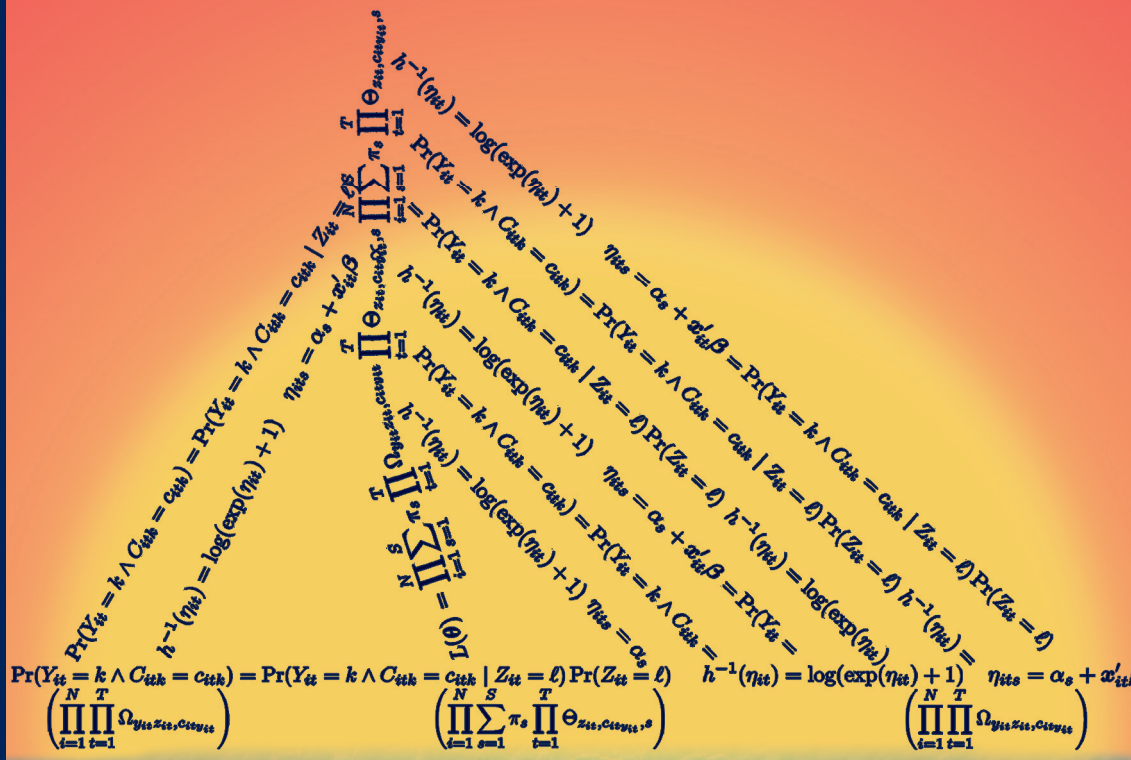
NIENKE DIJKSTRA

Statistical Modeling Approaches for Behavioral and Medical Sciences



Statistical Modeling Approaches for Behavioral and Medical Sciences

Statistical Modeling Approaches for Behavioral and Medical Sciences

Statistische modellen voor gedrags- en gezondheidswetenschappen

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
Rector Magnificus

Prof.dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on

Thursday 20 October 2022 at 13:00 hours

by

Nienke Fransje Simone Dijkstra
born in Capelle aan den IJssel, The Netherlands

Erasmus University Rotterdam

Doctoral committee

Promotors:	Prof.dr. P.J.F. Groenen Prof.dr. H. Tiemeier Prof.dr. A.R. Thurik
Other members:	Prof.dr. D. Fok Prof.dr. P.W. Jansen Prof.dr. M.J. de Rooij

Erasmus Research Institute of Management - ERIM
The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam
Internet: www.erim.eur.nl

ERIM Electronic Series Portal: repub.eur.nl

ERIM PhD Series in Research in Management, 539
ERIM reference number: EPS-2022-539-S&E
ISBN 978-90-5892-641-8

Cover image: Mirakels Ontwerp, Miranda Dood
Cover design: PanArt, www.panart.nl
Print: OBT bv, www.obt.eu

This publication (cover and interior) is printed on FSC® paper Magno Satin MC.

Table of contents

$$\Pr(Y_{it}=k\wedge C_{itk}=c_{itk})=\Pr(Y_{it}=k\wedge C_{itk}=c_{itk}\mid Z_{it}=\ell)\Pr(Z_{it}=\ell)\quad h^{-1}(\eta_{it})=\log(\exp(\eta_{it})+1)\quad \eta_{its}=\alpha_s+x_{it}'\beta\quad \Pr(Y_{it}=k\wedge C_{itk}=c_{itk})=$$

$$\left(\prod_{i=1}^{N}\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{it}w_{it}}\right)\qquad \left(\prod_{i=1}^{N}\sum_{s=1}^{S}\pi_s\prod_{t=1}^{T}\Theta_{z_{it},c_{it}w_{it},s}\right)\qquad \left(\prod_{i=1}^{N}\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{it}w_{it}}\right)$$

CHAPTER 1

Introduction

$= \Pr(Y_{it} = k \wedge C_{ith} = c_{ith} \mid Z_{it} = \ell) \Pr(Z_{it} = \ell) \quad h^{-1}(\eta_{it}) = \log(\exp(\eta_{it}) + 1) \quad \eta_{its} = \alpha_s + x'_{it}\beta \quad \Pr(Y_{it} = k \wedge C_{ith} = c_{ith}) = \Pr(Y_{it} =$

1.1 Motivation

Behavior and abilities are arguably the most fundamental concepts in life. Behavior is observable through someone's actions. Underlying a person's behavior are their beliefs and attitudes towards someone or something. Abilities are the personal qualities and skills needed to do something, in as far as they are innate traits and are not acquired through formal education.

The concepts behavior and abilities, and their causes and consequences, are studied in many disciplines, such as economics, psychology, sociology, and health sciences. For example, Thaler (2016) discusses the field of behavioral economics where the rational agent with well-defined preferences and unbiased beliefs, also known as the Home economicus, is replaced by the Homo sapiens or Humans; agents that sometimes show irrational and less optimal behavior.

Furthermore, behavioral and social studies have repeatedly shown that surveys and experiments are good predictors of actual behaviors. For example, patient individuals, generally, have higher savings and educational attainment, and social preferences are correlated with prosocial behaviors, like donating, volunteering, and helping others (Falk et al., 2015). Moreover, general risk tolerance, measured with either questionnaires or experimental tasks, is a good predictor of risky behaviors such as portfolio allocation, occupational choice, substance usage, and becoming an entrepreneur (Beauchamp et al., 2017; Dohmen et al., 2011).

Also a person's abilities are important to study as they partly define who we are and what we can achieve. In particular, cognitive ability is related to educational attainment, occupational success, health, and life expectancy (Batty et al., 2007; Gottfredson, 1997; Junger and van Kampen, 2010). As most abilities are innate traits, they are inheritable which underlines the importance of genetic studies. It is, for instance, well known that heredity, substantially, accounts for differences in cognitive abilities, like verbal and spatial reasoning (see, e.g., Plomin and DeFries, 1998). More recently, studies attempt to find which genes are responsible for the heritability of cognitive abilities by means of genomewide association studies (GWAS, Davies et al., 2011; Davis et al., 2010; Okbay et al., 2016; Plomin et al., 2013).

1.2 Research Topics

The research of behaviors and abilities is quickly changing. In the last decade, more and more emphasis is put on data collection and the benefits of large data sets have become more evident. Hence, like in many other fields, the data sets are rapidly growing in size, both in terms of observations and variables. Moreover, longitudinal studies gained increasing popularity. Following individuals for a longer period has been more and more appreciated. Many studies have been launched to investigate the development, movement, and progress of individuals on a wide range of domains. A few examples of large (population based) data sets including behavioral and medical data are the Adolescent Brain Cognitive Development (ABCD) Study (Garavan et al., 2018), the UK Biobank (Sudlow et al., 2015), the Generation R Study (Kooijman et al., 2016), the Rotterdam Study (also known as ERGO, Ikram et al., 2017), and

the Health and Retirement Study (Sonnega et al., 2014). This thesis includes data from both the ABCD Study and the Generation R Study.

An important example of the change in research of behaviors and abilities is the research on the connection between genes and behavior. For example, Linnér et al. (2019) conducts, in a sample of over one million individuals, a GWAS of general risk tolerance and risky behaviors, such as dangerous driving, smoking, and unsafe sex. Furthermore, the genetic association between educational attainment and general intelligence has been widely studied (Okbay et al., 2016; Rietveld et al., 2014; Sniekers et al., 2017).

Also the analysis of brain data, and in particular the analysis of data obtained from Magnetic Resonance Images (MRI), has flourished. Whereas previously brain data, typically grouped in anatomically divided regions, was analyzed through a simple linear regression, nowadays researchers apply novel machine learning techniques. Another development and benefit of the usage of machine learning techniques is the switch from the univariate perspective, where researchers concentrate on one region or vertex (i.e., a derivative of a voxel, a 3D pixel) in the brain at the time, to the multivariate perspective, where the whole brain (approximately 300,000 vertices) is analyzed at once. A convolutional neural network is a perfect example of a machine learning technique that gained popularity and is particularly well suited for analyzing images (Messina et al., 2021; Seong et al., 2018; Sturmfels et al., 2018).

The increase in number of variables (i.e., going from univariate approaches to multivariate approaches and going from aggregated regions to many more vertices) requires larger sample sizes for studies to retain the statistical power to be able to accurately detect small effect sizes (Arbabshirani et al., 2017). For example, Jollans et al. (2019) shows the performance of several machine learning techniques on simulated neuroimaging data with various sample sizes, number of features, and effect sizes, and advocates for a sample size larger than $N = 400$ observations. Moreover, Peng et al. (2021) acknowledges the problem of small samples and shows that with $N = 14503$ real structural MR images the age of individuals can be accurately predicted using a convolutional neural network.

The increase in sample sizes enables the use of advanced statistical methods. The larger the data sets the more they allow for the estimation of flexible and complex models. For example, the power of a small data set may not be enough to establish an interaction effect, whereas the effect could become evident in a larger data set. In addition, a large data set enables to include individual or group specific effects. The present thesis employs modern statistical models in the field of behavioral and medical sciences with the aim to generate more reliable inferences and more accurate predictions than would have been the case with established methods.

1.3 Part I: Risk Behavior Measured With Experimental Tasks

This thesis is divided in two parts. In the first part, I develop the Censored Mixture Model (CMM) to analysis risk behavior and show two applications. In Chapter 2,

the CMM is presented and its performance on a data set containing 3404 children who played the Columbia Card Task (CCT) is demonstrated. Chapter 3 emphasizes the difference between decision making under risk and uncertainty. Again using the CCT and a modified version of the CMM, this chapter associates individual-level aspects of entrepreneurship with both decision making under risk and uncertainty.

Questionnaires and experimental tasks are widely used to measure someone's risk attitude. Many experimental tasks have been developed, ranging from lottery tasks, where participants choose to participate in different lotteries, to gambling tasks, where participants bet on a outcome, to sequential risk tasks, where participants repeat a certain action. Two widely used sequential risk tasks are the Balloon Analogue Risk Task (BART, Lejuez et al., 2002) and Columbia Card Task (CCT, Figner et al., 2009), where participants repeatedly inflate a balloon or turn over cards, respectively. Both the payoff and the risk increases the longer a participant continues. These tasks have shown to be able to predict real-world risk taking behavior (see e.g., Lejuez et al., 2003; Pripfl et al., 2013). However, modeling these tasks comes with a few challenges.

First, the analyses are often based on the assumption of a smooth (normal) distribution of the residuals. However, certain outcomes are more attractive to participants than others. For example, in some sequential risk tasks participants have to select a number of repetitions of a certain action, this number indicates the level of risk someone is prepared to take (e.g., the number of pumps in the BART or the number of cards turned over in the CCT). It is well known that even numbers and multiples of five are more often selected than odd numbers (Baird et al., 1970). This pattern leads to inflated values in the outcome distribution.

The second challenge concerns censored observations. By definition, most sequential risk tasks may randomly end prematurely, such as when the balloon pops in the BART or when a loss card is encountered in the CCT, leading to censored observations. Typically, the researcher is interested in the level of risk a participant is willing to take and the censoring obscures this. Treating the censored observations as uncensored would lead to biases in the results.

Third, a priori unknown groups of participants can react differently to certain risk-levels. Similarly, individuals may have different tendencies towards rewards and/or losses. The model should be flexible enough to allow for these unobserved individual behaviors.

Chapter 2 proposes the Censored Mixture Model (CMM) to solve these three potential problems emerging with modeling risk taking. This chapter describes the CMM and includes an application with the Columbia Card Task.

Another situation where risk taking becomes important is for entrepreneurs. Chapter 3 contains an application of the CMM showing the relationship between individual entrepreneurial orientation (IEO) and both decision making under risk and uncertainty in a student sample. The importance of studying entrepreneurial orientation on an individual level has been demonstrated repeatedly. Bolton (2012) already show in a sample of students that IEO and its subscales positively associate with intentions for entrepreneurship, a finding that is supported by Koe (2016) and Sahoo and Panda (2019). Hence, IEO is considered to be a determinant of actual

entrepreneurial involvement.

Furthermore, risk and uncertainty are considered core concepts in entrepreneurship theories (Kihlstrom and Laffont, 1979; Kirzner, 1973). However, the empirical evidence about the relationships between risk and uncertainty preferences on the one hand and entrepreneurship on the other hand is rather mixed as a result of the use of artificial tasks, possibly lacking external or ecological validity, such as self-reports and lotteries (for recent reviews see Holm et al., 2013; Koudstaal et al., 2016).

