, this will not be very problematic in this case since the database used here has 6155 observations. If the errors will be shown heteroscedastic I will re-estimate them using the White heteroscedasticity consistent errors.

3.4 Assumptions

For this research I assume that the DNB Household Survey (see next section) is representative to the Dutch population as this is a goal for the DNB itself, and by executing this panel for more than 20 years now I assume that this indeed is true. I furthermore assume that the respondents filled in the survey independently. Lastly, as described in the literature section, I assume LOC to be strictly exogenous, that is, it does not respond to individuals' life experiences. This is in line with the latest literature as discussed in the literature section.

[1] See article 3:120 Dutch personal income tax act.

CHAPTER 4 DATA

This section first describes the dataset used in detail and the process of data preparation follows. As described in the methodology section, some variables are transformed from categorical variables to continuous values. The section ends describing the influence of missing values on the validity of my results.

4.1 Data description

In my research I use the De Nederlandsche Bank (DNB) Household Survey panel data, which is a panel (CentERdata panel) that since 1993 yearly collects economic data with the purpose of studying the economic and psychological determinants of saving behavior of Dutch households. This is a probability-based panel and to ensure the panel reflects a good representation of the Dutch population, people who do not have a computer were given the use of a computer and Internet access. Since this panel data is widely used for research purposes and contains around 2000 observations per year, I consider my sample set to be representative to the Dutch population.

The DNB Household Survey data is available to researchers on specific terms. The questionnaire is presented to all household members with age above sixteen. My focus on LOC limits the data to the latest available years. DNB introduces the LOC measures in 2009, however they have all missing values in the 2009 dataset. Since 2009, DNB only runs the LOC measures every odd year. After having contact with Senior Researcher Miquelle Marchand from the CentERdata research department I have been provided two reasons. The reasons consist of a combination of cost savings and the fact that the DNB expects LOC not to change considerable within the timespan of one year. This results in three useful years 2015 (5133 observations), 2013 (4952 observations) and 2011 (4173 observations). Excluding missing values for all relevant variables (see next section) results in 2370 observations for 2015, 1886 observations for 2013 and 1899 observations for 2011.

All euro values are adjusted for inflation by CPI index = 100 in 2015, rescaled from originally 1900 accessed from the Central Bureau of Statistics (CBS) to show real instead of nominal changes.

Summary statistics of all LOC variables in the DNB dataset, shown in table 2, indicate that means of all LOC variables are relatively stable over the three years. The standard deviations also stay quite constant. Mean values showing low variability suggest that someone's LOC does not change considerably over my total timespan of 5 years. Scatterplots confirm this by showing that both internal and external LOC is spread evenly trough the complete sample set.

Table 2. Summary statistics for Locus of Control variables per year

Variable	2015	2013	2011
locus01	4,130 (1,537)	4,332 (1,524)	4,310 (1,574)
locus02	3,726 (1,456)	3,820 (1,502	3,751 (1,540)
locus03	4,283 (1,409)	4,362 (1,359)	4,270 (1,445)
locus04	3,526 (1,588)	3,531 (1,597)	3,703 (1,641)
locus05	5,190 (1,096)	5,142 (1,085)	5,264 (1,106)
locus06	4,986 (1,273)	4,955 (1,226)	5,052 (1,236)
locus07	4,841 (1,288)	4,874 (1,290)	4,939 (1,278)
locus08	3,312 (1,402)	3,227 (1,380)	3,177 (1,412)
locus09	4,057 (1,630)	4,009 (1,558)	4,223 (1,623)
locus10	3,632 (1,476)	3,667 (1,476)	3,670 (1,499)
locus11	3,066 (1,402)	2,940 (1,374)	2,977 (1,447)
locus12	3,105 (1,476)	2,989 1,476)	3,057 (1,499)
locus13	2,937 (1,552)	2,842 (1,584)	2,984 (1,660)
Nr of observations	2.370	1.886	1.899

Table 2. The graph shows summary statistics of all LOC variables over the years 2015, 2013 and 2011 respectively. The columns show mean values and standard deviations between brackets. The total number of observations is displayed in the last row.

Table 3 shows summary statistics of the constructed variables used in the cross-sectional analysis. The table ends with the selected characteristics of the survey respondents. Notable is the high standard deviation for the median value of savings. This indicates that there is a large variability of savings within all survey respondents. Medians are taken for savings and income to account better for outliers. Median income is reasonably consistent over the years. Mean age is high in all years for filling in the surveys. As shown in column 2 to 4, the mean internal LOC varies more over the years than the mean of external LOC. Of all survey respondents, little more than 80 percent owns a house.

Table 3. Summary statistics for dependent and independent variables included in the cross-sectional analysis.

Variable	2015	2013	2011
savings	6.625 (80.300)	4.268 (62.716)	7.365 (90.406)
internal locus of control	8,67 (3,95)	10,38 (3,94)	8,67 (3,77)
external locus of control	15,15 (2,58)	14,93 (2,68)	15,75 (2,78)
age	53,43 (17,91)	52,46 (17,96)	54,05 (16,74)
income	31.374 (27.997)	32.968 (29.821)	32.226 (29.824)
homeownership	0,82 (0,38)	0,82 (0,39)	0,81 (0,39)
education	3,80 (2,61)	4,05 (2,63)	3,85 (2,68)
household members	2,66 (1,25)	2,55 (1,18)	2,51 (1,16)
nr of children in household	0,77 (1,09)	0,67 (1,03)	0,65 (1,00)
degree of urbanization	3,10 (1,27)	3,06 (1,29)	3,03 (1,29)
composition of household	2,38 (0,85)	2,33 (0,86)	2,32 (0,87)
most involved financial admin	0,52 (0,50)	0,53 (0,50)	0,53 (0,50)
main wage earner	0,52 (0,50)	0,53 (0,50)	0,53 (0,50)
partner present in household	0,84 (0,37)	0,83 (0,37)	0,82 (0,38)

Table 3. Summary statistics for all variables included in the cross-sectional analysis. The table shows median values for the variables savings and income accompanied by the standard deviation between brackets. Medians are shown since these variables are continuous and this provides a more reliable outcome than means. For all other variables mean values are shown, with the standard deviation between brackets.

4.2 Data preparation

In the following section I describe the preparation of the dataset that includes the transformation of categorical variables and my considerations according to the influence of missing data on validity of my results.

4.2.1 Categorical variables

Since the database contains categorical variables that are not suited for cross-sectional analysis, I transform these variables to numerical values. The DNB itself uses this method for their research purposes for years, indicating it is a transparent, effective way to minimize the bias it may create.

All categorical variables used in my OLS regression need to be transposed. This is mostly applicable to variables forming ones financial assets and financial liabilities. First, one is asked to indicate an exact number, e.g. what is the amount in euros on date X on your first savings account. Then, if one does not indicate an exact number here, the follow up question is to indicate into which of the categories one's balance go. The categories for asset and liability variables are displayed in figure 2 and the categories for the income variable in figure 3. Every time a category is indicated, the average of this category is further used to calculate assets and liabilities. At the highest category, the lower bound is the transformed value. Since a very low amount of transformations took place into this segment, this bias that might be created using this method will be minimal.

Table 4 shows the exact amount and percentage of transformed observations. As shown, this transformation allows me to use more valuable information on discretionary savings and other characteristics. Since average transformation percentages are low, I expect no considerable influence on the validity of my results.

Overview of methodology used to transform categorical variables to form savings rates.

Into which of the categories mentioned below did the balance of your account go on 31 December year X?	To transform the categorical variables the following table is used, according to DNB methodology:
1. Less than 50 Euro	1. 25 Euro
2. Between 50 Euro and 250 Euro	2. 150 Euro
3. Between 250 Euro and 500 Euro	3. 375 Euro
4. Between 500 Euro and 750 Euro	4. 625 Euro
5. Between 750 Euro and 1.000 Euro	5. 875 Euro
6. Between 1.000 Euro and 2.500 Euro	6. 1750 Euro
7. Between 2.500 Euro and 5.000 Euro	7. 3750 Euro
8. Between 5.000 Euro and 7.500 Euro	8. 6250 Euro
9. Between 7.500 Euro and 10.000 Euro	9. 8750 Euro
10. Between 10.000 Euro and 11.500 Euro	10. 10.750 Euro
11. Between 11.500 Euro and 14.000 Euro	11. 12.750 Euro
12. Between 14.000 Euro and 17.000 Euro	12. 15.500 Euro
13. Between 17.000 Euro and 20.000 Euro	13. 18.500 Euro
14. Between 20.000 Euro and 25.000 Euro	14. 22.500 Euro
15. 25.000 Euro or more	15. 25.000 Euro
-9 Don't know	

Figure 2. Overview of methodology used to transform categorical variables. Retrieved from the DNB 2015 Codebook. This is applicable for the following variables: per401, doo401, fin301, ps201, fam201, stu201, cred3, and301 (debt variables) and bet151, bet152, bet153, bet154, spa141, spa142, spa143, spa144, bel121, bel122, bel143, aan801, aan802, aan803, aan803, aan804, obl5, opt4a, opt8a (asset variables), for all years.

Overview of methodology to transform the categorical income variable.

Can you give an estimation of your taxable income for year X?	To transform the categorical variables the following table is used, according to DNB methodology:
1. Less than 2.500 Euro	1. 1.250 Euro
2. Between 2.500 Euro and 5.000 Euro	2. 3.750 Euro
3. Between 5.000 Euro and 10.000 Euro	3. 7.500 Euro
4. Between 10.000 Euro and 15.000 Euro	4. 12.500 Euro
5. Between 15.000 Euro and 20.000 Euro	5. 17.500 Euro
6. Between 20.000 Euro and 30.000 Euro	6. 25.000 Euro
7. Between 30.000 Euro and 40.000 Euro	7. 35.000 Euro
8. Between 40.000 Euro and 50.000 Euro	8. 45.000 Euro
9. Between 50.000 Euro and 75.000 Euro	9. 62.500 Euro
10. More than 75.000 Euro	10. 75.000 Euro
-9 Don't know	

Figure 3. Overview of methodology to transform the income variable in20 to form a measure for income. This is retrieved from the DNB 2015 codebook. This is only applicable to the variable *in20* and for all years.