To bypass these limitations, in Chapter 3, two versions of the CCT are conducted in a student sample. The advantage of the CCT is that with a minor adjustment of the task, decision making under risk as well as uncertainty can be elicited. Moreover, the CCT provides immediate feedback, eliciting affective engagement.

1.4 Part II: Predicting Cognitive Ability Using Brain Morphology

In the second part of this thesis, I compare two machine learning approaches with various simpler commonly used baseline approaches in their performance to predict cognitive ability using brain morphology. Cognition is one of the most important functions of the brain and many studies have associated intellectual ability to neuroimaging features, such as brain morphology (Fields, 2008; Reiss et al., 1996). In particular, it is repeatedly demonstrated that cortical thickness and cortical surface area are predictive of cognitive ability (Colom et al., 2013; Karama et al., 2009; Menary et al., 2013; Vuoksimaa et al., 2016).

Such associations between brain morphology and the performance on neuropsychological tests are called structure-function associations. Understanding this link between features of brain structure and neuropsychological tests at population level, can provide insights in typical brain development. Once the baseline for typical brain development is established, we can examine atypical brain development related to cognitive deficit. Moreover, understanding the neurodevelopment at population level allows for the identification of abnormal neurobiological processes in child psychopathology.

Much of the existing literature has been concentrated on either parcellating the brain into anatomically divided regions, known as regions of interest (ROIs) or analyzing the brain in a vertex-wise framework, where every vertex is analyzed individually. The first approach has been criticized for the assumption that any underlying association with a phenotype is constrained by anatomical boundaries and for the loss of crucial information when summarizing across large areas of neural tissue. The shortcoming of the second approach is that effects of multiple vertices combined are ignored and that not all detailed information is used for predicting the phenotype, which is most likely disadvantageous to the predictive performance.

To overcome these two shortcomings, a twofold approach is proposed in Chapter 4. First, detailed information about the whole brain (i.e., almost 150000 vertices per hemisphere) is analyzed, to release the assumption that a particular cognitive phenotype spatially correlates with the brain based on arbitrary anatomical parcellations.

Second, machine learning techniques that capture all information in a multivariate way are applied, to improve the predictions by analyzing all available information at once.

Chapter 4 concentrates on two machine learning techniques, namely adaptive lasso and better subset regression, with the purpose to predict non-verbal IQ. The predictive performance of these techniques is compared with the predictive performance of two commonly used methods that suffer from the criticism described above. First, the aggregated regions of interest are implemented in a multivariate regression. The downside of this approach is that detailed information is summarized based on anatomical boundaries, potentially leading to the loss of crucial information. The second technique is to implement many univariate regressions in which one vertex at the time is used to predict the phenotype non-verbal IQ. Evidently, a multivariate approach in which all vertices are included simultaneously leads to better (or at least equally good) predictions.

1.5 Individual Contributions and Publication Status

My PhD was funded by the Research Excellence Initiative (REI) of the Erasmus University Rotterdam grant awarded to Professors Thurik, Groenen, Tiemeier and Franken (project number 265.403). This section discusses my contributions to the chapters in the present thesis. The current chapter (Chapter 1) and Chapter 5, I wrote independently, although I received valuable feedback from my supervisors and Dr. Muetzel.

The research idea of Chapter 2 was proposed by Professor Tiemeier and he had a supervisory role throughout the project. The manuscript is written by myself and I conducted the analysis. Professor Groenen advised in the technical details. Dr. Figner helped positioning the paper within the psychological literature on risk taking and contributed in final checks of the paper. Also, Professor Thurik provided valuable textual feedback on the manuscript. The data acquisition was done by the Generation R team.

The basis of Chapter 3 was the CMM and the data collected under supervision of Kristel de Groot. Together with Dr. Rietveld, we developed the research question and wrote the manuscript. I conducted the data analyses and was mainly responsible for writing the methods and results section. Thereafter, we alternately revised and improved the manuscript.

Dr. Muetzel proposed to me to apply machine learning techniques to neuroimaging data (Chapter 4). Together with Dr. Muetzel, I wrote the analysis plan which was, thereafter, discussed and evolved further with Professors Groenen and Tiemeier. I performed the statistical analyses and wrote the manuscript. Dr. Muetzel and Professors Groenen and Tiemeier had a supervisory role and provided valuable feedback, as did Professor Thurik.

The publication status of each chapter is shown below. Additionally, an overview of where I have presented the projects during my PhD is printed.

Chapter 2 A Censored Mixture Model for Modeling Risk Taking

- Presented at Data Science, Statistics, and Visualization (DSSV, 2018), Compstat (2018), Society for Epidemiologic Research (SER, 2019), International Meeting of Psychometric Society (IMPS, 2019), and Tilburg University (2019)
- Published in Psychometrika

Chapter 3 Entrepreneurial Orientation and Decision Making Under Risk and Uncertainty: Experimental Evidence From the Columbia Card Task

- Presented at Erasmus University (2020)
- Manuscript submitted to Applied Psychology

Chapter 4 Predicting Cognition of Children Through Cortical Surface Area

- Manuscript in preparation

PART I

Risk Behavior Measured With Experimental Tasks

$$\Pr(Y_{it}=k\wedge C_{itk}=c_{itk})=\Pr(Y_{it}=k\wedge C_{itk}=c_{itk}\mid Z_{it}=\ell)\Pr(Z_{it}=\ell)\quad h^{-1}(\eta_{it})=\log(\exp(\eta_{it})+1)\quad \eta_{its}=\alpha_s+x_{it}'\beta\quad \Pr(Y_{it}=k\wedge C_{itk}=c_{itk})=$$

$$\left(\prod_{i=1}^{N}\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{it}w_{it}}\right)\qquad \left(\prod_{i=1}^{N}\sum_{s=1}^{S}\pi_s\prod_{t=1}^{T}\Theta_{z_{it},c_{it}w_{it},s}\right)\qquad \left(\prod_{i=1}^{N}\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{it}w_{it}}\right)$$

CHAPTER 2

A Censored Mixture Model for Modeling Risk Taking

$= \Pr(Y_{itt} = k \wedge C_{itth} = c_{itth} \mid Z_{itt} = \ell) \Pr(Z_{itt} = \ell) \quad h^{-1}(\eta_{itt}) = \log(\exp(\eta_{itt}) + 1) \quad \eta_{ittg} = \alpha_g + x'_{itt}\beta \quad \Pr(Y_{itt} = k \wedge C_{itth} = c_{itth}) = \Pr(Y_{itt} =$

Abstract

Risk behavior has substantial consequences for health, well-being, and general behavior. The association between real-world risk behavior and risk behavior on experimental tasks is well documented but their modeling is challenging for several reasons. First, many experimental risk tasks may end prematurely leading to censored observations. Second, certain outcome values can be more attractive than others. Third, a priori unknown groups of participants can react differently to certain risk-levels. Here, we propose the Censored Mixture Model (CMM) which models risk taking while dealing with censoring, attractiveness to certain outcomes, and unobserved individual risk preferences, next to experimental conditions.

2.1 Introduction

Taking a particular risk can have substantial consequences on health, well-being, and general behavior and, hence, is examined in many scientific fields, such as psychology, criminology, and economics. Risk taking can be measured by surveys or using experimental tasks. Although risk behavior across different experimental tasks does not highly correlate (Pedroni et al., 2017), studies have shown a moderate, but meaningful association between risk behavior measured in various experimental tasks and real-world risk taking. For example, Lejuez et al. (2003) and Pripfl et al. (2013) show that smokers take significantly more risk on, respectively, the Balloon Analogue Risk Task (BART) and the Columbia Card Task (CCT) than non-smokers. Likewise, using a survey, Collins et al. (1987) show a relationship between risk taking/rebelliousness and smoking at an older age.

There are four types of experiments commonly used to measure risk behavior. The first is based on lotteries, where an explicit description of the outcome and probabilities are given. Typically, participants have to state their preference between, for example, option A: 50% chance of winning 10 euro, and option B: 30% chance of winning 30 euro. Typically, in these tasks it is straightforward to decompose the underlying constructs of risk taking. However, they are often criticized for being too artificial and lacking external validity.

An example of the second type of experiments is the Iowa Gambling Task (Bechara et al., 1994), where participants can win or lose money by picking (many) cards from four decks, each card having a win and loss value. The expected value and probability distribution of the values of the cards in the four decks are unknown to participants at the start, but can be learned during the task. This task has shown to successfully predict real-world risk taking behavior. However, it is virtually impossible to decompose the underlying constructs of participants' risk taking behavior, such as risk preferences (Schonberg et al., 2011). Risk preferences are confounded with the learning curve, because participants have to unravel the expected return and the probability distribution of the decks while playing the game. In addition, it is difficult to distinguish whether the behavior is driven by risk attitude or sensitivity to reward or punishment.

The third type of task paradigm is based on gambling and includes, among others, the Cambridge Gambling Task (Rogers et al., 1999) and the Game of Dice Task (Brand et al., 2005), where participants have to bet on the color of randomly drawn cards or on the outcome of a roll of a dice, respectively. The probability of the possible outcomes is known, so there is no learning effect present. However, these tasks have the disadvantage that they do not allow to disentangle the effects of risk and of attractiveness of a higher expected pay-off value.

Lastly, in sequential risk tasks participants are asked to repeat a certain action. These tasks include, among others, the Balloon Analogue Risk Task (BART, Lejuez et al., 2002), the Columbia Card Task (CCT, Figner et al., 2009), the Angling Risk Task (Pleskac, 2008), and the Devil Task (Slovic, 1966), where the repeated actions are inflating a balloon, turning over cards, catching fish, and pulling knife switches, respectively. The risk increases the longer a participant continues. Although these

sequential risk tasks do not suffer from the issues described under the first three types of experiments, they have their own challenges, which makes modeling risk taking complex. Below we discuss two of these challenges.

First, the analyses are often based on the assumption of a smooth (normal) distribution of the residuals. However, certain outcomes are more attractive to participants than others. For example, in some sequential risk tasks participants have to select a number of repetitions of a certain action, this number indicates the level of risk someone is prepared to take (e.g., the number of pumps in the BART, the number of cards turned over in the CCT, or the number of fish caught in the Angling Risk Task). It is well known that even numbers and multiples of five are more often selected than odd numbers (Baird et al., 1970). This pattern leads to inflated values in the outcome distribution. Similarly, within surveys some outcomes tend to be more attractive than others. Imagine a longitudinal study measuring drug usage. Asking people how often they use drugs, typically, also leads to even numbers and multiples of five and ten (Klesges et al., 1995).

The second challenge of sequential risk tasks concerns censored observations. In the imaginary longitudinal drug study, participants can be easily lost, leading to incomplete information and censored observations. Moreover, most sequential risk tasks by definition may randomly end prematurely, such as when a loss card is encountered in the CCT. Typically, the researcher is interested in the level of risk a participant is willing to take and the censoring obscures this.

To deal with censored observations, Lejuez et al. (2002) suggest to use the adjusted score in the BART (average inflations over the unpopped balloons). However, Pleskac et al. (2008) have shown that this score is biased and propose the automatic BART, where participants have to choose a number of inflations before the trial starts and censoring is no issue. Figner et al. (2009) argue that people behave differently in a preplanned (i.e., the automatic BART) and impulsive (i.e., the original BART) decision making situation and developed the Columbia Card Task (CCT) to investigate the difference between deliberative and affective decision making. The CCT is a computer-based card game, and participants can win or lose money by turning over cards. A major advantage of the CCT, over the other sequential risk tasks, is that the CCT orthogonally varies the risk-relevant factors gain amount, loss amount, and loss probability. Moreover, these factors are known to the participant, so the risk taking is not confounded with the effect of learning these values.

So far, none of the existing studies have provided a statistical model that addresses all the issues introduced above, that is censored observations, attractiveness to certain outcomes, and unobserved heterogeneity. However, there are two studies worth mentioning. First, Wallsten et al. (2005) introduce a learning model for the BART that accommodates censoring. This model, however, is not directly applicable to tasks where the probability of losing is known, like the CCT and some tournaments in the Angling Risk Task, because it is built on the fact that participants have prior believes about these probabilities and update their believes throughout the game. Second, Weller et al. (2019) implement an interesting model that addresses the censoring in the CCT and allows individual unobserved heterogeneity through a random effects model. A potential disadvantage of this model is the assumption on

the distribution of the random effect. We propose a model that makes less parametric assumptions and models the unobserved heterogeneity in a more free way. We refer also to this paper for its excellent overview of the literature on risk decision making processes.

In the present study, we propose a Censored Mixture Model (CMM). This model is a specific form of survival analyses that are used often in scientific studies in medicine, sociology, and econometrics to predict the duration of time until a specific event, such as death, cancer diagnosis, divorce, graduation from school, or finding a job. A common trait across these studies is that when the data collection ends, the event has occurred for some individuals, but not for others. For example, some couples divorce during the study, while others do not. These couples may divorce later, but, by the time the data are analyzed, it is unknown when. These incomplete observations are called censored observations. In our CMM, censored observations are included in the model by using the information that the participant intended to take more risk than the observed risk level. The attractiveness of certain patterns in outcome values is covered by assigning additional probability mass to the inflated values in the distribution. Furthermore, we choose to model the unobserved individual tendency for risk taking through finite mixtures, which can approximate a variety of distributions. In addition, we choose a link function such that the regression coefficients have a linear interpretation on the interval $[0, \inf\rangle$. In short, the CMM models risk taking while dealing with censoring, attractiveness to certain outcomes, and unobserved individual risk preferences, next to analyzing the experimental conditions and having a straightforward linear interpretation of the effects.

Our model can be used to analyze all sequential risk tasks. In the present study, we analyze the hot version of the CCT, which measures affective decision making. In addition, results of the CMM applied to a BART data set are added In Appendix 2.D. The data of the CCT were collected as part of the Generation R Study which was designed to analyze early environmental and genetic determinants of growth, development, and health from fetal life until young adulthood. A total of 4538 nine-year old children participated in the CCT.

The remainder of this paper is structured as follows. It starts with a detailed explanation of the CCT and its challenges when modeling risk taking. Next, the data are discussed by means of the data collection process, cleaning procedure, and their characteristics. Subsequently, the structure of the model is discussed extensively. Last, we present the results and we will discuss the limitations.

2.2 Columbia Card Task

The Columbia Card Task is shown in Figure 2.1. There are 32 cards divided in win cards (happy faces) and loss cards (unhappy faces). At the beginning of a game round all cards are face down and participants are asked to turn over cards. By turning over a win card, the participant earns points and by turning over a loss card they lose points and the current game round ends. At every step, the participant has the choice between turning over another card and pressing the stop button to voluntarily stop this game round. It is also possible to stop immediately without turning over

any card. After a game has ended, the earned points are summed and the potential loss amount is subtracted.

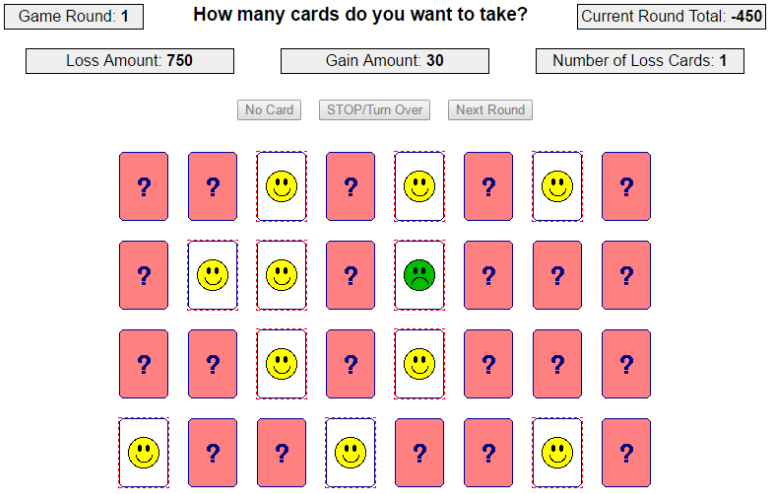


Figure 2.1: A screenshot of the first game round in the Columbia Card Task with the game settings: gain amount equal to thirty, loss amount equal to 750, and number of loss cards equal to one. In this game round, the participant first turned over ten win cards (happy faces). The eleventh card was a loss card (sad face), resulting in a total score in the current game round of $10 \times 30 - 750 = -450$.

The gain amount (points earned by turning over a win card), loss amount (points lost by turning over a loss card), and number of loss cards vary per round: the gain amount is either ten or thirty, the loss amount is either 250 or 750, and there are either 1 or 3 loss cards in the game (Figner and Weber, 2011). These three experimental conditions are displayed at the top of the screen and are known to the participant. Note that, in contrast to Figner et al. (2009), the loss cards are randomly distributed over the 32 cards. These three parameters lead to eight different game settings and within a block of eight trials the sequence of the game settings is random. Every participant plays at least two blocks of eight trials[1]. In other words, every game setting is played at least twice. Because of the different game settings, the CCT measures next to risk taking also the complexity of information use and the sensitivity to reward, punishment, and probability. With the three parameters (gain amount, loss amount, and number of loss cards) is it possible to assess which of these three parameters affects participants' choices.

The indicator for risk taking is the number of cards a participant intends to turn over. However, if a participant faces a loss card the game ends prematurely, the trial is censored and it is unknown how many cards a participant intended to turn over.

[1] At the beginning of the data collection we decided to shorten the test. Instead of three blocks with in total 24 trials, two blocks with in total 16 trials were played. There are 388 children, who played 24 trials. For the analysis only the first 16 trials of these children are used.

This should be considered in the analysis. Figner et al. (2009) manipulate the game such that in most trials the loss card is the last possible card to turn over and only analyze the uncensored trials. However, for this manipulation not to be discovered by participants, Figner et al. (2009) included extra trials where the loss card appeared at an early stage of the trial. This approach has several drawbacks. Besides the serious problem that such a setup uses deception, letting participants play extra rounds has the important disadvantage of being time consuming and hence more expensive. In addition, we show that the result in the previous trial effects the behavior in the current trial. Not correcting for the negative shock a loss card might have, could affect the results.

Another issue that should be accounted for is the attractiveness of certain outcome values. Figure 2.2 shows the distribution of the outcome, in this case, the number of cards turned over. The left graph only includes the uncensored trials and shows peaks at certain number of cards. The right panel suggests that the peaks are independent of the probability of being censored, because the censored trials (lower bars) do not show any irregular or unexpected values. Three categories of peaks can be distinguished. The first category is the excess of zeros. This inflation is probably caused by children who are very much risk averse and prefer not to play the game. The second category includes the peaks at four, eight, ten, twelve, sixteen, twenty, and twenty-four. Although ten is not a multiple of four, it seems to be an attractive number similar to the multiples of four; hence, it is included in this set. Recall the layout of the CCT from Figure 2.1, creating a geometric pattern, such as complete rows or columns, corresponds to turning over a number of cards equal to a multiple of four. The third category of excesses occurs with participants who are very risk seeking: if you managed to turn over 30 cards without facing a loss card, then why not as well try the 31st card? Note that Categories 1, 2, and 3 are also subsumed in Category 4 and that the weight of the observation is equally split over Category 1, 2, or 3 and Category 4.

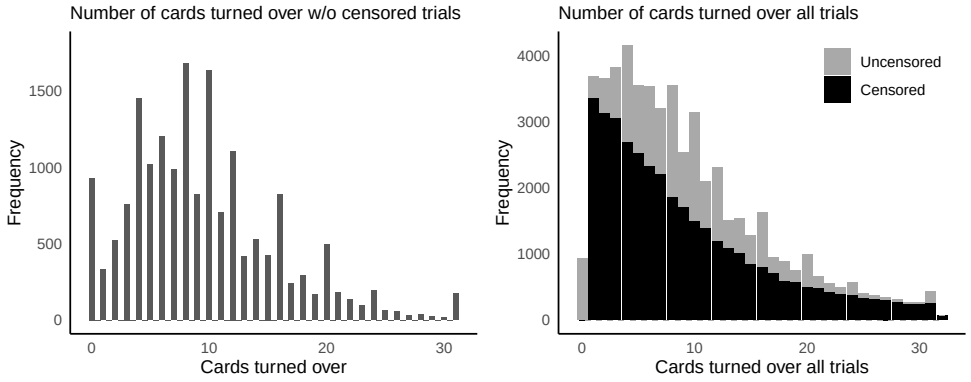


Figure 2.2: Distribution of the number of cards turned over.

To demonstrate that these categories of inflated values are independent over the

individuals, we compute the distribution of the outcomes per person over four categories. Figure 2.3 provides a parallel coordinates plot with these individual distributions in gray and the average over all distributions in black. Strict independence would be found if the gray lines coincide with the mean line in black. Since the individual distributions are relatively similar, we argue that the inflated values are independent over the individuals.

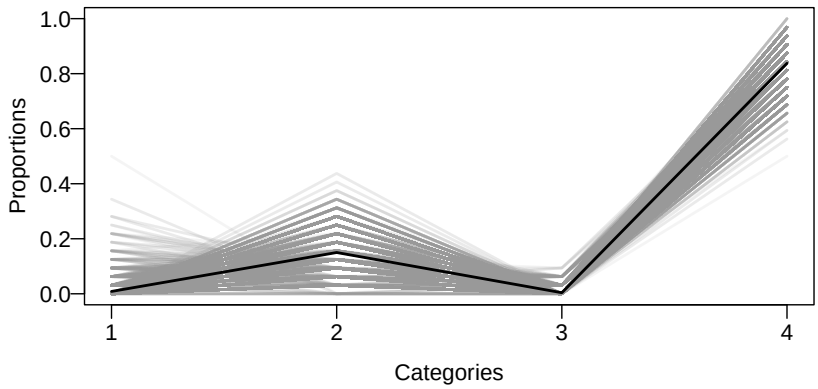


Figure 2.3: A parallel coordinates plot on proportions of outcomes in four categories: (1) zero cards turned over, (2) multiples of four, (3) 31 cards turned over, and (4) all possible outcomes (i.e., $\{0, 32\}$). Note that Categories 1, 2, and 3 are also subsumed in Category 4 and that the weight of the observation is equally split over Category 1, 2, or 3 and Category 4. The individual distributions are in gray and the average over all distributions is in black.

2.3 Data

What sets this research apart from previous studies, besides addressing all issues involved with modeling risk taking, is the large number of participants. The large sample size allows us to build a more flexible model that handles censoring, categorical background variables, such as individual characteristics, experimental conditions, and attractiveness to certain patterns and outcome values.

The current study is embedded in the Generation R Study, a large population based multi-ethnic cohort study (Kooijman et al., 2016). The Generation R Study was designed to analyze early environmental and genetic determinants of growth, development, and health from fetal life until young adulthood. The data collection is intense and includes multiple surveys with biological and observational assessments. The CCT is one of the observational assessments that was conducted on nine-year-old children (age 9.8 $\pm$0.26). The cohort includes almost ten thousand children at birth, of which 4538 children participated in the CCT. The data set is partitioned in a training set of 3404 children and an (prior to analysis) unseen test set of 1134 children.

The Generation R study has an open policy in regard to collaboration with other research groups (`http://www.generationr.nl/researchers/collaboration.html`). Requests for data access and collaboration are discussed in the Generation R Study Management Team.

The CCT was conducted as part of a series of assessments taking approximately three hours. At the beginning of the CCT, children were told that they would be rewarded with money based on their performance on the CCT. After all trials were played, three trials were randomly selected and were paid out in real money. The children had a start value of 200 cents (i.e., 2 euro) and the total points of the selected trials were added or deducted from this start value. Children could receive money, but did not have to pay any net losses.

The prevalence of censoring (i.e., the number of observation with incomplete data) in this data set is 68%. Therefore, treating the censored observations as uncensored would lead to severe biases in the results. Available background variables include children's age and IQ (102 $\pm$14.7), measured with the SON-R 2.5-7 at the age of six. Furthermore, information about the mother is available in ethnicity (Dutch = 59.8%, Dutch Antilles = 2.1%, African and Moroccan $<$ 5%, Asian (non-Western) and Turkish $<$ 6%, Surinamese = 7.1%, and other Western = 10,1%) and education (low = 6.7%, middle = 42.2%, high = 51.2%). The last background variable is the household income per month in euros ($<$ 2000 = 20.5%, 2000 $-$ 4000 = 43.8%, $>$ 4000 = 35.7%). Missing values in the background variables are imputed with single predictive mean matching (PMM) using age, gender, weight at birth, and IQ of the child, and the age at delivery, ethnicity, and education of the mother, and household income as predictors, and using the `mice` package in `R`.

The segments obtained with the CMM will be interpreted using the child behavior checklist (CBCL). This survey assesses child emotional and behavioral problems as perceived by the mother. The CBCL has been completed at the same time as the CCT, at age nine. For some children, the CBCL scores at age nine were missing but available at age six. For these children, their scores at age nine are imputed by single PMM using the score at age six and the covariates in the model. The 223 children that have scores neither at age six nor at age nine are excluded from this analysis.

As discussed, the CMM can also be used to analyze other risk tasks, for example the BART. Appendix 2.D contains the results of the CMM applied to a data set from a BART study (Dekkers et al., 2020).

2.4 Methods

The following section is concerned with the methods and techniques applied in this study. First, the structure of the model is discussed, extensively. Although the model is applicable to all sequential risk tasks, we construct the likelihood function through the CCT.

2.4.1 The Censored Mixture Model (CMM)

The CMM has its roots in the field of survival analysis. In the case of the CCT, we need to take care of three challenges: the censoring of the data, the attraction of particular outcomes, such as presented in Figure 2.2, and unobserved heterogeneity across individuals. By constructing a likelihood function, we can explain step-by-step how we incorporate these challenges into the CMM. Additionally, we show explicitly that the censoring is exogenous (also known as independent or non-informative censoring), which is a key assumption in survival analysis. The possible censoring is accommodated by adding the cumulative distribution function, equivalent to the survival function, to the likelihood function. Extra probability mass is assigned to the inflated values in the outcome distribution. Last, the finite mixtures account for the unobserved individual characteristics. Apart from this, we follow a generalized linear model approach, that is, we will assume that there is a linear combination of covariates that provides, after transformation by a link function, the mean of a distribution for every observed number of cards. We argue that the negative binomial distribution is appropriate and provide a link function that is close to linear for ease of interpretation. Below, a step wise explanation is given how these potential problems are solved by the CMM.

The observed variable to be modeled is the number of cards turned over y_{it} by individual i in trial t. Furthermore, we observe whether a trial is censored at card k, $c_{itk} = 1$, or not, $c_{itk} = 0$. However, we are interested in the latent random variable Z_{it}, indicating the number of cards someone intends to turn over. This variable is assumed to follow a known distribution (here we propose to use the negative binomial distribution). Now, the probability of the observed number of cards turned over can be expressed in terms of the latent random variable Z_{it}

$$\Pr(Y_{it} = k \wedge C_{itk} = c_{itk}) = \Pr(Y_{it} = k \wedge C_{itk} = c_{itk} \mid Z_{it} = \ell) \Pr(Z_{it} = \ell),$$

where C_{itk} indicates censoring at trial t for individual i at card k and ℓ is the intended number cards turned over.