Table 4. Results of transformation of income, asset and liability variables

Year	2011		2013		2015	
Variable name	Observations	Percentage	Observations	Percentage	Observations	Percentage
in20	877	21,0%	1232	24,9%	1428	27,8%
per401	17	0,4%	29	0,6%	31	0,6%
doo401	54	1,3%	36	0,7%	43	0,8%
fin301	4	0,1%	5	0,1%	11	0,2%
ps201	4	0,1%	3	0,1%	10	0,2%
fam201	14	0,3%	5	0,1%	27	0,5%
stu201	16	0,4%	60	1,2%	20	0,4%
cred3	20	0,5%	15	0,3%	32	0,6%
and301	2	0,1%	6	0,1%	14	0,3%
bet151	744	17,8%	930	18,8%	1126	21,9%
bet152	352	8,4%	476	9,6%	612	11,9%
bet153	90	2,2%	108	2,2%	157	3,1%
bet154	15	0,4%	19	0,4%	34	0,7%
spa141	474	11,4%	185	3,7%	639	12,4%
spa142	207	5,0%	285	5,8%	362	7,1%
spa143	75	1,8%	77	1,6%	122	2,4%
spa144	32	0,8%	34	0,7%	40	0,8%
bel121	98	2,3%	66	1,3%	89	1,7%
bel122	48	1,2%	31	0,6%	32	0,6%
bel123	9	0,2%	10	0,2%	9	0,2%
aan801	60	1,4%	47	0,9%	67	1,3%
aan802	26	0,6%	34	0,7%	35	0,7%
aan803	17	0,4%	20	0,4%	25	0,5%
obl5	19	0,5%	13	0,3%	12	0,2%
opt4a	2	0,1%	3	0,1%	3	0,1%
opt8a	1	0,0%	0	0,0%	3	0,1%

Table 4. This table describes the performed transformations in income, liability and asset variables used in the cross-sectional analysis. The horizontal dotted line separates the categories respectively. The first column shows all variables where transformation was necessary. For details on how and why, see the methodology section. Then, the next column shows the absolute number of observations that is transformed, followed by the percentage of the total observations included in the indicated year.

4.2.2 Missing observations

The impact of missing data on the validity of research findings depends on the mechanisms that led to the missing data, the patterns and the proportion of the data missing, where the pattern of the missing data has greater influence on research results than the proportion of data missing (Tabachnick & Fidell, 2001, p58). In the following section I will analyze for patterns and proportions of missing data.

This will be done by using a mean comparison tests, the independent two-sample two-sided t-tests in STATA, between the observations with complete data and data with missing values following Kim & Curry (1977). These insights will provide information about the pattern of missing data and on the validity of my results. This analysis is executed for all independent variables: locus of control, age, income, homeownership and education.

Locus of control

For year 2015 executing a t-test between the complete observations and the observation with missing loci variables results in a significant difference in mean savings. The group with missing loci variables (326 observations, 6,4%) has significant lower discretionary savings with a p-value of 0,0001 that indicates a significant mean difference at the 1% level.

In 2013, there are 726 (14,7%) missing observations. Executing another t-test between mean savings of missing loci observations and the rest of the sample results in a p-value very close to zero that indicates a significant mean difference at the 1% level. Household members that did not indicate the LOC variables have again significantly lower discretionary savings than the rest of the sample.

For year 2011 there are 423 (10,1%) missing observations. Executing the same t-test between mean savings of missing LOC observations and the rest of the sample results in a p-value of 0,0003 that indicates a significant mean difference at the 1% level. Household members that did not indicate the LOC variables have significantly lower discretionary savings than the rest of the sample.
These results imply that excluding the aforementioned 326 observations in 2015, 726 observations in 2013 and 423 observations in 2011 can lead to a selection bias since the excluded households have on average lower discretionary savings. This might bias the estimated effect of LOC a bit upwards, as this disproportionally excludes households with low savings.

Age

In STATA I generate the variable age derived from year of birth, which represents the age at the time of filling in the questionnaire. The DNB presents the questionnaire to household members with age > 16 only. For this reason, I check the impact of missing values and the impact of excluding observations with age < 16. These values might indicate a human error in filling in year of birth, or these household members are not yet financially autonomous.

For year 2015 this generates 3 (0,06%) missing values and 948 (18,5%) values with age < 16. The three missing age values are excluded without any consequences for representativeness. Executing a two-sample two-sided t-test between the mean discretionary savings of observations with age < 16 and the rest of the sample results in a p-value very close to zero, that indicates a significant difference from the sample mean at the 1% level. Household members with age < 16 have significant lower mean discretionary savings than household members with age > 16.

In 2013 there are 494 (10%) missing values and 419 (8,5%) values with age < 16. Executing a t-test between mean savings of missing age observations and the rest of the sample results in a p-value that indicates a significant difference from the sample mean at the 1% level. Household members that did not indicate their year of birth have significantly lower discretionary savings than the rest of the sample. Executing a second t-test between mean discretionary savings of observations with age < 16 and the rest of the sample results in a p-value that indicates a significant difference from the sample mean at the 1% level. Household members with age < 16 have significantly lower discretionary savings than the rest of the sample.

For year 2011 this generates 190 (4,6%) missing values and 490 (11,7%) values with age < 16. Executing the same t-test between mean discretionary savings of missing age observations and the rest of the sample results in a p-value of 0,1559 that indicates an insignificant mean difference. Executing a t-test between the mean discretionary savings of observations with age <16 and the rest of the sample results in a p-value of 0,0023 that indicates a significant difference from the sample mean at the 1% level. Household members with age < 16 have significantly lower discretionary savings than the rest of the sample. The observations with age < 16 have considerable substance to create a bias into the results. A robustness check will be performed with an additional dummy for underage saving to investigate this further.

Income

Income is the control variable measured by the variable *in18*. This variable measures the yearly total net income of the household member and is adjusted for inflation, displayed in 2015 euros.

The income variable holds 1493 (29%) missing observations and 583 (11,4%) 'don't know' observations for 2015. Executing a t-test between the mean discretionary savings of observations with missing income observations compared to complete income observations results in a p-value that indicates a significant difference from the sample mean at the 1% level. Household members with missing income data have significantly lower discretionary savings than household members with complete income data. Executing a t-test between the mean discretionary savings of 'don't know' observations and the complete income observations results in a p-value of 0,0495 that indicates a significant difference from the sample mean at the 5% level. Household members that do not know

their income have significantly lower discretionary savings than household members with complete income data.

For year 2013 the income variable has 2010 (40,6%) missing observations and 377 (7,6%) 'don't know' observations. The analysis results in a significant mean difference at the 1% level. Household members with missing income data have significant lower discretionary savings than household members with complete income data. Executing a t-test between the mean discretionary savings of 'don't know' observations and complete income observations results in a p-value of 0,0191 that indicates a significant difference from the sample mean at the 5% level. Household members with 'don't know' income values have significantly lower discretionary savings than household members with complete income data.

In 2011 the income variable has 1521 (36,4%) missing observations and 255 (6,1%) 'don't know' observations. Executing a t-test between the mean discretionary savings of observations with missing income observations compared to complete income observations results in a p-value that indicates a significant difference from the sample mean at the 1% level. Household members with missing income data have significant lower discretionary savings than household members with complete income data. Executing a t-test between the mean discretionary savings of 'don't know' observations and the complete income observations results in a p-value of 0,0234 that indicates a significant difference from the sample mean at the 5% level. Household members that do not know their income have significantly lower discretionary savings than household members with complete income data.

The results of these t-tests imply that excluding the aforementioned observations in all years can lead to a selection bias since the excluded households have on average lower discretionary savings. This might bias the estimated effect of LOC a bit upwards, as this disproportionally excludes households with low discretionary savings.

Homeownership

The control for homeownership is measured by the variable *woning* and indicates the type of accommodation for the household member.

In 2015, 4075 household members (79,4%) own a house, 1002 household members (19,5%) rent a house, and 56 (1,1%) has another type of housing which includes unknown, free living and sub-rented housing. No observations are missing. As I am only interested in renting versus owning a house, I delete these 56 observations and it is not likely to have any consequences on validity of my results.

Again, in 2013 there are no missing observations and 3933 (79,4%) household members own a house, 1012 (20,4%) household members rent a house and 7 (0,1%) has another type of housing.

No missing observations are present in 2011, where 3384 (81,1%) household members own a house, 781 (18,7%) household members rent a house and 8 (0,2%) has another type of housing. Excluding these 7 observations in 2013 and 8 observations in 2011 is not likely to have any consequences on validity of my results.

Education

The covariate for education is measured by the variable *oplmet* and indicates the highest level of education completed. All common education options are mentioned. The most common education completed in the Netherlands is havo[2], thus this will be used as baseline in my analysis.

In 2015 *oplmet* contains 31 (0,6%) missing observations and 38 (0,7%) observations that indicate 'other'. Executing a first t-test results in a p-value of 0,2731 that indicates an insignificant mean difference for the missing observations, whereas executing a second t-test for the observations that indicate 'other' results in a p-value of 0.2539 that indicates an insignificant sample mean difference. These total of 69 missing values are excluded from the sample set without consequences of representativeness.

Oplmet contains 7 (0,1%) and 10 (0,2%) missing observations for year 2013 and 2011 respectively. These missing observations are excluded from the sample without creating any bias in my results. *Oplmet* with values 'other' are 39 (0,9%) and 34 (0,4%) for year 2013 and 2011 respectively. The 2013 t-test results in a p-value of 0,7278 that indicates an insignificant difference from the sample mean. Excluding these 39 observations will not have consequences for validity of my results. The 2011 t-test results in a p-value of nearly zero, that indicates significant mean difference at the 1 % level. Household members that indicated 'other' have significantly higher discretionary savings than the rest of the sample.

[2] https://www.onderwijsincijfers.nl/kengetallen/sectoroverstijgend/nederlands-onderwijsstelsel/hoogst-behaalde-onderwijsniveau

CHAPTER 5 RESULTS

5.1 Condensed results

Cross-sectional analysis of the relationship between the discretionary savings and LOC is performed for three years using the following OLS regression:

$$s_i = \alpha + \beta_1 il_i + \beta_2 el_i + \beta_3 y_i + \beta_4 g_i + \beta_5 e_i + \beta_6 h_i + \sum_{i=7}^{k} \beta_i z_i + \varepsilon_i$$

Following literature as stated in the methodology section, I expect internal LOC to vary positively with discretionary savings. Income is expected to vary positively with discretionary savings, as well as age, homeownership and the degree of completed education.