More insight on the conditional probability $\Pr(Y_{it} = k \wedge C_{itk} = c_{itk} \mid Z_{it} = \ell)$ can be obtained by considering all possible outcomes of the game, that is, for all combinations of number of observed cards y_{it}, being censored or not ($c_{itk} = 1$ or 0), and the number of cards intended to turn over z_{it}. Table 2.1 provides these probabilities where the notation $p_k = \Pr(C_{itk} = 1 \mid C_{it\ell} = 0 \; \forall \; \ell < k)$ is used. Note that these probabilities are purely based on the game settings. Due to symmetry properties, this table can be summarized by

$$\Pr(Y_{it} = k \wedge C_{itk} = c_{itk} \mid Z_{it} = \ell) \quad = \quad \begin{cases} \Omega_{k\ell,c_{itk}} = \Omega_{kk,c_{itk}} & \forall \ell \geq k \text{ if } c_{itk} = 1, \\ \Omega_{k\ell,c_{itk}} = 0 & \forall \ell < k \text{ if } c_{itk} = 1, \\ \Omega_{k\ell,c_{itk}} = 0 & \forall \ell \neq k \text{ if } c_{itk} = 0 \end{cases}$$

where $\Omega_{k\ell,c_{itk}}$ corresponds to the fixed probabilities given in Table 2.1. Subsequently,

Table 2.1: The probability for each possible combination of the observed number of cards k, the intended number of cards ℓ, and being censored at card k (c_{itk}), that is, $\Pr(Y_{it} = k \wedge C_{itk} = c_{itk} \mid Z_{it} = \ell) = \Omega_{k\ell,c_{itk}}$.

		$z_{it} = \ell$					
$y_{it} = k$	c_{itk}	0	1	2	...	31	32
0	0	1	0	0	...	0	0
0	1	0	0	0	...	0	0
1	0	0	$1-p_1$	0	...	0	0
1	1	0	p_1	p_1	...	p_1	p_1
2	0	0	0	$(1-p_1)(1-p_2)$	...	0	0
2	1	0	0	$(1-p_1)p_2$	...	$(1-p_1)p_2$	$(1-p_1)p_2$
⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮
31	0	0	0	0	...	$\prod_{i=1}^{31}(1-p_i)$	0
31	1	0	0	0	...	$\prod_{i=1}^{30}(1-p_i)p_{31}$	$\prod_{i=1}^{30}(1-p_i)p_{31}$
32	0	0	0	0	...	0	0
32	1	0	0	0	...	0	$\prod_{i=1}^{31}(1-p_i)p_{32}$

the likelihood contribution for person i at trial t can be written as

$$
\begin{aligned}
L_{it} &= \sum_{\ell=0}^{32} \Pr(Y_{it} = k \wedge C_{itk} = c_{itk} \mid Z_{it} = \ell) \Pr(Z_{it} = \ell) \\
&= \begin{cases} \Omega_{k\ell,c_{itk}} \Pr(Z_{it} = \ell) = \Omega_{k\ell,c_{itk}} \Theta_{\ell,c_{itk}} & \text{if } c_{itk} = 0 \\ \displaystyle\sum_{m=k}^{32} \Omega_{km,c_{itk}} \Pr(Z_{it} = m) = \Omega_{k\ell,c_{itk}} \sum_{m=k}^{32} \Pr(Z_{it} = m) = \Omega_{k\ell,c_{itk}} \Theta_{\ell,c_{itk}} & \text{if } c_{itk} = 1. \end{cases}
\end{aligned}
$$

Thus, for $c_{itk} = 1$, $\Theta_{\ell,c_{itk}}$ equals one minus the cumulative distribution function, that is, $\Theta_{\ell,c_{itk}} = 1 - \sum_{\ell=0}^{k-1} \Pr(Z_{it} = \ell)$ and for $c_{itk} = 0$ we have $\Theta_{\ell,c_{itk}} = \Pr(Z_{it} = \ell)$. Thus for $c_{itk} = 1$, $\Theta_{\ell,c_{itk}}$ is equivalent to the survival function. Multiplying over the trials, the likelihood contribution of person i can be written as

$$
L_i = \prod_{t=1}^{T} \Omega_{y_{it}z_{it},c_{itz_{it}}} \Theta_{z_{it},c_{ity_{it}}}.
$$

Although the probability $\Pr(Z_{it} = \ell)$ seems to follow a known distribution, it does not follow a smooth distribution, see Figure 2.2. The left panel shows the number of cards a child intends to turn over, that is, proportionally the empirical equivalence of $\Pr(Y_{it} = k \mid C_{itk} = 0) = \Pr(Z_{it} = k)$. It is easy to see that some outcomes seem extra attractive. We choose to distinguish four of these cases: (a) $k = 0$, (b)

$k \in A$ with $A = \{4, 8, 10, 12, 16, 20, 24\}$, (c) $k = 31$, and (d) otherwise. To control for these four cases, we implement a multiple inflation model, where the observations belonging to each of these cases get extra probability mass through parameter ϕ_m with $m = 1, \ldots, 4$. The probability $\Pr(Z_{it} = \ell)$ is defined by

$$\Pr(Z_{it} = \ell) = \begin{cases} \phi_4 f(0) + \phi_1 & \text{if } \ell = 0, \\ \phi_4 f(\ell) + \frac{1}{|A|}\phi_2 & \text{if } \ell \in A, \\ \phi_4 f(31) + \phi_3 & \text{if } \ell = 31, \\ \phi_4 (1 - F(32)) & \text{if } \ell = 32, \\ \phi_4 f(\ell) & \text{for all other values } \ell, \end{cases}$$

where $f(\ell)$ and $F(\ell)$ are the probability mass function and cumulative distribution function, respectively, of a known distribution. Note that the weights ϕ_m have to be between zero and one, $0 \leq \phi_m \leq 1$, and sum to one, $\sum_{m=1}^{4} \phi_m = 1$.

Since the number of cards someone intends to turn over (z_{it}) is nonnegative and discrete, the distribution has to have these properties as well. We choose the negative binomial distribution, because it allows the variance to differ from the mean, in contrast to the Poisson distribution. The negative binomial distribution can be written as a Poisson–Gamma mixture. Specify the mean of the Poisson distribution as a combination of a deterministic function of the predictors, $\mu_{it} = g(\eta_{it})$, and a random component, $\nu \sim_{\text{i.i.d.}} g(\nu \mid \kappa)$. Let $g(\nu \mid \kappa)$ be the density of the Gamma distribution, then the resulting Poisson–Gamma mixture density can be rewritten as the negative binomial density. The probability mass function of this distribution is specified as follows

$$f(z_{it} \mid \mu_{it}, \delta) = \frac{\Gamma(\delta + z_{it})}{\Gamma(\delta) z_{it}!} \left(\frac{\delta}{\delta + \mu_{it}}\right)^{\delta} \left(\frac{\mu_{it}}{\mu_{it} + \delta}\right)^{z_{it}},$$

where $\delta = 1/\kappa$.

In a generalized linear model (GLM), the mean μ_{it} of the distribution is specified through an inverse link function

$$\mu_{it} = h^{-1}(\eta_{it}).$$

The mean μ_{it} of the negative binomial distribution must be larger than zero. Therefore, the inverse link function $h^{-1}(\eta_{it})$ should map $\eta_{it} \in \mathbb{R}$ to $\mathbb{R}^+$. In GLM, η_{it} is chosen as a linear combination of covariates $\boldsymbol{x}'_{it}$, that is,

$$\eta_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta}.$$

For ease of interpretation of the coefficients $\boldsymbol{\beta}$, we specify, the inverse link function by

$$h^{-1}(\eta_{it}) = \log(\exp(\eta_{it}) + 1)$$

so that whenever $\eta_{it} > 1$, the inverse link function becomes close to linear, yet $h^{-1}(\eta_{it}) > 0$ for any η_{it}, see Figure 2.4 (see, e.g., Ranganath et al., 2016).

The predictor variables are all gathered in the vector $\boldsymbol{x}_{it}$. Some predictor variables are categorical and we choose to represent each of them by their own dummy variable.

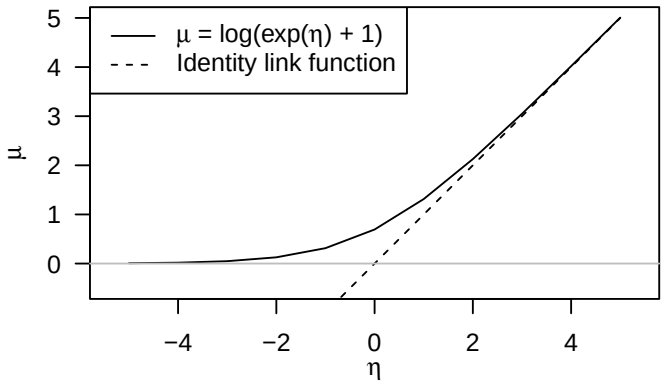


Figure 2.4: Both the proposed inverse link function, $\mu_{it} = h^{-1}(\eta_{it}) = \log(\exp(\eta_{it}) + 1)$, and the identity link function, $\mu_{it} = h^{-1}(\eta_{it}) = \eta_{it}$, on the domain [-5:5].

Without loss of generality, the weights corresponding to the dummy variables for each categorical variable must have sum zero, that is,

$$\boldsymbol{C}\boldsymbol{\beta} = \mathbf{0} \text{ with } \boldsymbol{C} = \begin{bmatrix} (\mathbf{1'1})^{-1/2}\mathbf{1'} & \mathbf{0'} & \mathbf{0'} \\ \mathbf{0'} & (\mathbf{1'1})^{-1/2}\mathbf{1'} & \mathbf{0'} \end{bmatrix},$$

where, for illustration, it is assumed that there are two categorical variables followed by numerical predictors, so that $\boldsymbol{C}$ consists of one row per categorical variable with ones at the positions of the weights and zero elsewhere. Note that the factors $(\mathbf{1'1})^{-1/2}$ are chosen for notational convenience. Again, without loss of generality, it is also assumed that the numerical predictor variables are z-scores (with mean zero and standard deviation one) so that the intercept α can be interpreted as an overall measure of risk taking for someone who has a neutral score on all predictors.

The complete likelihood over all N individuals becomes

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{N} L_i = \prod_{i=1}^{N}\prod_{t=1}^{T} \Omega_{y_{it} z_{it}, c_{ity_{it}}} \Theta_{z_{it}, c_{ity_{it}}} \\ &= \left(\prod_{i=1}^{N}\prod_{t=1}^{T} \Omega_{y_{it} z_{it}, c_{ity_{it}}} \right) \left(\prod_{i=1}^{N}\prod_{t=1}^{T} \Theta_{z_{it}, c_{ity_{it}}} \right) \\ &\propto \prod_{i=1}^{N}\prod_{t=1}^{T} \Theta_{z_{it}, c_{ity_{it}}}, \end{aligned}$$

where $\boldsymbol{\theta}$ is the vector of all unknown parameters. The factor $\prod_{i=1}^{N}\prod_{t=1}^{T} \Omega_{y_{it} z_{it}, c_{ity_{it}}}$ is irrelevant for maximizing the likelihood as it is constant, so that optimizing $\prod_{i=1}^{N}\prod_{t=1}^{T} \Theta_{z_{it}, c_{ity_{it}}}$ over $\boldsymbol{\theta}$ is sufficient. Note that this independence implies that

in the CMM the censoring is exogenous, meaning that the distribution of censoring does not provide information on the distribution of the number of cards turned over. We also assume that any carry-over effects are subsumed in the linear combination η_{it}, so that the conditional independence above still holds.

In a final step of the CMM, we wish to be able to model unobserved heterogeneity across individuals by adding finite mixtures with different intercepts per segment to the model, that is,

$$\eta_{its} = \alpha_s + \boldsymbol{x}_{it}'\boldsymbol{\beta},$$

with α_s the segment specific intercept. The relative size of the segment is estimated by π_s. Then, the likelihood function becomes

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{N}\sum_{s=1}^{S}\pi_s\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{ity_{it}}}\prod_{t=1}^{T}\Theta_{z_{it},c_{ity_{it}},s} \\ &= \left(\prod_{i=1}^{N}\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{ity_{it}}}\right)\left(\prod_{i=1}^{N}\sum_{s=1}^{S}\pi_s\prod_{t=1}^{T}\Theta_{z_{it},c_{ity_{it}},s}\right) \\ &\propto \prod_{i=1}^{N}\sum_{s=1}^{S}\pi_s\prod_{t=1}^{T}\Theta_{z_{it},c_{ity_{it}},s}, \end{aligned} \tag{2.1}$$

where $\boldsymbol{\theta}$ is understood to contain all unknown parameters. Note that $\Theta_{z_{it},c_{ity_{it}},s}$ has obtained an additional subscript s to indicate that this probability is dependent on the parameter α_s. Thus, the CMM needs to maximize $L(\boldsymbol{\theta})$ over $\boldsymbol{\theta}' = [\boldsymbol{\alpha}', \boldsymbol{\beta}', \delta, \boldsymbol{\phi}', \boldsymbol{\pi}']$ subject to $\boldsymbol{C\beta} = \mathbf{0}$, $\phi_m \geq 0$, $\mathbf{1}'\boldsymbol{\phi} = 1$, $\pi_s \geq 0$, and $\mathbf{1}'\boldsymbol{\pi} = 1$. More details about the estimation procedure can be found in Appendix 2.A.

2.5 Results

Before the Censored Mixture Model (CMM) can be applied to the data discussed in Section 2.3, several parameters need to be set. First, the maximization of the log likelihood function is performed through the `optimx` function in `optimx` package in `R`. All default settings are used except for the relative convergence tolerance, *reltol*, which is set more strictly such that the maximization has converged as soon as $\log L(\boldsymbol{\theta}^{(t)}) - \log L(\boldsymbol{\theta}^{(t-1)})$ is less than $10^{-10}(|\log L(\boldsymbol{\theta}^{(t)})| + 10^{-10})$ where t is the iteration counter. After convergence, one step of the Newton-Raphson method is performed using a numerically approximated Hessian with the aim of ensuring that the gradient is close to zero. To speed up the convergence, the start values of $\boldsymbol{\alpha}$ and $\boldsymbol{\phi}$ are based on educated guesses. For $\boldsymbol{\alpha}$, the start values are uniformly distributed over the possible outcomes, $\{0, 32\}$. The initial values of $\boldsymbol{\phi}$ are based on the observed proportion of excesses in Figure 2.2, that is, the difference of the observed proportion of the inflated outcome minus the interpolated value of the previous and next outcomes without inflation. To further improve the convergence speed, we trained our model on a small subsample, $N = 100$, and implemented these parameter estimates as start values of $\boldsymbol{\theta}$ in the model using the original sample.

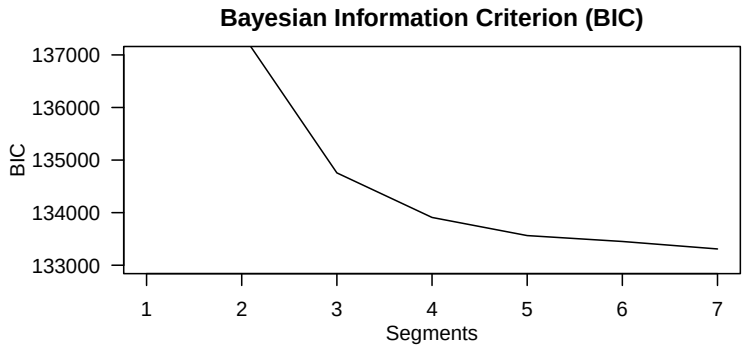


Figure 2.5: The Bayesian information criterion (BIC) of CMMs with $S = 2$ to 7 segments.

2.5.1 Selection of the Number of Segments

As the number of segments is unknown a priori, the model is computed for several numbers of segments S. We use several criteria to decide on a useful number of segments: the Bayesian information criterium (BIC), a minimum segment size, and the distinctiveness of the segment specific intercept. The BIC for various choices of S is shown in Figure 2.5. Since the number of observations is so large in this study, adding a segment hardly affects the BIC. Therefore, searching for the number of segments that would lead to a minimum BIC would require an unrealistically high number of segments. Therefore, we additionally check the size of the segments π_s and the segment specific intercepts α_s given in Table 2.2. We opt for segments that have $\pi_s > 5\%$ of the observations, that is, 170 children. Furthermore, we impose the segment specific intercepts α_s to be sufficiently different to avoid segments where the level of risk seeking as symbolized by their respective α_s is hardly different. Based on these three criteria, we choose to continue interpreting the model with $S = 4$ segments.

From Table 2.2, it is clear that Segment 1 is overall the smallest ($\pi_1 = 0.097$) and is characterized by children that are on average most risk averse as α_1 is the smallest of all segments. In contrast, the last segment contains children that are most risk seeking as their intercept $\alpha_4 = 37.52$ is even larger than the total number of cards that could be turned over in the game.

2.5.2 Segments Specific Results

For each individual, we can compute the a posteriori probability of belonging to a segment. Ideally, these probabilities are close to one for one of the segments and close to zero for the others thereby clearly assigning an individual to a segment. To see how distinctive the segments are, we consider the highest a posteriori probability for each individual and plot that in a histogram. Figure 2.6 shows this distribution and it is clear that indeed most children are assigned to a segment with a large probability.

Table 2.2: Segment probabilities π_s and segment specific intercepts α_s with the standard errors between brackets for CMMs with $S = 2$ to 7 segments. Note that the seven-segment solution is near the boundary and that α_6 in this solution is poorly estimated.

		Segment s						
S		1	2	3	4	5	6	7
2								
	π_s	0.394	0.606					
		(0.010)	(0.010)					
	α_s	9.90	26.35					
		(0.164)	(0.313)					
3								
	π_s	0.150	0.432	0.419				
		(0.007)	(0.010)	(0.011)				
	α_s	6.77	13.93	30.93				
		(0.148)	(0.184)	(0.408)				
4								
	π_s	0.097	0.275	0.357	0.271			
		(0.006)	(0.011)	(0.012)	(0.011)			
	α_s	5.85	11.04	18.68	37.52			
		(0.152)	(0.188)	(0.295)	(0.772)			
5								
	π_s	0.023	0.119	0.284	0.331	0.243		
		(0.003)	(0.007)	(0.011)	(0.012)	(0.012)		
	α_s	3.11	6.89	11.74	19.44	38.40		
		(0.167)	(0.148)	(0.187)	(0.322)	(0.904)		
6								
	π_s	0.002	0.103	0.206	0.256	0.249	0.164	
		(0.003)	(0.008)	(0.021)	(0.015)	(0.017)	(0.015)	
	α_s	2.67	6.55	10.64	15.58	23.97	46.82	
		(0.258)	(0.168)	(0.315)	(0.587)	(0.852)	(2.685)	
7								
	π_s	0.007	0.052	0.130	0.294	0.303	0.000	0.214
		(0.002)	(0.005)	(0.009)	(0.012)	(0.012)	(0.000)	(0.013)
	α_s	-0.66	5.03	8.36	12.99	21.07	22.70	40.83
		(0.225)	(0.163)	(0.188)	(0.225)	(0.438)	(148.3)	(1.328)

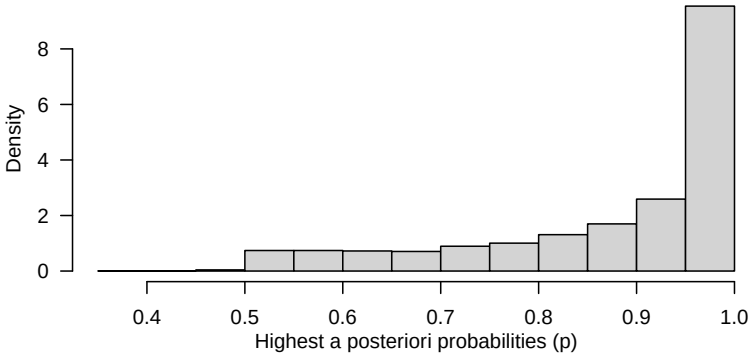


Figure 2.6: A histogram of the highest a posteriori segment probability of each individual.

Therefore, virtually each child can be assigned with high probability to one of the segments.

It is interesting to investigate how the segments differ on characteristics that have not been part of the model. In particular, how do the segments differ with regard to the occurrence of behavioral problems as measured by the CBCL? The resulting z-scores (with mean zero and standard deviation one) per segment weighted by the a posteriori probabilities per segment are presented in Table 2.3. Appendix 2.B discusses how to test for differences of weighted means. The stars in Table 2.3 denote whether one of the segment averages is significantly different from the overall average for this particular symptom.

The level of behavioral problems in all subscales except that of somatic complaints differs between the groups of children as defined by our segments. The risk averse children in Segment 1 and the risk seekers in Segment 4 on average have more behavioral problems than the children in Segment 2 and 3. Table 2.3 also suggests that children in Segment 3 who intend to turn over on average 18 cards score on average the lowest on all CBCL subscales. Furthermore, we can see from this table that a risk averse strategy is most profitable, as the average score is highest in Segment 1 (most risk averse segment) and lowest in Segment 4 (most risk seeking segment).

One of the contributions of the CMM model is that the intended number of cards to be turned over is estimated by the segment specific intercept α_s. Due to the censoring, children who intend to turn over a high number of cards will often not be able to do so. Therefore, the observed number of cards turned over under estimates the intended number of cards to be turned over. We can easily compare them using the forelast row in Table 2.3 with the α_s from Table 2.2. For example, the average number of observed cards turned over by children in Segment 1 is 5.0 whereas average number of card intended is 5.9. For Segments 2, 3, and 4, these values are 7.7, 10.0, and 11.7 observed and 11.0, 18.7, and 37.5 intended. Indeed, a large under estimation of risk seeking is obtained when only considering the observed number of cards turned over.

Table 2.3: Weighted z-scores per segment of CBCL subscales scores and other CCT characteristics. A Wald test is performed to check for a significant difference between the segments. One star denotes $0.05 \leq p < 0.10$, two $0.01 \leq p < 0.05$, and three $p < 0.01$. The 223 children without a CBCL score measured at either six or nine years old are excluded.

	Segment s				
	1	2	3	4	Total
π_s	0.10	0.28	0.36	0.27	
CBCL subscales					
Internalizing *	0.06	0.00	-0.05	0.05	0.00
Externalizing **	0.08	-0.03	-0.04	0.06	0.00
CBCL symptom subscales					
Anxiety *	0.07	0.01	-0.06	0.04	0.00
Social withdrawal *	0.02	0.02	-0.06	0.05	0.00
Somatic complaints	0.03	-0.02	-0.01	0.02	0.00
Social problems ***	0.09	-0.04	-0.05	0.08	0.00
Thought problems **	0.10	-0.02	-0.05	0.05	0.00
Attention problems ***	-0.03	-0.04	-0.06	0.13	0.00
Delinquent behavior **	0.00	-0.03	-0.03	0.08	0.00
Aggressive behavior *	0.10	-0.02	-0.04	0.04	0.00
Average score ***	-87.0	-123.7	-170.7	-230.1	-165.7
# cards turned over ***	5.0	7.7	10.0	11.7	9.3
# censored trials ***	5.8	8.5	11.4	14.3	10.8

2.5.3 Regression Coefficients

The regression coefficients $\boldsymbol{\beta}$ are presented in Table 2.4. The two numerical variables (age and IQ) are standardized to z-scores prior to the analysis. Whether or not a loss card was drawn in the last two games is recorded by the following variables: previous loss yes, previous loss no, second previous loss yes, and second previous loss no. As our link function in Figure 2.4 is close to the identity function for values larger than 1, the coefficients can be interpreted on the scale of the number of cards turned over. For categorical variables, we chose mean weights of the categories belonging to a single variable to be zero so that the intercept can be interpreted as the average score in the segment for a neutral child. As a consequence, the difference in weights between two categories is the corresponding effect, for example, girls on average turn over $0.286 + \mid -0.286 \mid = 0.572$ cards more than boys.

Furthermore, age and IQ have a negative association with the number of cards turned over. Also, a higher household income is related to higher levels of risk taking. Children with a mother with a Dutch or Asian ethnicity turn over fewer cards than the base average.

Due to the different game settings, we are able to investigate the effect of the loss

probability and the sensitivity to reward and punishment. According to the model, the number of loss cards has the strongest effect on the number of cards turned over. In a game with three loss cards on average 1.7 cards less are turned over, than in a game with one loss card. The game setting loss amount also shows the expected direction of effect. In a trial with a high loss amount, the expected number of cards turned over is lower than in a trial with a low loss amount. Unexpectedly, the predicted number of cards turned over is lower in a trial with a high gain amount than it is in a trial with a low gain amount.

Moreover, the results in the previous round have a strong association with observed behavior in the current round. If a loss card was encountered in the previous round, on average 1.6 cards less are turned over. The experience of a loss card two trials earlier also relates to the intention to turn over one card less in the current trial. Note that these variables capture the immediate impact of a negative experience (e.g., turning over a loss card), not the learning effect. We argue that there is no learning effect in the current data set, since the average number of cards turned over per trial varies between 10.5 cards in trial 1 and 8.9 cards in trial 15.

We included interaction terms between the game settings and sex. According to the estimates, the combination of boy and a loss amount of 250 accounts for an additional 0.063 ($= -0.286 + 0.195 + 0.154$) cards to be turned over. In a trial with loss amount 750, a boy is expected to turn over 0.635 ($=\mid -0.286 - 0.195 - 0.154 \mid$) cards less than the base average (i.e., the segment specific intercepts). Hence, the effect the loss amount has on the number of cards turned over by boys is 0.698 ($= 0.063+ \mid -0.635 \mid$). This effect is smaller for girls, namely 0.572 ($= (0.286 + 0.195 - 0.154) + (0.286 - 0.195 + 0.154)$). Therefore, boys seem to be more sensitive to punishment in the CCT than girls are. Moreover, boys are also more influenced by the number of loss cards (2.038 vs. 1.360), whereas girls seem to be more sensitive to reward than boys are (0.346 vs. 1.026).

Table 2.4: Regression coefficients with their standard errors. Within a categorical variable the sum of coefficients sum to zero and the continues variables age and IQ are standardized.