Table 5 describes the total results for all three years and table 6 describes the results the replicated cross-sectional analysis for the baseline models of years 2015, 2013 and 2011. The tables show the coefficients accompanied by the t-values between brackets. *** Indicates significance at the 1% level, ** indicates significance at the 5% level and * indicates significance at the 10% level. In the last rows, sample size, the F statistic, R-squared, adjusted r-squared and the root mean square error provide some extra diagnostics to interpret the results more completely. The analyses used all household members that have non-missing values for all variables and covariates of interest. Full regression outputs are available in the appendix.

Table 5. Results of cross-sectional analysis for total sample

Variable	1	2	3
internal locus of control	1.498,04*** (7,21)	1500,50*** (7,19)	1.494,23*** (7,18)
external locus of control	-339,62 (-1,13)	n/a	n/a
age	840,37*** (13,25)	829,72*** (13,96)	834,89*** (14,09)
log income	6.708,35*** (6,18)	6.741,72*** (6,18)	6.656,63*** (6,15)
homeownership	8.607,58*** (4,66)	8.626,87*** (4,72)	9.212,79*** (5,18)
Education:			

special education	-329,05 (-0,06)	-426,93 (-0,08)	-464,42 (-0,08)
primary education	3.871,01 (1,39)	3.777,68 (1,36)	3.624,13 (1,31)
HAVO/VWO/pre-university education	11.003,70*** (4,26)	11.103,44*** (4,29)	11.185,18*** (4,33)
MBO / senior vocational training	2.314,37 (1,24)	2,404,92 (1,30)	2.328,44 (1,27)
HBO / vocational education	12.621,80*** (4,80)	12.788,25*** (4,87)	12.666,08*** (4,86)
WO / university education	20.358,00*** (4,29)	20.603,25*** (4,33)	20.006,71*** (4,16)
no education	55.348,44*** (3,97)	55.067,53*** (3,97)	56.081,13*** (4,01)
household members	-1.392,49 (-0,72)	-2.080,73** (-2,47)	-2.842,66*** (-4,71)
nr of children in household	-542,92 (-0,34)	n/a	n/a
degree of urbanization	780,12 (1,02)	782,86 (1,04)	n/a
composition of household	-1.702,50 (-1,18)	-1.637,08 (-1,26)	n/a
most involved financial admin	-5.713,75*** (-3,00)	-5.576,12*** (-3,06)	-5.380,85*** (-3,00)
main wage earner	-6.730,92*** (-3,35)	-6,579,33*** (-3,43)	-6.389.28*** (-3,38)
partner present in household	-1.508,91 (-0,42)	n/a	n/a
sample size	6.155	6.155	6.215
F statistic	(19, 6135) 25,56)	(16, 6138) 28,66	(14,6200) 32,85
R-squared	0,0727	0,0725	0,0723
Adj. r-squared	0,0698	0,0703	0,0706
root MSE	70.149	70.138	69.828

Table 5. This table shows the results of the replicated cross-sectional analysis for the total sample. The results of column 1 are the baseline model, column 2 and 3 show results from models only including significant or nearly becoming significant covariates. Column 1, 2 and 3 show the coefficient and t-statistic between brackets resulting from the regression. In the last rows, sample size, the F statistic, R-squared, adjusted r-squared and the root mean square error provide some extra diagnostics to interpret the results more completely. *** indicates significance at the 1% level, ** indicates sigificance at the 5% level and * indicates significance at the 10% level.

Table 6. Results of cross-sectional analysis per year

Variable	2015	2013	2011
internal locus of control	1.630,3*** (4,93)	1.044,3*** (2,86)	1.961,0*** (4,46)
external locus of control	208,9 (0,38)	-1.040,8** (-2,55)	-362,7 (-0,61)
age	807,9*** (8,05)	789,9*** (8,95)	997,3*** (6,50)
log income	6.578,1*** (3,12)	8.090,1*** (5,57)	6134,7*** (2,98)
homeownership	6.648,8** (2,19)	6.841,6** (2,28)	13.371,5*** (3,50)
Education:			
special education	1.435,3 (0,15)	405,0 (0,03)	-2.633,4 (-0,31)
primary education	6.479,7 (1,15)	5.930,3 (1,21)	-281,4 (-0,13)
HAVO/VWO/pre-university education	13.857,9*** (2,90)	9.331,0** (2,20)	11.005,4*** (2,39)
MBO / senior vocational training	1.230,1 (-0,42)	5.477,6 (1,63)	5.116,2 (1,47)
HBO / vocational education	11.509,5*** (2,65)	14.066,9*** (3,20)	14.448,9*** (2,71)
WO / university education	9.290,6 (0,97)	16.456,2*** (2,70)	40.931,9*** (4,53)
no education	103.475,7*** (22,39)	30.230,9*** (6,53)	47.269,3** (2,52)
household members	742,2 (0,21)	-3.946,6 (-1,13)	3,6 (0,00)
nr of children in household	-1.591,7 (-0,61)	-54,7 (-0,02)	-774,3 (-0,26)
degree of urbanization	3.273,2** (2,13)	-273,9 (-0,23)	-1.398,1 (-1,11)
composition of household	-4.041,7 (-1,32)	1.627,5 (0,71)	-2.689,0 (-1,51)
most involved financial admin	-4.700,9 (-1,39)	-3.714,3 (-1,24)	-9.838,8*** (-2,68)
main wage earner	-5.102,1 (-1,45)	-6.509,4** (-2,07)	-9.818,8** (-2,45)
partner present in household	-8.061,6 (-1,16)	1.085,2 (0,20)	1.473,2 (0,22)
sample size	2.370	1.886	1899

F statistic	(19, 2350) 329,50	(19, 1866) 22,47	(19, 1879) 8,46
R-squared	0,0583	0,0911	0,1044
Adj. r-squared	0,0527	0,0843	0,0977
root MSE	79.282	60.645	71.986

Table 6. This table shows the results of the replicated cross-sectional analysis of the years 2015, 2013 and 2011. The results of the baseline models with all included covariates are shown. Column 2 – 4 show the coefficient, accompanied by t-statistic resulting from the regression. In the last rows, sample size, the F statistic, R-squared, adjusted r-squared and the root mean square error provide some extra diagnostics to interpret the results more completely. *** indicates significance at the 1% level, ** indicates sigificance at the 5% level and * indicates significance at the 10% level.

In the analyses carried out were positive significant effects for all three years for internal LOC on discretionary savings with coefficients of 1.630, 1.044 and 1.961 in the years 2015, 2013 and 2011 respectively. These coefficients show relative low variability and are in all three years significant and positive, which suggests that this effect is to these aspects stable over time. This effect is significant for all years at the 1% level. More positive and significant effects over all three years on discretionary savings are found for age (coefficients for 2015, 2013, 2011: 808, 790 and 997). The coefficients for age over discretionary savings are highly stable over the three years of measurement and the high t-values (8,05, 8,95 and 6,50 in 2015, 2013 and 2011 respectively) indicate reliable results. In addition, the effect of income on discretionary savings is significant in all three years at the 1% level, with coefficients of 6.578, 8.090 and 6.135 in year 2015, 2013 and 2011 respectively. The effect of owning a house versus renting a house also results in a positive and significant effect on discretionary savings for all three years. Education shows a positive significant effect on discretionary savings for pre-university and vocational education.

The effect of no education at all on discretionary savings is disproportionally high and significant for all three years. The coefficients show high variability from 103.475 in 2015, 30.230 in 2013 to 47.269 in 2011. As this contradicts intuition, several robustness checks are performed in section 5.4.

The effect of external LOC on discretionary savings is not significant over all years. The coefficients show high variability and are positive in 2015 (209) and negative in 2013 and 2011 (-1.041 and 363). Characteristics of the household show that the number of household members does have a significant effect on discretionary savings, the same effect holds for the number of children in the household. The effects of being a main wage earner and being most involved with the financial administration are non-significant, yet the coefficient is negative in all three years. The degree of urbanization shows

high variability within the three years with coefficients of 3.273 in 2015, -274 in 2013 and -1.398 in 2011. All three models show an adjusted r-squared of 0,0527, 0,0843 and 0,0977, which indicates that around 5 to 10% of the total variance is explained.

5.2 Constructing LOC variables using factor analysis

This section describes the process of factor analysis to form my measure for internal and external LOC. Since an individuals' degree of internal or external LOC cannot be asked for explicitly, I expect there to be underlying factors in the thirteen separate variables that measure LOC. To check this, I construct a table of correlation coefficients shown in table 7. The table shows some correlations close to or above the threshold of 0,5. These above threshold correlations are centered at the variables locus05 – locus07 and at locus12 - locus13. There are correlations that are considerably low.

Table 7. Correlation coefficients of LOC variables in 2015

	locus01	*locus02*	*locus03*	*locus04*	*locus05*	*locus06*	*locus07*	*locus08*
locus01	1.0000							
locus02	0.3668	1.0000						
locus03	0.4006	0.4651	1.0000					
locus04	0.2258	0.3729	0.3241	1.0000				
locus05	0.1829	0.1327	0.2783	0.1914	1.0000			
locus06	0.2347	0.2196	0.2918	0.1916	**0.4841**	1.0000		
locus07	0.1645	0.2515	0.2562	0.2596	**0.3771**	**0.5452**	1.0000	
locus08	-0.0976	-0.1144	-0.1325	-0.1597	-0.1410	-0.1560	-0.1535	1.0000
locus09	0.0486	0.0885	0.0407	0.0911	0.0746	0.1078	0.1367	0.0073
locus10	-0.0739	-0.0710	-0.0909	-0.1429	-0.1314	-0.1353	-0.0938	0.4484
locus11	-0.0343	0.0098	-0.0614	0.0055	-0.1895	-0.0701	-0.0583	0.2734
locus12	-0.0623	-0.0403	-0.1117	-0.0672	-0.1731	-0.0582	-0.0344	0.3003
locus13	-0.0239	-0.0388	-0.0849	-0.0643	-0.1616	-0.0414	-0.0325	0.2900

	locus09	*locus10*	*locus11*	*locus12*	*locus13*
locus09	1.0000				
locus10	0.0620	1.0000			
locus11	-0.0221	0.3438	1.0000		
locus12	-0.0212	0.3335	**0.4937**	1.0000	
locus13	0.0267	0.3323	**0.3807**	**0.4929**	1.0000

Table 7. Correlation coefficients of LOC variables in 2015. The table shows the correlation coefficients between the variable in column 1 and row 2, for example. The diagonal displays correlations of one; this is the correlation with itself. The upper half of the correlation table is blank, as this is a mirror image of the lower half.