Background variables	β - coefficients (st error)		Game settings	β - coefficients (st error)	
Age	-0.012	(0.071)	Gain amount (10)	0.343	(0.039)
Boy	-0.286	(0.079)	Gain amount (30)	-0.343	(0.039)
Girl	0.286	(0.079)	Loss amount (250)	0.195	(0.038)
IQ	-0.539	(0.095)	Loss amount (750)	-0.195	(0.038)
Ethnicity mother			Loss cards (1)	0.850	(0.039)
Dutch	-1.170	(0.157)	Loss cards (3)	-0.850	(0.039)
Asian	-0.875	(0.258)	Previous loss yes	-0.823	(0.040)
African	0.570	(0.393)	Previous loss no	0.823	(0.040)
Moroccan	0.477	(0.338)	Second previous loss yes	-0.502	(0.040)
Dutch Antilles	-0.139	(0.379)	Second previous loss no	0.502	(0.040)
Surinamese	0.288	(0.378)	Interaction terms		
Turkish	0.527	(0.342)	Gain amount (10) : Boy	-0.170	(0.039)
Other Western	0.322	(0.265)	Gain amount (30) : Boy	0.170	(0.039)
Education mother			Gain amount (10) : Girl	0.170	(0.039)
No or primary education	0.571	(0.219)	Gain amount (30) : Girl	-0.170	(0.039)
Secondary education	-0.231	(0.143)	Loss amount (250) : Boy	0.154	(0.038)
Higher education	-0.340	(0.137)	Loss amount (750) : Boy	-0.154	(0.038)
Household income per month in euro's			Loss amount (250) : Girl	-0.154	(0.038)
< 2000	-0.134	(0.163)	Loss amount (750) : Girl	0.154	(0.038)
$2000 - 4000$	-0.231	(0.114)	Loss cards (1) : Boy	0.169	(0.039)
> 4000	0.365	(0.123)	Loss cards (3) : Boy	-0.169	(0.039)
			Loss cards (1) : Girl	-0.169	(0.039)
			Loss cards (3) : Girl	0.169	(0.039)

2.5.4 Model Performance

Our model gives of each child on each trial a probability distribution for the number of cards turned over. To obtain a sense how well the model fits the observed uncensored number of cards turned over, we compute a point estimate as the expected value of that distribution. Then, the model performance can be judged in terms of the difference between observed and expected number of cards turned over. Appendix 2.C provides more details on how these expectations are computed. Predictions can be generated with our CMM. The in-sample root mean square error (RMSE) is equal to 8.5 and the mean absolute deviation (MAD) of the residuals is equal to 5.4. On a scale of 0-32 cards that can be turned over, these average deviations seem reasonable. To test the external validity of the model, the data set is randomly partitioned in a training set ($N = 3404$) and a prior to analysis unseen test set ($N = 1134$). The out-of-sample RMSE is equal to 8.4 and the MAD is equal to 5.3, showing little difference between in-sample and out-of-sample accuracy.

Another way to evaluate the model performance is by comparing the distributions of the empirical and predicted number of cards turned over for the training and the test data, similar to Figure 2.2. We break down the comparison into a censored and uncensored case. For a fair comparison between the observed and predicted number of cards turned over, one has to multiply the distribution of the predicted outcome with the probability of being (un)censored. These probabilities can be derived from Table 2.1, where p_k is equal to

$$p_k = \frac{1}{32-k-1} \qquad \text{if \#loss cards } = 1$$
$$p_k = \frac{3}{32-k-1} \qquad \text{if \#loss cards } = 3.$$

The empirical probability of the number of cards intended to turn over in the training set is shown in the left panel of Figure 2.7. The comparison with the right panel with the CMM predicted probabilities shows that these predictions are quite accurate. The left panel of Figure 2.8 shows the empirical probability per card of being censored in the training data and the right panel shows these values as predicted. Again, the distribution of the predicted values is similar to those observed. To guard against overfitting, we provide the same plots for the test set of 1049 children in Figures 2.9 and 2.10. The same interpretation holds as for the training data: there are some minor deviations from the observed distribution, but overall the test set prediction of these distributions is quite accurate.

A widely used test to compare two distributions is the Chi-square goodness of fit test, which tests whether the observed sample is drawn from the predicted distribution. The in-sample Chi-square statistic is equal to $X^2 = 793.8$ ($p < 0.001$, df $= 31$). Note that only the uncensored observations are used to compute this statistic. It is well known that the Chi-square goodness of fit test is very sensitive to the number of observations (Cochran, 1952); therefore, we also look at the correlation between the observed and predicted probabilities. Figure 2.11 displays a strong correlation of 0.97 between the predicted and observed probabilities. The out-of-sample correlation is equal to 0.97. In addition, as a (dis)similarity measure we added the Hellinger

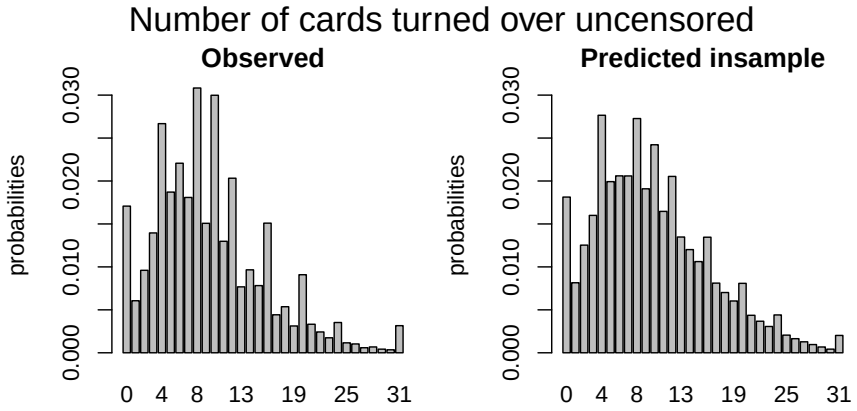


Figure 2.7: Distribution of the empirical (left panel) and predicted by the CMM (right panel) number of cards turned over for the uncensored observations in the training data.

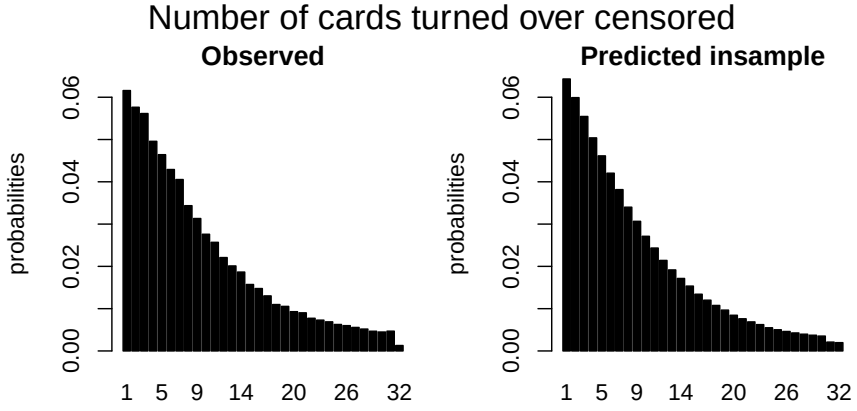


Figure 2.8: Distribution of the empirical (left panel) and predicted by the CMM (right panel) number of cards turned over corrected for the probability of being censored per card in the training data.

distance between the predicted and observed distribution. The Hellinger distance is related to the Euclidean distance, so the closer the value to zero the similar the two distributions are. Both the in-sample and out-of-sample Hellinger distance is rounded equal to 0.08.

One of the model assumptions is conditional independence implying that any carry over effects are subsumed in the linear combination. To empirically justify this assumption, we additionally estimated the model with dummy variables up to ten

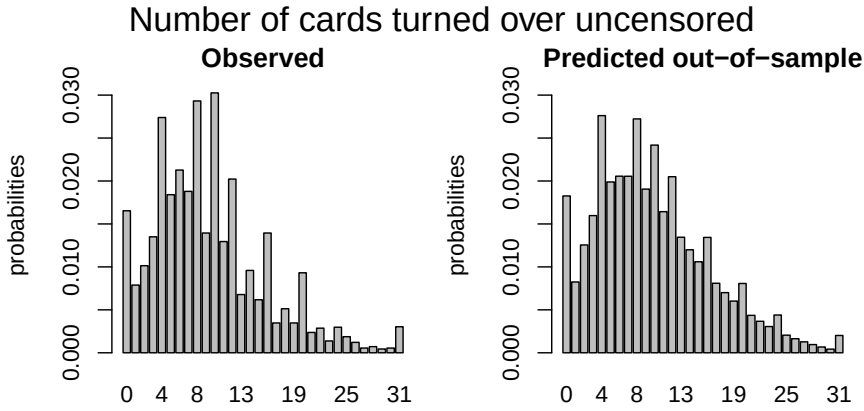


Figure 2.9: Distribution of the empirical (left panel) and predicted number of cards turned over by the CMM (right panel) for the uncensored observations in the test data.

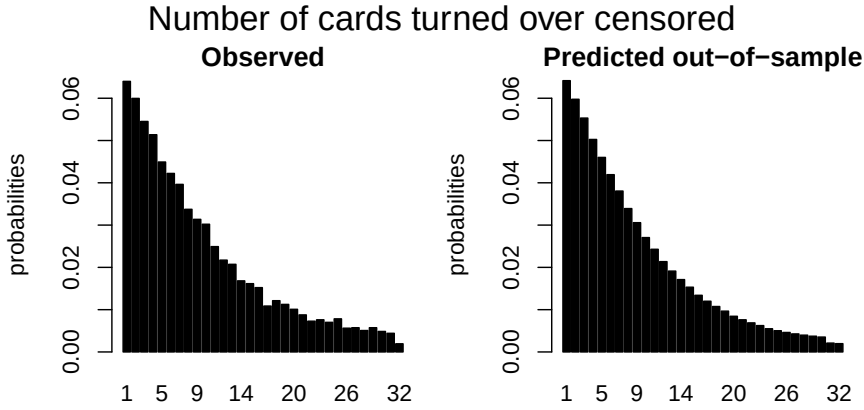


Figure 2.10: Distribution of the empirical (left panel) and predicted by the CMM (right panel) number of cards turned over corrected for the probability of being censored per card in the test data.

lags whether or not a loss card was encountered. The strongest effect occurred immediately after the loss card was encountered. We found a monotonically decreasing effect of the lag of the loss trial, which can be seen as a support for the conditional independence assumption.

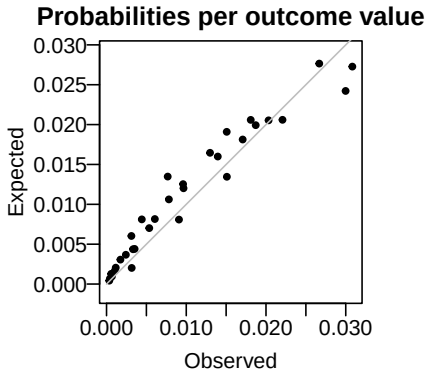


Figure 2.11: Scatterplot of the observed and expected probabilities per outcome value {0, 31}.

2.5.5 Segment Specific Effects for Covariates

In addition to segment specific intercepts, the CMM also allows for segment specific effects for covariates. Here we discuss the results of a CMM with four segments and segment specific effects for the game settings gain amount, loss amount, and number of loss cards. The linear combination in this model becomes

$$\eta_{its} = \alpha_s + \tilde{\boldsymbol{x}}'_{it}\tilde{\boldsymbol{\beta}}_s + \boldsymbol{x}^{*\prime}_{it}\boldsymbol{\beta}^*,$$

where $\tilde{\boldsymbol{x}}'_{it}$ contains the game settings gain amount, loss amount, and number of loss cards, $\tilde{\boldsymbol{\beta}}_s$ are the segment specific effects for these game settings, and $\boldsymbol{x}^{*\prime}_{it}$ contains the covariates excluding the game setting variables. Note that we did not include interaction terms between the game settings and sex in this analysis, because these effects are captured by the segment specific effects.

The segment probabilities π_s, segment specific intercepts α_s, and segment specific effects $\tilde{\boldsymbol{\beta}}_s$ are presented in Table 2.5. The segment sizes and segment specific intercepts are comparable to the ones belonging to the CMM with four segments and only segments specific intercepts, see Table 2.2. It is clear that children in the fourth segment use to the information provided the most. In addition, the increase in information use is evident for all the game settings. So, these segments do not make a clear distinction between children who are, for example, more sensitive to reward than to punishment. Furthermore, children in the first segment only take into account the number of loss cards.

Table 2.5: Results of the four segment CMM with both segment specific intercepts α_s and segment specific effects of the game setting parameters $\tilde{\boldsymbol{\beta}}_s$. The standard errors are given in parentheses.

	Segment s							
	1		2		3		4	
π_s	0.089	(0.027)	0.268	(0.072)	0.362	(0.068)	0.280	(0.041)
α_s	5.239	(0.705)	10.409	(0.435)	17.914	(0.435)	35.127	(0.435)
Gain amount (10)	-0.043	(0.060)	0.327	(0.063)	1.347	(0.119)	1.351	(0.435)
Gain amount (30)	0.043	(0.060)	-0.327	(0.063)	-1.347	(0.119)	-1.351	(0.435)
Loss amount (250)	0.062	(0.059)	0.330	(0.063)	0.224	(0.110)	1.030	(0.435)
Loss amount (750)	-0.062	(0.059)	-0.330	(0.063)	-0.224	(0.110)	-1.030	(0.435)
Loss cards (1)	0.437	(0.062)	1.028	(0.065)	1.480	(0.117)	3.384	(0.464)
Loss cards (3)	-0.437	(0.062)	-1.028	(0.065)	-1.480	(0.117)	-3.384	(0.464)

This model can be compared to the models including only segment specific intercepts by means of the BIC. The BIC for this model including segment specific effects for the game settings in addition to segment specific intercepts is equal to 133,771. The BICs of the other models are presented in Figure 2.5. In particular, the BIC value of 133,909 for the four segments intercepts only model is slightly worse.

The other results from this analysis (i.e., regression coefficients, segment specific results, and model fit) are very similar to the ones presented above and, therefore, are included in Appendix 2.E.

The four-segment model with segment specific effects for the game settings discussed in this section is just one of many that can be considered. In practice, various models can be compared that differ in the number of segments and the covariates that are segment specific. These models can be compared based on, for example, their BIC value.

2.5.6 Parameter Recovery

We did a small simulation study to investigate the parameter recovery. This recovery is limited to the game setting parameters (i.e., the parameters belonging to the gain amount, loss amount and number of loss cards), as other predictors can be seen as covariates correcting the individual's response for background variables. Data for $N = 500$ children who each played $T = 8$ trials are generated for two sets of game setting parameters β_1, β_2, and β_3. The data are sampled from $S = 2$ mixtures with segment specific intercepts α_1 and α_2. The probabilities ϕ_1, ϕ_2, ϕ_3, and ϕ_4 for the attractiveness for certain outcomes are set, such that they resemble the attractiveness in Figure 2.2. Censoring is based on the probability of encountering a loss card as determined by the game setting of either one or three loss cards. In the simulation study, we use a slightly different parametrization from Appendix 2.A to ensure uniqueness of the parameters. Particularly, we use $\alpha_1 = \tilde{\alpha}_1^2$ and $\alpha_2 = \tilde{\alpha}_1^2 + \tilde{\alpha}_2^2$ and optimize over $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$. Without loss of generality, we enforce $\tilde{\alpha}_1 \geq 0$ and $\tilde{\alpha}_2 \geq 0$. These settings lead to a total of ten unique parameters, that is, β_1, β_2, β_3, $\tilde{\alpha}_1$, $\tilde{\alpha}_2$, δ, τ_1, τ_2, τ_3, and σ.

The definitions of the probabilities in the maximum likelihood function in the CMM served as the basis of the data generating process. For each of the parameter sets 500 replication data sets were drawn. Table 2.6 presents the two sets of true parameter values, the mean, median, and standard deviation of the estimated parameters, the average of the estimated standard errors (MeanSE), the RMSE, MAD, and coverage percentage of all model parameters. The RMSE, MAD, and coverage percentage are calculated with respect to the true parameter values and the coverage percentage is based on the 95% confidence intervals. For computational speed, we used the true parameter values as start values in the optimization process. The present recovery of the model parameters is good. Consistent with expectations when using 95% confidence intervals, almost all parameters in both sets have a coverage percentage around 95%. Furthermore, this parameter recovery allows us to investigate the performance of the BIC statistic in detecting the true number of segments. We compared the BIC statistic of the one segment models with the ones from the two segment models. In all replications in both sets, the two segment model is favored over the one segment model, thereby confirming the usefulness of the BIC statistic in identifying the number of segments.

2.6 Discussion

The Censored Mixture Model (CMM) is developed to solve three potential problems emerging with modeling risk taking and is applied to the Generation R data set, an exceptionally large data set with 3404 children that each completed 16 rounds of the Columbia Card Task (CCT). First, to accommodate the potential censoring that often occurs in sequential risk tasks, a cumulative distribution function is added to the likelihood function to compute the probability of turning over more cards than observed. In the Generation R data set, the prevalence of censoring is 68%. Ignoring the censoring would seriously underestimate the intended level of risk taking as more than two third of the data would not be used. Second, the inflated values in the outcome distribution are handled by assigning additional probability mass to these outcomes in the likelihood function. Figures 2.7, 2.8, 2.9, and 2.10 clearly show peaks at certain outcome values in both the observed and predicted graphs. Without the additional probability mass for the inflated values, the distributions in the predicted graphs would have been smoother and, hence, less similar to the observed graphs. Finally, four segments with a segment specific intercept are added to the model to account for unobserved heterogeneity across individuals. The distribution of posterior probabilities in Figure 2.6 clearly points out that the four segments are distinctive as large probabilities (say above 80%) are most prevalent. In case the individuals in the sample all have the same tendency for risk taking, this graph would be centered around 0.25, indicating that the individuals are assigned to all segments with equal probability.

Similar challenges as with the CCT occur when analyzing risk taking through other sequential risk tasks, like the BART and Angling Risk Task. Hence, our CMM can also be used to analyze these risk tasks. Only minor adjustments to the model are necessary. For example, additional probability mass is assigned to inflated outcome

Table 2.6: Recovery results of the CMM for two sets of true parameter values with $N = 500$ children each playing $T = 8$ trials. Reported are the true parameter value, mean, median, and standard deviation (SD) of the estimated parameters, the average of the estimated standard errors (MeanSE), root mean square error (RMSE), mean absolute deviation (MAD), and coverage percentage.

Parameter	True value	Mean	Median	SD	MeanSE	RMSE	MAD	Coverage
Set 1								
β_1	2.00	1.98	1.96	0.34	0.35	0.34	0.27	0.95
β_2	-3.00	-3.02	-3.01	0.40	0.37	0.40	0.32	0.94
β_3	-6.00	-6.02	-5.98	0.44	0.42	0.44	0.35	0.95
$\tilde{\alpha}_1$	3.39	3.39	3.39	0.09	0.08	0.09	0.07	0.94
$\tilde{\alpha}_2$	2.92	2.92	2.91	0.10	0.10	0.10	0.08	0.96
δ	3.00	3.03	3.02	0.21	0.22	0.22	0.17	0.96
τ_1	-3.50	-3.51	-3.50	0.14	0.14	0.14	0.11	0.94
τ_2	-1.90	-1.90	-1.90	0.15	0.15	0.15	0.12	0.95
τ_3	-2.10	-2.12	-2.10	0.21	0.21	0.21	0.17	0.95
σ	-0.40	-0.40	-0.41	0.19	0.18	0.19	0.15	0.94
Set 2								
β_1	6.00	6.00	6.00	0.62	0.61	0.62	0.50	0.95
β_2	-4.00	-4.00	-3.99	0.56	0.54	0.55	0.45	0.95
β_3	-7.50	-7.55	-7.52	0.59	0.60	0.59	0.48	0.95
$\tilde{\alpha}_1$	3.87	3.87	3.88	0.12	0.12	0.12	0.09	0.94
$\tilde{\alpha}_2$	2.83	2.86	2.85	0.16	0.16	0.16	0.12	0.96
δ	3.00	3.03	3.02	0.23	0.22	0.23	0.18	0.96
τ_1	-3.50	-3.50	-3.50	0.13	0.13	0.12	0.10	0.96
τ_2	-1.90	-1.90	-1.89	0.15	0.14	0.15	0.12	0.93
τ_3	-2.10	-2.14	-2.10	0.30	0.29	0.30	0.23	0.96
σ	-0.40	-0.39	-0.41	0.39	0.35	0.39	0.30	0.91

values: for the CCT these values are $\{0, 4, 8, 10, 12, 16, 20, 24, 31\}$, which corresponds to creating geometric patterns, such as complete rows or columns. However, for other sequential risk tasks these values are likely to be different, depending upon the shape of the distribution of the outcome values and, hence, need to be adapted in the model.

We want to briefly discuss several elements of our CMM applied to our specific data set. The selection of the number of segments in the CMM was based on a three-way procedure: (a) the Bayesian information criterion (BIC) values of the different models are compared, (b) the segment specific intercepts had to be distinctive among the segments, and, (c) the smallest segment had to contain at least five percent of the sample. Although we are confident that a model with four segments is optimal in our case, a different strategy could have led to a different number of segments and, hence, slightly different results.

The CMM with four segments showed some interesting results in terms of segment characteristics. Both the most risk averse and risk seeking segments, respectively, Segments 1 and 4, have the highest level of behavioral problems measured by the

Table 2.7: Optimal number of cards to turn over when maximizing the expected value.

1 loss card			
		Loss amount	
		250	750
Gain	10	7	0
amount	30	23	6

3 loss cards			
		Loss amount	
		250	750
Gain	10	0	0
amount	30	4	0

child behavioral checklist (CBCL).

Furthermore, from Table 2.4, where the regression coefficients are shown it is clear that the number of loss cards has the highest effect on the number of cards turned over, compared to the gain amount and loss amount. This result is in accordance with many other studies using the CCT (Holper and Murphy, 2014; Kluwe-Schiavon et al., 2015; Penolazzi et al., 2012). Looking at the risk neutral strategy based on the expected values (Table 2.7), it is clear that turning over zero cards is often most profitable. It would be interesting to see whether different game settings lead to the same results. Note that we used the same game settings as (Figner and Weber, 2011). Additionally, children with a high IQ turn over fewer cards than the average and children from a family with a low household income turn over more cards than the average. Lastly, children with a Dutch or Asian background are more risk averse, compared to children with another ethnic background.

We found a good model fit represented by a correlation of 0.97 between the observed and predicted probabilities of the number of cards at an aggregate level and a root mean square error (RMSE) of 8.5 and mean absolute deviation (MAD) of 5.5 derived from point estimates for the number of cards turned over. The out-of-sample correlation is equal to 0.97, the RMSE is equal to 8.4 and the MAD is equal to 5.3, showing little difference between in-sample and out-of-sample accuracy. Note that the current predictions are for unseen individuals. However, using a part of the trials for computing a posteriori segment probabilities would allow to make predictions for future trials taking better stock of the unobserved heterogeneity in the data.

Next to segment specific intercepts, the CMM also allows for segment specific effects for the game settings. We found that individuals in the most risk seeking segment, and thus with the highest segments specific intercept, also have the highest effects of the segment specific game setting weights. Furthermore, in all segments, the number of loss cards is taken into account. The difference between the segments lies in the use of the gain amount and loss amount.

A parameter recovery study showed a good recovery of the model parameters. Almost all parameters have a coverage percentage around 95%. Furthermore, this simulation study shows that the BIC is a useful statistic to identify the number of segments.

Recently, new light was shed on the debate concerning parameter reliability, that

is, whether behavioral tasks provide reliable estimates and whether they have good construct validity. Pedroni et al. (2017) argue that people's risk preferences are inconsistent across behavioral tasks. Holzmeister and Stefan (2021), meanwhile, provide evidence that participants are well aware of the variation in risk level associated with their choices and question the statement of inconsistent risk preferences. Note that the proposed CMM is not aimed at obtaining risk scores for individuals. Instead, our model provides useful conclusions on risk taking at the level of the entire sample.

It would also be interesting for further research to collect more information on the underlying decision process. For example, the time between actions (e.g., turning over cards) can be used to investigate a potential fatigue effect. In addition, more time between actions at the end of a trial could indicate that a participant doubts between continuing and stopping. Moreover, the pattern and sequence of the cards turned over could further justify our assumption that participants of the CCT create geometric patterns for turning over cards. Furthermore, the segment specific effects can be extended to other variables. For instance, the previous loss variables can be made segment specific to allow more freedom in capturing the carry-over effects.

A limitation of the current approach is that the negative binomial distribution has an infinite upper bound. This property implies that there is probability mass after 32, meaning that the model allows for turning over cards in a non-feasible region. However, all the point estimates (for computations, see Appendix 2.C) are within the range $\{0, 32\}$. Therefore we argue that this is not a major issue. As an alternative to the negative binomial distribution, future studies can use a truncated version, that is, consider the negative binomial distribution conditional on the number of cards turned over being smaller or equal to 32 for the CCT. Taken together, we believe that the Censored Mixture Model proposed in this paper is an important tool in the analysis of risk taking.

Appendix

2.A Maximization of the Likelihood Function

To be able to maximize the likelihood over $\boldsymbol{\theta}$, it is useful to have no constraints and to ensure that $\boldsymbol{\theta}$ is unique. To do so, several reparametrizations are needed and we will represent that by the vector function $\boldsymbol{h}()$. First, to ensure the restrictions $0 \leq \phi_m \leq 1$ and $\sum_{m=1}^{4} \phi_m = 1$, we define

$$\boldsymbol{\phi} = \boldsymbol{h}_{\boldsymbol{\phi}}(\boldsymbol{\tau}) = \begin{cases} \exp(\tau_m)/(1 + \sum_{j=1}^{3} \exp(\tau_j)) & \text{for } m \in \{1, 2, 3\}, \\ 1 - \sum_{m=1}^{3} \exp(\tau_m)/(1 + \sum_{j=1}^{3} \exp(\tau_j)) & \text{for } m = 4. \end{cases}$$

This reparametrization allows the constrained optimization over $\boldsymbol{\phi}$ to be replaced by the unconstrained optimization over $\boldsymbol{\tau}$. Secondly, a similar reparametrization is used for avoiding the sum one and nonnegativity constraints on the S segment probabilities in $\boldsymbol{\pi}$ by $S-1$ values through $\boldsymbol{\pi} = \boldsymbol{h}_{\boldsymbol{\pi}}(\boldsymbol{\sigma})$.