This provides reason to expect two common factors. Executing factor analysis shows two retained factors for all three years satisfying the Kaiser-Criterion with Eigenvalues higher than one, which means that they account for more variability then the original variables do. These results are presented in figure XX and as shown, the factor structure for all three years is very similar, indicating low variability between years and making it legitimate to perform factor analysis for all years.

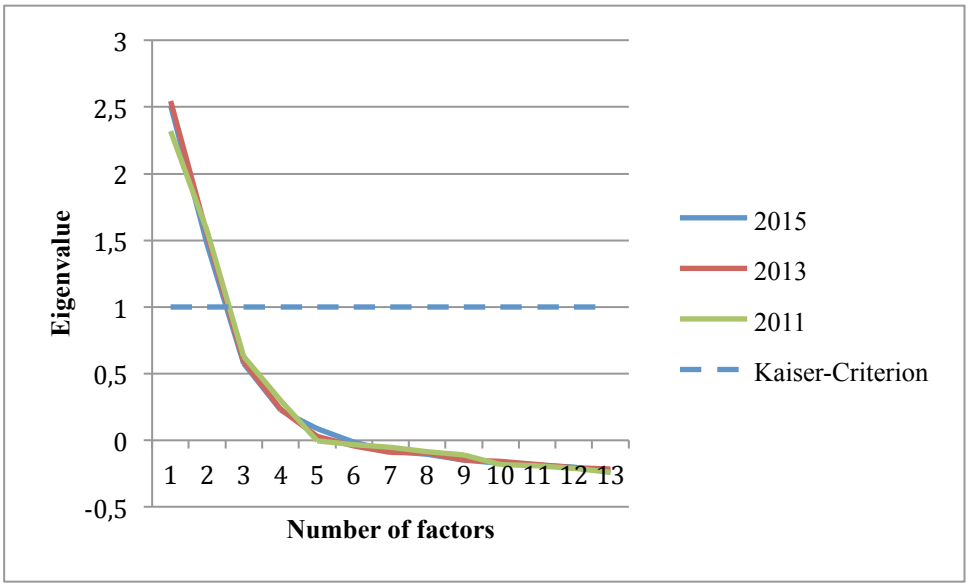

Figure 4. Eigenvalues of factors in LOC variables for the years 2015, 2013 and 2011. The blue dotted line indicates the significance level of 1, determined by the Kaiser Criterion. This indicates that in each year two factors, which explain more variability than the original variables, are retained.

For 2015, factor one has an eigenvalue of 2,51 and factor two has an eigenvalue of 1,47. In 2013 and 2011, the factors 1 have an eigenvalue of 2,54 and 2,32 respectively. The factors 2 have eigenvalues of 1,54 and 1,57, respectively.

To interpret these factors, I analyze their factor loadings. Table 8 shows factor loadings for each year. The bold numbers are the most significant influencers in forming the factors. The bold and * indicate that factor one is mostly formed by *locus03, locus06* and *locus07*. These variables have labels such as "When I get what I want, it's usually because I worked hard for it" and "My life is determined by my

own actions". As these variables are internally focused, it is suited to label this first factor *Internal Locus of Control.* The bold and ** indicate that factor two is mostly formed by *locus11*, *locus12* and *locus13*. These variables have labels such as "It's not always wise for me to save because many things turn out to be a matter of luck" and "It is chiefly a matter of fate whether I become rich or poor". These labels are externally focused, and point towards a non-causal relationship between lifetime outcomes and their own actions. Therefore it is appropriate to label the second factor *External Locus of Control.* It is worth noticing that all variables have considerable high unique variances (>0,5). This provides reason to perform an additional check how suitable my data is for factor analysis. I use the Kaiser-Meyer-Olkin (KMO) measure for sample adequacy.

Table 8. Factor loadings

variable	2015		2013		2011	
	factor 1	factor 2	factor 1	factor 2	factor 1	factor2
locus01	0,3949	0,2545	0,4242	0,2827	0,4118	0,2620
locus02	0,4469	0,3331	0,4631	0,3172	0,4623	0,2895
locus03	**0,5251***	0,2903	**0,5330***	0,2786	**0,5129***	0,2979
locus04	0,4163	0,2235	0,4760	0,1264	0,3784	0,1368
locus05	**0,5220***	0,1136	**0,5134***	0,1338	**0,4704***	0,1587
locus06	**0,5515***	0,2958	**0,5838***	0,3087	**0,5079***	0,3865
locus07	**0,5047***	0,3017	**0,5702***	0,2231	**0,4707***	0,3238
locus08	-0,4442	0,2896	-0,4547	0,3834	-0,4766	0,3067
locus09	0,1062	0,1206	0,0767	0,0735	0,1201	0,0855
locus10	0,4307	0,3821	-0,3739	0,4286	-0,3443	0,3999
locus11	-0,3813	**0,4845****	-0,3276	**0,4497****	-0,3600	**0,4726****
locus12	-0,4333	**0,5125****	-0,3893	**0,5658****	-0,4769	**0,5365****
locus13	-0,3846	**0,4741****	-0,3120	**0,4970****	-0,3336	**0,5097****

Table 8. This table shows the factor loadings of the two retained factors in the years 2015, 2013 and 2011. The bolded measures highlight the highest loadings, and thus the most influential variables. The bold and * indicate the most influential variables for factor 1, the bold and ** indicate the most influential variables for factor 2.

The KMO measure shows how much the factors can explain each individual variable and is displayed in figure 5. The overall KMO adequacy is 0,78 for 2015 and 0,79 and 0,76 for 2013 and 2011, respectively. This is above the common threshold of > 0,5 and indicates nearly meritorious data. The only outlier is *locus09*, however this variable is not the main influential part in determining the factors.

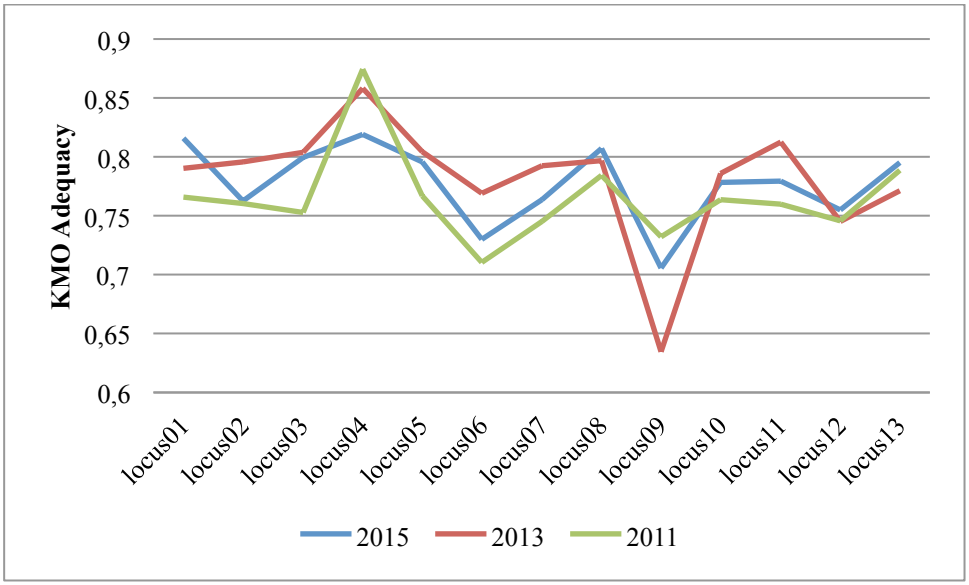

Figure 5. This figure shows the Kaiser-Meyer-Olkin adequacy for all LOC variables. This measure indicates the adequacy of the two factors explaining each individual variable. The y-axis shows the KMO adequacy, the x-axis shows the thirteen LOC variables.

5.3 Estimates of the model specified per year

5.2.1 2015

In this paragraph I describe the results for year 2015. The results of the analysis in 2015 are shown in table 9. Table 9, column 1 shows the results of the baseline model with all covariates included. Coefficients are provided and t-values are included between brackets. Column 2 presents the second model, in which I include significant covariates and covariates that are nearby significance, indicated by their p-value close to 0,100. Column 3 shows the model with only significant covariates. The baseline model shows the highest adjusted r-squared (0,0527) and thus is best fitted to the dataset. The F-statistic indicates that the model was significant as a whole [F (19, 2350) = 329,50].

The results show a consecutive positive significant effect for internal LOC on discretionary savings. This effect has a coefficient of 1.630,3 and is significant at the 1% level. Age is also positively significant at the 1% level with a coefficient of 807,9 and a t-value of 8,05. Going forward to row 4, the logarithmic income shows a consistent positive and significant effect at the 1% level with a coefficient of 6.578,1. The results show other significant effects for pre-university education (coefficient = 13.857,9, t-value = 2,90), vocational education (coefficient = 11.509,5, t-value = 2,65) and no education at all (coefficient = 103.475,7, t-value = 22,39). This unexpected significant effect of no education at all on savings is significant at the 1% level with a disproportional high t-value compared to the other results. The coefficient is also not likely to be proportionate. These specific results will be subject to a robustness check (see section 5.4). The degree of urbanization of the residence of the household member has a significant positive effect on household discretionary savings significant at the 5% level.

Table 9. Results of cross-sectional analysis for 2015

Variable	1	2	3
internal locus of control	1.630,3*** (4,93)	1.619,9*** (4,89)	1.602,0*** (4,81)
external locus of control	208,9 (0,38)	245,8 (0,45)	n/a
age	807,9*** (8,05)	800,0*** (9,10)	765,3*** (9,29)
log income	6.578,1*** (3,12)	6.501,7*** (3,01)	6.506,2*** (2,98)
homeownership	6.648,8** (2,19)	5.247,7* (1,82)	5.965,7** (2,07)
Education:			
special education	1.435,3 (0,15)	2.211,3 (0,21)	2.697,1 (0,26)

primary education	6.479,7 (1,15)	6.716,9 (1,21)	7.977,7 (1,45)
HAVO/VWO/pre-university education	13.857,9*** (2,90)	13.983,4*** (2,91)	13.245,5*** (2,77)
MBO / senior vocational training	-1.230,1 (-0,42)	-1.736,5 (-0,55)	-3.092,0 (-1,01)
HBO / vocational education	11.509,5*** (2,65)	11.324,5*** (2,61)	9.586,4** (2,33)
WO / university education	9.290,6 (0,97)	9.066,6 (0,94)	6.866,7 (0,72)
no education	103.475,7*** (22,39)	102.268,3*** (21,37)	104.067,3*** (21,34)
household members	742,2 (0,21)	n/a	n/a
nr of children in household	-1.591,7 (-0,61)	n/a	n/a
degree of urbanization	3.273,2** (2,13)	3.031,7** (2,04)	3.074,8** (2,07)
composition of household	-4.041,7 (-1,32)	-5.205,3*** (-2,74)	-4.347,4** (-2,41)
most involved financial admin	-4.700,9 (-1,39)	-3.675,8 (-1,14)	n/a
main wage earner	-5.102,1 (-1,45)	-4.048,3 (-1,21)	n/a
partner present in household	-8.061,6 (-1,16)	n/a	n/a
sample size	2370	2370	2370
F statistic	(19, 2350) 329,50	(16,2353) 372,04	(13, 2356) 470,90
R-squared	0,0583	0,0573	0,0560
Adj. r-squared	0,0527	0,0517	0,0504
root MSE	79.282	79.270	79.275

Table 9. This table shows the results of the cross-sectional analysis of the year 2015. The results of column 1 are the baseline model, column 2 and 3 show results from models only including significant or nearly becoming significant covariates. Column 1, 2 and 3 show the coefficient and t-statistic between brackets resulting from the regression. In the last rows, sample size, the F statistic, R-squared, adjusted r-squared and the root mean square error provide some extra diagnostics to interpret the results more completely. *** indicates significance at the 1% level, ** indicates sigificance at the 5% level and * indicates significance at the 10% level.

5.2.2 2013

The results for 2013 are shown in table 10. Equally to the earlier mentioned results, column 1 shows the results of the baseline model, column 2 shows the model with significant covariates included and covariates that are close to significance. Column three shows the model only with significant covariates. Coefficients are provided and t-values are included between brackets. The F statistic shows again that the model was significant as a whole with F (19,1866) = 22,47. Adjusted r-squared shows that model 1 is the best fit with the data.

As can be seen in table 10, the results show a consequent positive and significant effect at the 1% level for internal LOC, external LOC, age and income. The coefficients are relatively stable over the three models. In contrast to 2015, this year homeownership has a positive significant effect at the 5% level on discretionary savings with a coefficient of 6.841. Moreover, university education shows this year a significant effect on discretionary savings with a t-value of 2,70 and a coefficient of 16.456.

Pre-university education and vocational education show a positive significant effect on discretionary savings with coefficients of 9.331 and 14.067 respectively. The results for not having an education at all are again significant at the 1% level with a coefficient of 30.230. This results seems less inflated than the 2015 result, however it still is disproportional compared to other results. Lastly, in 2013 there is a significant negative effect of being the main wage earner on discretionary savings with a coefficient of -6.509 and t-value of -2,07.

Table 10. Results of cross-sectional analysis for 2013

Variable	1	2	3
internal locus of control	1.044,3*** (2,86)	1.044,0*** (2,87)	1.039,7*** (2,86)
external locus of control	-1.040,8** (-2,55)	-1.078,9*** (-2,70)	-1.075,9*** (-2,69)
age	789,9*** (8,95)	785,6*** (9,34)	777,9*** (9,31)
log income	8.090,1*** (5,57)	7.995,4*** (5,70)	8.039,6*** (5,74)
homeownership	6.841,6** (2,28)	6.952,8** (2,33)	7.209,6** (2,43)
Education:			
special education	405,0 (0,03)	1.057,7 (0,09)	2.006,8 (0,17)

primary education	5.930,3 (1,21)	6.030,6 (1,25)	6.608,9 (1,37)
HAVO/VWO/pre-university education	9.331,0** (2,20)	9.473,8** (2,24)	9.191,9** (2,18)
MBO / senior vocational training	5.477,6 (1,63)	5.248,5 (1,57)	5.040,5 (1,52)
HBO / vocational education	14.066,9*** (3,20)	14.238,4*** (3,23)	13.907,4*** (3,18)
WO / university education	16.456,2*** (2,70)	16.443,5*** (2,83)	16.079,0*** (2,77)
no education	30.230,9*** (6,53)	28.644,4*** (6,47)	29.001,4*** (6,58)
household members	-3.946,6 (-1,13)	-3.095,5*** (-3,17)	-2,888,0*** (-2,98)
nr of children in household	-54,7 (-0,02)	n/a	n/a
degree of urbanization	-273,9 (-0,23)	n/a	n/a
composition of household	1.627,5 (0,71)	n/a	n/a
most involved financial admin	-3.714,3 (-1,24)	-4.034,7 (-1,43)	n/a
main wage earner	-6.509,4** (-2,07)	-6.772,4** (-2,29)	-8.277,4*** (-2,66)
partner present in household	1.085,2 (0,20)	n/a	n/a
sample size	1.886	1.902	1.902
F statistic	(19, 1866) 22,47	(15, 1886) 27,38	(14, 1887) 29,06
R-squared	0,0911	0,0909	0,0902
Adj. r-squared	0,0843	0,0842	0,0835
root MSE	60.645	60.376	60.385

Table 10. This table shows the results of the cross-sectional analysis of the year 2013. The results of column one are the baseline model, column 2 and three show results from models only including significant or nearly becoming significant covariates. Column 1, 2 and 3 show the coefficient and t-statistic between brackets resulting from the regression. In the last rows, sample size, the F statistic, R-squared, adjusted r-squared and the root mean square error provide some extra diagnostics to interpret the results more completely. *** indicates significance at the 1% level, ** indicates sigificance at the 5% level and * indicates significance at the 10% level.

5.2.3 2011

Table 11 shows the results for 2011. The F statistic for this year is F = (19,1879) = 8,46, indicating that the model jointly together is significantly different from zero. The results show again positive and significant effects on discretionary savings at the 1% level for internal LOC (coefficient: 1.961), age (coefficient: 997), income (coefficient: 6135) and homeownership (coefficient: 13.372). Pre-university, vocational and university education show a positive significant effect on discretionary savings at the 1% level. Again the effect of no education on discretionary savings is significant, here at the 5% level with a coefficient of 47.269.

Household characteristics as the number of household members, number of children, degree of urbanization of the residence, the composition of the household and whether there is a partner present in the household do not show any significant effects on discretionary savings. If someone is the most involved in the financial administration in the household, this affects discretionary savings negatively with a coefficient of -9.839, significant at the 1% level with a t-value of 2,68. The effect of being the main wage earner in a household on discretionary savings is negative and significant for all three models.

Table 11. Results for cross-sectional analysis for 2011

Variable	1	2	3
internal locus of control	1.961,0*** (4,46)	1.960,9*** (4,50)	1.984,8*** (4,51)
external locus of control	-362,7 (-0,61)	-352,6 (-0,61)	n/a
age	997,3*** (6,50)	1.023,3*** (7,52)	1.013,3*** (7,72)
log income	6134,7*** (2,98)	6.166,7*** (2,97)	6.309,5*** (3,08)
homeownership	13.371,5*** (3,50)	13.563,9*** (3,83)	12.683,4*** (3,73)
Education:			
special education	-2.633,4 (-0,31)	-3.112,9 (-0,36)	-4.257,3 (-0,50)
primary education	-281,4 (-0,13)	-273,8 (-0,07)	-49,9 (-0,01)
HAVO/VWO/pre-university education	11.005,4*** (2,39)	10.918,8** (2,36)	10.928,9** (2,37)
MBO / senior vocational training	5.116,2 (1,47)	5.070,8 (1,45)	5.146,2 (1,47)

HBO / vocational education	14.448,9*** (2,71)	14.581,0*** (2,78)	15.271,5*** (2,82)
WO / university education	40.931,9*** (4,53)	41.047,6*** (4,58)	42.240,4*** (4,67)
no education	47.269,3** (2,52)	47.244,9** (2,50)	45.955,3** (2,47)
household members	3,6 (0,00)	n/a	n/a
nr of children in household	-774,3 (-0,26)	n/a	n/a
degree of urbanization	-1.398,1 (-1,11)	-1.421,5 (-1,17)	n/a
composition of household	-2.689,0 (-1,51)	-2.853,0* (-1,78)	-2.908,4* (-1,83)
most involved financial admin	-9.838,8*** (-2,68)	-10.128,1*** (-2,87)	-10.099,2*** (-2,86)
main wage earner	-9.818,8** (-2,45)	-10.131,1*** (-2,64)	-10.123,5*** (-2,67)
partner present in household	1.473,2 (0,22)	n/a	n/a
sample size	1.899	1.899	1.908
F statistic	(19, 1879) 8,46	(16, 1882) 8,98	(13, 1894) 9,50
R-squared	0,1044	0,1043	0,0998
Adj. r-squared	0,0977	0,0976	0,0931
root MSE	71.986	71.932	71.906

Table 11. This table shows the results of the cross-sectional analysis of the year 2011. The results of column one are the baseline model, column 2 and three show results from models only including significant or nearly becoming significant covariates. Column 1, 2 and 3 show the coefficient and t-statistic between brackets resulting from the regression. In the last rows, sample size, the F statistic, R-squared, adjusted r-squared and the root mean square error provide some extra diagnostics to interpret the results more completely. *** indicates significance at the 1% level, ** indicates sigificance at the 5% level and * indicates significance at the 10% level.

5.3 Regression diagnostics

Results may be misleading if the underlying data do not meet the underlying OLS assumptions. These assumptions contain information about linearity, normality, homogeneity of variance, multicollinearity and error independency. In this section I will check for these assumptions.

The first assumption contains linearity between the dependent and independent variables and is tested by scatterplot of internal locus of control over discretionary savings. No sign of clear linearity is visible in the plots of all three years. The patterns show low variability among the complete dataset with a few data points that show higher dispersion. In 2013, the degree of internal LOC is more condensed compared to the other years. Graphical representation of higher powers of the internal LOC shows dispersion that does not fit to these models. This indicates that the linear model fits best to the dataset.

Normality of residuals is required for hypothesis testing since this assumes that the p-values for t-tests and F-values are valid. OLS regressions require that the errors are independent and identically distributed. Residuals are predicted in STATA and first a Kernel density plot makes a comparison between the normal density distribution and the actual distribution of the residuals. This plot shows a more condensed distribution of the residuals compared to the normal distribution in all years. A standardized normal probability plot shows signs of heteroscedacity since the actual distribution does not follow the normal distribution. The pnorm measure is sensitive to non-normality in the middle range of the data, while qnorm is sensitive to non-normality in the tails. The qnorm distribution shows a spread of the residuals that is wider towards the beginning and end of the graph. I can conclude that the residuals are not normally distributed and might suffer from heteroscedacity.

To check for heteroscedasticity of the residuals, the White test is performed and results in Prob > F = 0.0000 to reject the homoscedacity hypothesis. An RVF plot shows a non-constant pattern of predicted residuals in all three years. Residuals are narrower at the left side and show more dispersion towards the right-end of the graph. This indicates that errors differ in variance, and this is another indication of heteroscedasticity in my dataset.

With a Variance Inflation Factor (VIF) analysis I check the variables for multicollinearity. As the degree of multicollinearity increases, the regression model estimates become unstable and the standard errors for the coefficients can be inflated. High (> 10) VIF factors could indicate inflated p-values, which is bad for the validity of my results. The VIF for all variables in 2015, 2013 and 2011 is very close to 1, and none of the VIFs exceeds 2. This indicates that these regression models do not suffer from multicollinearity.

The DNB dataset does not meet all required assumptions for linear regression, so some adjustments are made to minimize the biased estimates of coefficients: in all three years errors are found to be heteroscedastic and therefore all regressions are performed with errors replaced with White's robust standard errors.

5.4 Robustness checks on the data

As explained in section 4.2.2, including only the observations with complete values for all covariates might introduce some selection bias into my results since in some cases quite high percentages of all data points in a specific year with significant mean differences are excluded. In this section I perform several robustness checks to limit this influence and to inspect whether my results are stable and reliable. These robustness checks simultaneously provide more thorough understanding of the disproportionally high positive and significant effect of no education at all on discretionary savings.

5.4.1 excluding income as a covariate

First I will analyse whether the influence of income has a significant impact on my results. By including income as a covariate, in 2015 2.076 observations are excluded due to missing values or 'don't know' values. Keeping these observations into the dataset and performing the cross-sectional regression again without income as a covariate leads to the results shown in table 9. Full regression outputs of this robustness check are available in the appendix in figure S. The overall model is significant with F (18, 3833) = 11,37. The effect of internal LOC on discretionary savings has become stronger with a coefficient of 1.512 and a t-value of 6,18 that indicates significance at the 1% level. The coefficient stays nearly the same; the covariate income included provides a coefficient of 1.630, versus the 1.512 of this robustness check. This result is favourable since it shows that the inclusion of income as a covariate nearly did not change the main coefficient of interest. The effects of age and homeownership on discretionary savings are additionally significant at the 1% level with coefficients of 598 and 8.116 respectively. Pre-university and vocational educations also still have a significant effect on discretionary savings at the 1% level with large coefficients of 12.069 and 7.587 respectively. The degree of urbanization is still significant for 2015 at the 5% level with a coefficient of 2.411.

This first robustness check also dampens the highly significant result of not having a completed education at all on discretionary savings. Including the observations that did not enter or did not know their income data had significantly lower discretionary savings than the rest of the sample. Including these significantly lower income observations results in a t-value of 2,13 that indicates a significant effect at the 5% level with a coefficient of 57.230. Noteworthy to add is the high 95% confidence

interval of this variable that ranges from 4.497 to 109.963, which is very wide compared to the confidence intervals of other covariates.

Table 12. Robustness checks

Variable	1	2	3	4
internal locus of control	1.630,3*** (4,93)	1.511,8*** (6,18)	1.388,3*** (4,37)	1.633,2*** (4,94)
external locus of control	208,9 (0,38)	-4,8 (-0,01)	561,3 (1,13)	208,0 (0,37)
age	807,9*** (8,05)	597,7*** (8,50)	829,3*** (8,26)	807,0*** (8,04)
dummy age < 16	n/a	n/a	-10.473,8* (-1,76)	n/a
log income	6.578,1*** (3,12)	n/a	5.050,2** (2,30)	6.560,2*** (3,12)
homeownership	6.648,8 (2,19**)	8.116,2 (3,95***)	5.739,7 (1,97**)	6.648,4 (2,19**)
dummy special housing	n/a	n/a	n/a	-35.400,0 (-2,51**)
Education:				
special education	1.435,3 (0,15)	-1.159,4 (-0,19)	836,2 (0,09)	1.429,1 (0,15)
primary education	6.479,7 (1,15)	3.156 (0,86)	8.568,6 (1,56)	6.492,8 (1,16)
HAVO/VWO/pre-university education	13.857,9*** (2,90)	12.068,6*** (3,52)	13.688,4*** (2,86)	13.898,0*** (2,91)
MBO / senior vocational training	-1.320,1 (-0,42)	-1.328.9 (-0,66)	-1.178,3 (0,38)	-1.343,54 (-0,43)
HBO / vocational education	11.509,5*** (2,65)	7.586,5*** (2,62)	12.569,4*** (2,88)	11.477,8*** (2,65)
WO / university education	9.290,6 (0,97)	7.768,4 (1,21)	11.029,6 (1,12)	9.279,3 (0,97)
no education	103.475,7*** (22,39)	57.230,6** (2,13)	11.108,5 (1,40)	103.452,2*** (22,45)
household members	742,2 (0,21)	-257,0 (-0,12)	2.269,6 (0,80)	773,1 (0,22)
nr of children in household	-1.591,7 (-0,61)	-1.503,3 (-0,89)	-1.548,3 (-0,80)	-1.620,1 (-0,63)
degree of urbanization	3.273,2** (2,13)	2.411,0** (2,39)	3.874,7** (2,50)	3.272,4** (2,13)

composition of household	-4.041,7 (-1,32)	-2.905,1 (-1,59)	-5.197,2* (-1,92)	-4.069,3 (-1,34)
most involved financial admin	-4.700,9 (-1,39)	-3.581,9 (-1,61)	-4.805,1 (-1,43)	-4.704,6 (-1,39)
main wage earner	-5.102,1 (-1,45)	-3.970,3* (-1,72)	-5.314,2 (-1,52)	-5.100,8 (-1,45)
partner present in household	-8.061,6 (-1,16)	-4.192,6 (-0,94)	-9.767,3 (-1,53)	-8.063,4 (1,16)
sample size	2.370	3.852	2.814	2.377
F statistic	(19, 2350) 329,50	(18, 3833) 11,37	(20, 2.793) 11,19	(20, 2.356) 342,11
R-squared	0,0583	0,0492	0,0532	0,0585
Adj. R-squared	0,0527	0,0460	0,0481	0,0525
root MSE	79.282	66.458	80.992	79.182

Table 12. This table shows the results of the performed robustness checks on the 2015 data. Model 1 shows the baseline model. Column 2 shows the results of the model excluding income as a covariate. Column 3 shows the results of the model including a dummy for underage saving. Column 4 shows the results of the model including a dummy for special housing. All columns show the coefficient and t-statistic between brackets resulting from the regression. In the last rows, sample size, the F statistic, R-squared and the root mean square error provide some extra diagnostics to interpret the results more completely. *** indicates significance at the 1% level, ** indicates sigificance at the 5% level and * indicates significance at the 10% level.

5.4.2 including age below 16 as a covariate

As age is constructed by taking the entered year of birth at the time of filling in the questionnaire, a lot of inappropriate observations arise. A total of 951 observations result in missing values (3) and age below 16 (948). This is inconsistent with the expectations since DNB states to present the questionnaire to household members with age > 16 only. These values might indicate a human error in filling in the year of birth or it might be that these household members are not yet financially autonomous. Household members with age < 16 have significant lower mean income at the 1% level than household members with age > 16. Including these 948 observations with a dummy for underage saving and performing a cross-sectional regression again results in a significant model with F (20,2793) = 11,19 (see table 9 for all results).

The effect of internal LOC on discretionary savings is significant at the 1% level with a coefficient of 1.388. Other positive significant effects are age (coefficient 829, t-value 8,26), income (coefficient 5.050, t-value 2,30), homeownership (coefficient 5.740, t-value 1,97), pre-university education (coefficient 13.688, t-value 2,86), vocational education (coefficient 12.569, t-value 2,88) and the degree of urbanization (coefficient 3.874, t-value 2,50).

5.4.3 including special housing as a covariate

Special housing includes free living and sub-rented housing. Including only 7 extra observations with a dummy for special housing results in a coefficient of -35.400, significant at the 5% level. The other coefficients of interest stay relatively stable and significant. This additional covariate does not dampen the disproportional effect of no education on discretionary savings and results in a coefficient of 103.452, significant at the 1% level. This is close to the baseline model however with a slightly lower adjusted r-squared.

CHAPTER 6 CONCLUSION

In this paper I used a unique dataset with 6155 observations from the DNB Household Survey, containing detailed information on Dutch households, to explore the effects of internal LOC on discretionary savings. I hypothesized that internal LOC has a positive effect on discretionary savings. To my knowledge I am the first to empirically examine and document the effects of this relationship. The results of the performed analysis suggest that discretionary savings vary positively with the degree of internal LOC with coefficients being 1.630 in 2015, 1.044 in 2013 and 1.961 in 2011, which retains the stated hypothesis.

This result is stable to the aspects that it is positive and significant over three years of measurement and are robust after controlling for external LOC, age, income, homeownership, number of household members, number of children in the household, degree of urbanization of the residence, the composition of the household, whether or not someone is the most involved in the financial administration, whether or not someone is the main wage earner, and whether or not there is a partner present in the household. Furthermore, I find evidence that age, income, homeownership, pre-university- and vocational education have a positive and significant effect over all three years on discretionary savings, which suggests that these aspects also play a role in determining one's discretionary saving behaviour.

In my analysis, I constructed a measure for internal and external LOC by factor analysis. These factors satisfy to the Kaiser criterion and the Kaiser-Meyer-Olkin measure for sampling adequacy. The measure for discretionary savings excludes pension savings, as I only focused on the proportion that is voluntary withheld, i.e. where psychic costs tend to occur. This measure is constructed by transforming some categorical variables into continuous variables where the average of the category indicated is used to calculate total discretionary savings. The effect of missing observations is explored and the OLS regressions are performed using STATA.

The results in this paper are consistent with the results of enhanced perceived self-control on saving behavior (Thaler & Shefrin, 1988) and the general findings of Rosenbaum (1980) that indicate people having a more internal LOC tend to have more self-control. The results contribute to the literature of saving and consumption behaviors, more specifically to the influence of personality traits on discretionary saving. Second, it extends the understanding of the effects of internal LOC among Dutch households with specific characteristics. In the context of behavioural finance, these findings provide tools for creating effective nudges on stimulating discretionary saving.

6.1 Limitations and suggestions for future research

This research also contains its limitations. I am aware of the possible selection bias created in my results. First, the selection bias created excluding all observations with missing savings, as this might disproportionally exclude observations with low income, low age or low savings. Missing observations is more a rule than exception in survey data research and the fact that this database is collected by the DNB and not by myself results in this limitation to be inevitable. The possible bias created by the transformation of categorical variables into continuous variables could be tackled by constructing a model with mixed categorical and continuous dependent variables in Matlab or R, however this could not be done within the timespan of this research. Excluding missing income observations might have created a slight upward bias, as it disproportionally excluded observations with lower income. This effect of creating a bias is also found in excluding observations with missing LOC variables, as these observations had significantly lower discretionary savings.

In general, using questionnaires for research may introduce some bias since people might answer incorrectly by mistake or on purpose. This might especially be the case at the age variable, where in 2015, 948 (18,5%) of observations show age below 16 even though the DNB states to present the questionnaire only to household members with age above 16. To some extend, the robustness checks performed limit this possible bias as it indicates that my results are still robust after controlling for age below 16, including all observations with missing income and controlling for special housing. My analyses are executed replacing the residuals with White's heteroscedasticity corrected errors since the original dataset did not show residuals to be homoscedastic.

The independent variables internal LOC and external LOC are constituted by factor analysis. Though factor analysis is a good way to measure latent variables, I am aware that the interpretation and procedure of the factor analysis rests on some degree of subjectivity. For example, determining the number of retained factors could be done by several criteria.

The baseline model shows the best fit with the data, however it is notable that the adjusted r-squared is quite low for the presented model. This indicates that the model captures only 5 to 10% of the variation in the data. There is a proportion of the variance that cannot be explained by my model and this probably can be attributed to idiosyncratic shocks or individual preferences not captured by the model. However, since this model is not used for forecast purposes and only for academic purposes, I do not consider this caveat to be of serious importance.

The effect of no completed education at all on discretionary savings being highly positive and significant in all three years is something that I do not understand. It could be that there is an underlying and undiscovered reason that individuals with no education have exceptionally high discretionary savings, or it might be that these individuals are too embarrassed to tell the truth and therefore provide untrue values. The model controlling for underage saving shows interesting results: it drops the effect to not being significant and the coefficient for underage saving is negatively significant with a coefficient of -10.474. This suggests that the individuals who are below the age of 16 also did not completed any education and might be the cause of the disproportionally high effect of no education on discretionary savings in the baseline models.

The results of this paper do not answer to the question whether or not these effects are stable over time. The results do show stability in the aspects of showing positive and significant effects of internal LOC on discretionary saving; the degree of stability over time is subject for future research. The results of this paper could be extended using the aforementioned model that uses categorical and continuous dependent variables and needs to be modeled in Matlab or R. The research also could be extended using a longer time span or more international datasets.

REFERENCES

Abay, K. A., Berhane, G. and Assefa, T. W. (2016, June). Locus of Control, Hyperbolic Preferences and Demand for Commitment and Savings: Evidence from Rural Ethiopia. Paper presented at the UNU Wider Conference, University of Copenhagen. Retrieved from https://www.wider.unu.edu/sites/default/files/Events/PDF/Slides/1_abay.pdf

Almlund, M., Lee Duckworth, A., Heckman, J. and Kautz, T. (2011). Personality psychology and Economics. In Hanushek, E. A., Machin, S. and Woessmann, L. (Eds) *Handbook of the Economics of Education*, Amsterdam: Elzevier. 1 – 181 Retrieved from: http://www.nber.org/papers/w16822

Ameriks, J., Caplin, A. and Leahy, J. (2003). Wealth Accumulation and the Propensity to Plan. *The Quarterly Journal of Economics 68*, 1007-1047

Becker, A., Deckers, T., Dohmen, T.J., Falk, A. and Kosse, F. (2012). 'The relationship between economic preferences and psychological personality measures', mimeo, Institute for the Study of Labor (IZA)

Behrman, J. R., Mitchell, O. S., Soo, C. K. and Bravo, D. (2012). How Financial Literacy Affects Household Wealth Accumulation. *The American Economic Review 102(3),* 300 – 304

Bernheim, B. D., Skinner, J., and Weinberg, S. (2001). What Accounts for the Variation in Retirement Wealth among U. S. Households. *The American Economic Review 91(4)*, 832–857

Campbell, J. Y. and Mankiw, N. G. (1989). Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence. In Blanchard, O. J. and Fisher, S. (Eds) *NBER Macroeconomics Annual, 4,* 185 – 246. Retrieved from: http://www.nber.org/chapters/c10965

Carroll, C. D. (2001). A Theory of the Consumption Function, With and Without Liquidity Contraints. *Journal of Economic Perspectives 15(3),* 23 – 45

Choi, J., Laibson, D., Madrian, B., and Metrick, A. (2003). Optimal Defaults. *American Economic Review Papers and Proceedings, 93(2)*, 180-185

Cobb-Clark, D. A. and Schurer, S. (2013). Two Economists' Musings on the Stability of Locus of Control. *The Economic Journal, 123, 358 – 400.* Doi:10.1111/ecoj.12069

Coleman, M. and Deleire, T. (2003). An economic model of locus of control and the human capital investment decision. *Journal of Human Resources, 38(3)*, 701–21

Cronqvist, H., & Siegel, S. (2013, November). The origins of savings behavior. In *AFA 2011 Denver Meetings Paper*. Retrieved from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1649790

Dynan, K. E., Skinner, J. and Zeldes, S. P. (2004). Do the rich save more? *Journal of Political Economy 112 (2)*, 397- 444

Friedman, M. A. (1957). The Permanent Income Hypothesis. In: *A Theory of the Consumption Function.* 20 – 37. Princeton: Princeton University Press. Retrieved from: http://www.nber.org/chapters/c4405

Gatz, M. and Karel, J. (1993). Individual change in perceived control over 20 years. *International Journal of Behavioral Development, 16(2),* 305–322

Goldsmith, A., Veum, J.R. and Darity, W. (1996). The psychological impact of unemployment and Joblessness. *Journal of Socio-Economics, 25(3),* 333–358

Heineck, G. and Anger, S. (2010). The returns to cognitive abilities and personality traits in Germany. *Labour Economics, 17(3),* 535–546

Jappelli, T. and Modigliani, F. (1998). The Age-Saving Profile and the Life-Cycle Hypothesis. In The Collected Papers of Franco Modigliani. MIT Press

Kain, J.F. and Quigley, J. M. (1971). Housing Market Discrimination, Home-ownership and Savings Behavior. *The American Economic Review, 62(3),* 263 - 177

Kim, J. O., & Curry, J. (1977). The treatment of missing data in multivariate analysis. *Sociological Methods & Research, 6* (2), 215-240

Kimball, M.S. and Shumway, T. (2009). Fatalism, Locus of Control, and Retirement Saving. University of Michigan

Laibson, D. I. (1997). Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics, 112(2)*, 443-477. doi:10.1162/003355397555253

Laibson, D. I., Repetto, A., and Tobacman, J. (1998). Self-Control and Saving for Retirement. *Brookings Papers on Economic Activity, 1998(1),* 91 – 172

Li, J. (2012). Precautionary saving in the presence of labor income and interest rate risks. *Journal of Economics, 106(3),* 251 - 266. doi:10.1007/s00712-011-0244-6

Modigliani, F. and Brumberg, R. (1954). Utility analysis and the Consumption Function: An Interpretation of Cross-section Data. In The Collected Papers of Franco Modigliani. MIT Press

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied*, *80*(1), 1

Shefrin, H. M. and Thaler, R. H. (1988). The Behavioral Life-Cycle Hypothesis. *Economic Inquiry, 26(4), 609 – 643*

Tabachnick, B. G., Fidell, L. S. and Osterlind, S. J. (2001). Using Multivariate Statistics.

Thaler, R. H. and Shefrin, H. M. (1981). An Economic Theory of Self-Control. *The Journal of Political Economy, 89(2), 392 – 406*

Venti, S.F. and Wise, D.A. (1998). The Cause of Wealth Dispersion at Retirement: Choice or Chance? *The American Economic Review, 88,* 185 – 191

APPENDIX - *Regression output for baseline model 2015*

Linear regression

Number of obs	= 2,370
F(19, 2350)	= 329.50
Prob > F	= 0.0000
R-squared	= 0.0583
Root MSE	= 79282

networth2015	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
internallocus	1630.308	330.7038	4.93	0.000	981.8065	2278.81
externallocus	208.93	557.0474	0.38	0.708	-883.4254	1301.285
age	807.8653	100.3603	8.05	0.000	611.0613	1004.669
login18	6578.099	2106.601	3.12	0.002	2447.108	10709.09
woning	6648.834	3031.754	2.19	0.028	703.6425	12594.03
oplmet						
(Voortgezet) ..	1435.344	9761.389	0.15	0.883	-17706.49	20577.17
Kleuter-, lag..	6479.677	5613.836	1.15	0.249	-4528.909	17488.26
HAVO/VWO / pr..	13857.86	4784.585	2.90	0.004	4475.409	23240.3
MBO of het le..	-1320.064	3152.989	-0.42	0.675	-7502.994	4862.866
HBO (eerste o..	11509.46	4349.683	2.65	0.008	2979.841	20039.07
Wetenschappel..	9290.567	9590.406	0.97	0.333	-9515.971	28097.1
Did not have..)	103475.7	4622.174	22.39	0.000	94411.75	112539.7
aantalhh	742.179	3556.383	0.21	0.835	-6231.795	7716.153
aantalki	-1591.675	2601.204	-0.61	0.541	-6692.568	3509.218
sted	3273.164	1537.285	2.13	0.033	258.589	6287.739
woonvorm	-4041.712	3063.184	-1.32	0.187	-10048.54	1965.112
account	-4700.866	3379.97	-1.39	0.164	-11328.9	1927.167
kostwin	-5102.069	3515.533	-1.45	0.147	-11995.94	1791.8
partner	-8061.611	6958.301	-1.16	0.247	-21706.66	5583.437
_cons	-102891.2	19627.98	-5.24	0.000	-141381.2	-64401.25

Regression output for baseline model 2013

Linear regression

Number of obs	= 1,886
F(19, 1866)	= 22.47
Prob > F	= 0.0000
R-squared	= 0.0911
Root MSE	= 60645

networth2013	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
internallocus	1044.275	365.5707	2.86	0.004	327.3045	1761.245
externallocus	-1040.783	408.5898	-2.55	0.011	-1842.125	-239.4423
age	789.9916	88.25129	8.95	0.000	616.91	963.0732
login18	8090.129	1453.054	5.57	0.000	5240.348	10939.91
woning	6841.551	3004.768	2.28	0.023	948.4913	12734.61
oplmet						
(Voortgezet) speciaal onder..	404.9578	11914.81	0.03	0.973	-22962.8	23772.72
Kleuter-, lager- of basison..	5930.327	4893.126	1.21	0.226	-3666.248	15526.9
HAVO/VWO / pre-university e..	9330.977	4237.406	2.20	0.028	1020.423	17641.53
MBO of het leerlingwezen / ..	5477.586	3355.114	1.63	0.103	-1102.584	12057.76
HBO (eerste of tweede fase)..	14066.86	4398.133	3.20	0.001	5441.088	22692.64
Wetenschappelijk onderwijs ..	16456.21	6103.35	2.70	0.007	4486.105	28426.32
Did not have education (yet)	30230.94	4629.17	6.53	0.000	21152.04	39309.83
aantalhh	-3946.627	3486.291	-1.13	0.258	-10784.07	2890.813
aantalki	-54.6638	2995.793	-0.02	0.985	-5930.122	5820.794
sted	-273.8874	1169.462	-0.23	0.815	-2567.479	2019.704
woonvorm	1627.469	2296.8	0.71	0.479	-2877.098	6132.037
account	-3714.349	2993.364	-1.24	0.215	-9585.042	2156.344
kostwin	-6509.428	3144.073	-2.07	0.039	-12675.7	-343.1593
partner	1085.184	5433.042	0.20	0.842	-9570.293	11740.66
_cons	-97509.94	17844	-5.46	0.000	-132506.2	-62513.66

Regression output for baseline model 2011

Linear regression

Number of obs	1,899
F(19, 1879)	8.46
Prob > F	0.0000
R-squared	0.1044
Root MSE	71986

networth2011	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
internallocus	1961.026	440.0608	4.46	0.000	1097.967	2824.085
externallocus	-362.7498	591.4862	-0.61	0.540	-1522.789	797.2892
age	997.3101	153.5167	6.50	0.000	696.2289	1298.391
login18	6134.675	2061.918	2.98	0.003	2090.785	10178.57
woning	13371.48	3819.137	3.50	0.000	5881.284	20861.67
oplmet						
(Voortgezet) speciaal onderwijs (low)	-2633.413	8632.099	-0.31	0.760	-19562.92	14296.1
Kleuter-, lager- of basisonderwijs (low)	-281.3983	3913.346	-0.07	0.943	-7956.36	7393.563
HAVO/VWO (high)	11005.43	4601.956	2.39	0.017	1979.947	20030.91
MBO of het leerlingwezen (mediate)	5116.17	3477.25	1.47	0.141	-1703.508	11935.85
HBO (eerste of tweede fase) (high)	14448.85	5333.733	2.71	0.007	3988.188	24909.51
Wetenschappelijk onderwijs WO (high)	40931.93	9027.122	4.53	0.000	23227.7	58636.17
Did not have education (yet)	47269.25	18754.74	2.52	0.012	10486.93	84051.56
aantalhh	3.555552	3413.073	0.00	0.999	-6690.256	6697.367
aantalki	-774.3258	3036.189	-0.26	0.799	-6728.984	5180.332
sted	-1398.142	1263.187	-1.11	0.269	-3875.539	1079.255
woonvorm	-2688.994	1776.053	-1.51	0.130	-6172.237	794.2495
account	-9838.797	3669.633	-2.68	0.007	-17035.78	-2641.813
kostwin	-9818.778	4015.668	-2.45	0.015	-17694.42	-1943.14
partner	1473.235	6564.861	0.22	0.822	-11401.95	14348.42
_cons	-99268.83	24192.41	-4.10	0.000	-146715.7	-51822.01

Regression output for income robustness check

Linear regression

Number of obs	= 3,852
F(18, 3833)	= 11.37
Prob > F	= 0.0000
R-squared	= 0.0492
Root MSE	= 66458

networth2015	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
internallocus	1511.877	244.7869	6.18	0.000	1031.952	1991.802
externallocus	-4.805819	360.6917	-0.01	0.989	-711.9718	702.3602
age	597.7462	70.28201	8.50	0.000	459.9525	735.5399
woning	8116.231	2052.434	3.95	0.000	4092.263	12140.2
oplmet						
(Voortgezet) speciaal onderwijs / (co..	-1159.376	5982.528	-0.19	0.846	-12888.62	10569.87
Kleuter-, lager- of basisonderwijs / ..	3156.313	3681.399	0.86	0.391	-4061.374	10374
HAVO/VWO / pre-university education	12068.64	3428.106	3.52	0.000	5347.55	18789.72
MBO of het leerlingwezen / senior voc..	-1328.867	2004.292	-0.66	0.507	-5258.447	2600.714
HBO (eerste of tweede fase) / vocatio..	7586.456	2900.234	2.62	0.009	1900.306	13272.61
Wetenschappelijk onderwijs WO / unive..	7768.392	6409.271	1.21	0.226	-4797.516	20334.3
Did not have education (yet)	57230.57	26896.67	2.13	0.033	4497.411	109963.7
aantalhh	-257.004	2139.507	-0.12	0.904	-4451.685	3937.677
aantalki	-1503.342	1693.751	-0.89	0.375	-4824.081	1817.398
sted	2411	1007.961	2.39	0.017	434.8096	4387.191
woonvorm	-2905.086	1831.212	-1.59	0.113	-6495.328	685.1564
account	-3581.862	2230.678	-1.61	0.108	-7955.293	791.5679
kostwin	-3970.272	2307.559	-1.72	0.085	-8494.433	553.8882
partner	-4192.568	4479.821	-0.94	0.349	-12975.63	4590.493
_cons	-26250.52	7754.351	-3.39	0.001	-41453.57	-11047.48

Regression output for age robustness check

Linear regression

Number of obs	= 2,814
F(20, 2793)	= 11.19
Prob > F	= 0.0000
R-squared	= 0.0532
Root MSE	= 80992

networth2015	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
internallocus	1388.247	317.8344	4.37	0.000	765.0332	2011.461
externallocus	561.2598	498.4894	1.13	0.260	-416.185	1538.705
age	829.2639	100.3861	8.26	0.000	632.4253	1026.102
dumage	-10473.84	5954.505	-1.76	0.079	-22149.51	1201.839
login18	5050.179	2192.239	2.30	0.021	751.6061	9348.751
woning	5739.669	2919.994	1.97	0.049	14.10588	11465.23
oplmet						
(Voortgezet) speciaal onderwijs / (con..	836.2099	9837.13	0.09	0.932	-18452.57	20124.99
Kleuter-, lager- of basisonderwijs / k..	8568.584	5491.289	1.56	0.119	-2198.81	19335.98
HAVO/VWO / pre-university education	13688.42	4778.442	2.86	0.004	4318.781	23058.05
MBO of het leerlingwezen / senior voca..	-1178.266	3130.379	-0.38	0.707	-7316.356	4959.825
HBO (eerste of tweede fase) / vocation..	12569.41	4363.451	2.88	0.004	4013.494	21125.32
Wetenschappelijk onderwijs WO / univer..	11029.61	9828.983	1.12	0.262	-8243.196	30302.41
Did not have education (yet)	11108.46	7937.839	1.40	0.162	-4456.162	26673.09
aantalhh	2269.591	2823.58	0.80	0.422	-3266.923	7806.105
aantalki	-1548.251	1924.66	-0.80	0.421	-5322.15	2225.648
sted	3874.712	1548.661	2.50	0.012	838.0752	6911.348
woonvorm	-5197.178	2700.576	-1.92	0.054	-10492.5	98.14807
account	-4805.089	3356.789	-1.43	0.152	-11387.13	1776.949
kostwin	-5314.175	3497.673	-1.52	0.129	-12172.46	1544.111
partner	-9767.315	6402.211	-1.53	0.127	-22320.86	2786.229
_cons	-82534.15	19918.12	-4.14	0.000	-121589.9	-43478.42

Regression output for other housing robustness check

Linear regression

Number of obs	= 2,377
F(20, 2356)	= 342.11
Prob > F	= 0.0000
R-squared	= 0.0585
Root MSE	= 79182

networth2015	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
internallocus	1633.247	330.4081	4.94	0.000	985.3257	2281.167
externallocus	208.0436	554.8236	0.37	0.708	-879.9497	1296.037
age	806.966	100.3964	8.04	0.000	610.0915	1003.84
login18	6560.228	2102.977	3.12	0.002	2436.35	10684.11
woning	6648.394	3031.196	2.19	0.028	704.3056	12592.48
dumwoning	-35400.03	14111.6	-2.51	0.012	-63072.48	-7727.584f
oplmet						
(Voortgezet) speciaal onderwijs / (continue..	1429.059	9762.758	0.15	0.884	-17715.43	20573.55
Kleuter-, lager- of basisonderwijs / kinder..	6492.75	5594.335	1.16	0.246	-4477.581	17463.08
HAVO/VWO / pre-university education	13897.95	4770.247	2.91	0.004	4543.631	23252.27
MBO of het leerlingwezen / senior vocationa..	-1343.542	3140.803	-0.43	0.669	-7502.566	4815.482
HBO (eerste of tweede fase) / vocational co..	11477.75	4336.293	2.65	0.008	2974.402	19981.1
Wetenschappelijk onderwijs WO / university ..	9279.345	9591.991	0.97	0.333	-9530.276	28088.97
Did not have education (yet)	103452.2	4608.776	22.45	0.000	94414.48	112489.8
aantalhh	773.1391	3526.757	0.22	0.826	-6142.731	7689.009
aantalki	-1620.104	2582.2	-0.63	0.530	-6683.724	3443.517
sted	3272.408	1536.584	2.13	0.033	259.2104	6285.606
woonvorm	-4069.323	3028.347	-1.34	0.179	-10007.82	1869.178
account	-4704.603	3376.548	-1.39	0.164	-11325.92	1916.712
kostwin	-5100.807	3512.137	-1.45	0.147	-11988.01	1786.392
partner	-8063.364	6939.027	-1.16	0.245	-21670.6	5543.87
_cons	-102653.5	19595.52	-5.24	0.000	-141079.8	-64227.27