Thirdly, there are sum zero constraints $\boldsymbol{C}_j\boldsymbol{\beta}_j = \boldsymbol{0}$ on the weights corresponding to the dummy coding of the categories belonging to each categorical variable j. Instead of using these constraints, one of the categories per categorical variable is assigned as a reference category and consequently that particular weight in $\boldsymbol{\beta}_j$ are set to zero, effectively excluding these weights from the optimization. This implies that the effect of the reference categories is included in the intercepts $\boldsymbol{\alpha}_u$. For notational convenience, it is useful to gather the intercepts in $\boldsymbol{\alpha}$ and weights $\boldsymbol{\beta}$ into a single vector $\boldsymbol{\gamma}$, that is, $\boldsymbol{\gamma}_u = [\boldsymbol{\alpha}_u', \boldsymbol{\beta}_u']'$ where $\boldsymbol{\beta}_u$ is the vector of weights with the values corresponding to reference categories fixed to zero. The transformation needed from the unique vector of parameters $\boldsymbol{\gamma}_u = [\boldsymbol{\alpha}_u', \boldsymbol{\beta}_u']'$ to the non-unique vector $\boldsymbol{\gamma} = [\boldsymbol{\alpha}', \boldsymbol{\beta}']'$ is illustrated by the following example with $S = 3$, two categorical variables, and two numerical variables. Then,

$$\boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \boldsymbol{h}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_u) = \begin{bmatrix} \boldsymbol{I} & K_1^{-1}\boldsymbol{1}\boldsymbol{1}' & K_2^{-1}\boldsymbol{1}\boldsymbol{1}' & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} - K_1^{-1}\boldsymbol{1}\boldsymbol{1}' & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} - K_2^{-1}\boldsymbol{1}\boldsymbol{1}' & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & 1 & 0 \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & 0 & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_u \\ \boldsymbol{\beta}_{u1} \\ \boldsymbol{\beta}_{u2} \\ \beta_{u3} \\ \beta_{u4} \end{bmatrix} = \boldsymbol{A}\boldsymbol{\gamma}_u,$$

where K_j is the number of categories for categorical variable j and the matrices $\boldsymbol{I}, \boldsymbol{0}$, and $\boldsymbol{1}\boldsymbol{1}'$ are understood to be adapted to the corresponding sizes depending on the relevant lengths of the vectors $\boldsymbol{\alpha}_u$ and $\boldsymbol{\beta}_{uj}$.

Let $\boldsymbol{\theta}_u' = [\boldsymbol{\gamma}_u', \delta, \boldsymbol{\tau}', \boldsymbol{\sigma}']$ be the unconstrained vector of all uniquely defined parameters. Then, the reparametrization function $\boldsymbol{h}(\boldsymbol{\theta}_u)$ can be written as

$$\boldsymbol{\theta} = \boldsymbol{h}(\boldsymbol{\theta}_u) = [\boldsymbol{h}_{\boldsymbol{\gamma}}'(\boldsymbol{\gamma}_u), \delta, \boldsymbol{h}_{\boldsymbol{\phi}}'(\boldsymbol{\tau}), \boldsymbol{h}_{\boldsymbol{\phi}}'(\boldsymbol{\sigma})]'.$$

Thus, the constrained maximization of $\boldsymbol{\theta}$ in (2.1) is equivalent to the unconstrained maximization

$$L(\boldsymbol{\theta}) = L(\boldsymbol{h}(\boldsymbol{\theta}_u) = L_u(\boldsymbol{\theta}_u).$$

To maximize $L_u(\boldsymbol{\theta}_u)$ over $\boldsymbol{\theta}_u$, we use the BFGS quasi-Newton algorithm as implemented in the `optimx` package in `R`. After convergence, one step of the Newton-Raphson method with a numerically approximated Hessian is performed to ensure the gradient to be close to zero.

It is well known that for maximum likelihood estimation, at a maximum $\boldsymbol{\theta}_u^*$, the parameters are normally distributed $\boldsymbol{\theta}_u \sim N(\boldsymbol{\theta}_u^*, \boldsymbol{\Sigma}_u)$ where $\boldsymbol{\Sigma}_u = -(\nabla^2 L_u(\boldsymbol{\theta}_u^*)^{-1})$ is the inverse of the negative Hessian of $L_u()$ evaluated at $\boldsymbol{\theta}_u^*$. As we would like to have the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}^* = \boldsymbol{h}(\boldsymbol{\theta}_u^*)$, the Delta method is applied so that

$$\boldsymbol{\theta}^* \sim N(\boldsymbol{h}(\boldsymbol{\theta}_u), \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \tag{2.2}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \nabla \boldsymbol{h}'(\boldsymbol{\theta}_u) \boldsymbol{\Sigma}_u \nabla \boldsymbol{h}(\boldsymbol{\theta}_u)$ and $\nabla \boldsymbol{h}(\boldsymbol{\theta}_u)$ is the Jacobian matrix of $\boldsymbol{h}()$, that is,

$$\nabla \boldsymbol{h}(\boldsymbol{\theta}_u) = \begin{bmatrix} \nabla \boldsymbol{h}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_f^*) & 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 0 & \nabla \boldsymbol{h}_{\boldsymbol{\phi}}(\boldsymbol{\tau}_u^*) & \mathbf{0} \\ \mathbf{0} & 0 & \mathbf{0} & \nabla \boldsymbol{h}_{\boldsymbol{\pi}}(\boldsymbol{\sigma}_u^*) \end{bmatrix},$$

where $\nabla \boldsymbol{h}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_f^*) = \boldsymbol{A}$, $\nabla \boldsymbol{h}_{\boldsymbol{\phi}}(\boldsymbol{\tau}_u^*)$, and $\nabla \boldsymbol{h}_{\boldsymbol{\pi}}(\boldsymbol{\sigma}_u^*)$ are the Jacobians for $\boldsymbol{h}_{\boldsymbol{\gamma}}()$, $\boldsymbol{h}_{\boldsymbol{\phi}}()$, and $\boldsymbol{h}_{\boldsymbol{\pi}}()$, respectively. The standard errors of the model parameters in $\boldsymbol{\theta}$ are the square roots of the diagonal elements of the covariance matrix in (2.2).

2.B Significance Testing for Weighted Means

Obtaining a test for the differences between weighted means as presented in Table 2.3, some non-standard steps are needed. Let $\boldsymbol{P}$ be the $N \times S$ matrix of the a posteriori probabilities from the CMM. Then, regress the a posteriori probabilities in $\boldsymbol{P}$ on a CBCL symptom subscale without intercept through OLS. The obtained regression coefficients $\boldsymbol{\psi}$ can be transformed to the weighted averages as follows

$$\boldsymbol{\psi}^* = \boldsymbol{B}\boldsymbol{\psi}, \text{ with } \boldsymbol{B} = \text{Diag}(\boldsymbol{P}'\mathbf{1})^{-1}\boldsymbol{P}'\boldsymbol{P},$$

where the operator Diag(.) transforms a vector into a diagonal matrix. We can do a Wald test with null hypothesis that the weighted means are the same ($\boldsymbol{\psi}^* = c\mathbf{1}$). The standard errors needed for the Wald test can be derived from the diagonal elements of the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\psi}^*} = \boldsymbol{B}'\boldsymbol{\Sigma}_{\boldsymbol{\psi}}\boldsymbol{B}$, where $\boldsymbol{\Sigma}_{\boldsymbol{\psi}}$ is the original covariance matrix obtained from the linear regression.

2.C CMM Expected Values

As the CMM provides a probability distribution for each number of cards turned over on each trial, obtaining predictions from a Censored Mixture Model (CMM) is not straightforward. Therefore, we choose the expectation as a point estimate for the predicted value. These expectations can be obtained from the following steps. First, the estimated regression coefficients $\hat{\alpha}_s$ and $\hat{\beta}$ are used to compute the linear combination

$$\hat{\mu}_{its} = \log(\exp(\hat{\alpha}_s + \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}}) + 1).$$

Second, for each individual, in each trial, and for each segment we can compute the probabilities of all possible outcomes $y \in \{0, \ldots, 32\}$,

$$\Pr(Z_{it} = \ell \mid \hat{\mu}_{its}, \hat{\delta}) = \begin{cases} \hat{\phi}_4 f(0 \mid \hat{\mu}_{its}, \hat{\delta}) + \hat{\phi_1} & \text{if } \ell = 0, \\ \hat{\phi}_4 f(\ell \mid \hat{\mu}_{its}, \hat{\delta}) + \frac{1}{|A|}\hat{\phi_2} & \text{if } \ell \in A, \\ \hat{\phi}_4 f(31 \mid \hat{\mu}_{its}, \hat{\delta}) + \hat{\phi_3} & \text{if } \ell = 31, \\ \hat{\phi}_4 f(32 \mid \hat{\mu}_{its}, \hat{\delta}) & \text{if } \ell = 32, \\ \hat{\phi}_4 f(\ell \mid \hat{\mu}_{its}, \hat{\delta}) & \text{for all other values } \ell, \end{cases} \tag{2.3}$$

where $f(\ell \mid \hat{\mu}_{its}, \hat{\delta})$ is the probability mass function of the negative binomial distribution, $A = \{4, 8, 10, 12, 16, 20, 24\}$, $|A|$ is the cardinality of set A, and $\hat{\phi}_j$ is the estimate of ϕ_j. The distributions in Figures 2.7, 2.8, 2.9, and 2.10 are obtained by weighting for the segments, using the estimated prior segment probability $\hat{\pi}_s$, and then summing over the individuals, trials, and segments

$$h_{it}(y) = \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{S} \hat{\pi}_s \Pr(Z_{it} = \ell \mid \hat{\mu}_{its}, \hat{\delta}) \quad \forall\ \ell \in \{0, \ldots, 32\}.$$

Note that the right graphs of these figures are multiplied by the probability of being (un)censored to get the same scale as the left graph.

The expected values per person and per trial can be obtained from the probabilities in (2.3), that is,

$$\hat{y}_{it} = \sum_{s=1}^{S} \hat{\pi}_s \left(\sum_{\ell=0}^{M} \ell \Pr(Z_{it} = \ell \mid \hat{\mu}_{its}, \hat{\delta}) + M\hat{\phi}_1(1 - F(M+1 \mid \hat{\mu}_{its}, \hat{\delta})) \right).$$

The probability mass above M, denoted by $1 - F(M+1 \mid \hat{\mu}_{its}, \hat{\delta})$, is added to the expected values as if it were the probability mass at M. Since the negative binomial distribution has an infinite upper bound, M should be large to obtain an accurate expected value. The probability mass above $M = 100$ is smaller than 0.0001 and so will not have a meaningful effect on the expected value. Therefore, we choose $M = 100$. Again, the expected values are multiplied by the probability of being uncensored to allow for a fair comparison with the observed outcome.

2.D Censored Mixture Model Applied to the Balloon Analogue Risk Task

The Censored Mixture Model (CMM) can be more generally applied than only to the Columbia Card Task (CCT). To show this, we have also analysed the Balloon Analogue Risk Task (BART) with the CMM. In this supplementary material we shall apply the CMM to a small data set and report the fit measures and overall distributions.

The data used in this supplementary study is provided by Dekkers et al. (2020). The data contains 180 boys, aged 12-19 years old, and 81 adolescents have ADHD.

Next, we have information about the Social Economic Status (SES), IQ, and ethnicity (Dutch 76.7%, other western 55.6%, and non-Western 17.8%). Additionally, we use the variables previous loss and second previous loss as predictors. These variables indicate whether the balloon in the (second) previous trial exploded or not. This leads to a total of $12+2S-1$ ($= S$ segment specific intercepts $\boldsymbol{\alpha} + 8$ regression coefficients $\boldsymbol{\beta} + 1$ scale parameter $\delta + 3$ weight parameters $\boldsymbol{\tau} + S - 1$ segment probabilities $\boldsymbol{\sigma}$) estimators that need to be estimated in a model with S segments. Every participant played the BART thirty times, meaning that we have 5400 ($= 180$ adolescents $\times$ 30 trials) observations to estimate these parameters. More information about the data collection can be found in Dekkers et al. (2020).

Similar as with the Columbia Card Task (CCT), the BART has certain outcomes that are more attractive than others. We distinguish four categories: (a) $k = 50$, (b) multiples of ten except for 50 (i.e., $k \in A$ with $A = \{10,\ 20,\ 30,\ 40,\ 60,\ 70,\ 80,\ 90,\ 100\}$), (c) multiples of five that are not multiples of ten (i.e., $k \in B$ with $B = \{5, 15, 25, 35, 45, 55, 65, 75, 85, 95\}$), and (d) otherwise. The same computation settings as in the Section 3.4 are used. Note that we did not split the data in a training and test set, and therefore have no out-of-sample performance measures, because the data set is too small.

The Bayesian information criterion (BIC) is lowest for the model with five segments, see Figure 2.12. Although the first segment in the model with five segments is small, see Table 2.8, we believe that the difference between the segment specific intercepts α_s, in particular the difference between the first segment and the other segments, are relevant. Therefore, we choose the model with five segments. Table 2.9 displays the game characteristics weighted by the a posteriori probabilities per segment. The stars denote that for all characteristics at least one of the segment averages is significantly different from the overall average. Furthermore, the regression coefficients are presented in Table 2.10. Similar as in the CCT, the results from the previous and second previous trial have a large impact on the behavior in the current trial. Last, we found a reasonable good correlation of 0.92, Figure 2.14, with an RMSE of 17.8 and MAD of 15.4. Figure 2.13 displays the distributions of the empirical and predicted number of pumps for the uncensored observations. The Hellinger distance between these two distributions is 0.27.

Table 2.8: Segment probabilities π_s and segment specific intercepts α_s, with the standard errors between brackets for CMMs with S = 2 to 6 segments.

		Segment s					
S		1	2	3	4	5	6
2							
	π_s	0.324	0.676				
		(0.042)	(0.042)				
	α_s	35.696	65.563				
		(1.614)	(2.065)				
3							
	π_s	0.112	0.323	0.565			
		(0.028)	(0.042)	(0.046)			
	α_s	24.125	45.88	70.347			
		(1.394)	(1.799)	(1.777)			
4							
	π_s	0.062	0.301	0.456	0.181		
		(0.019)	(0.040)	(0.051)	(0.044)		
	α_s	12.445	33.921	55.52	78.583		
		(1.895)	(1.820)	(2.101)	(3.772)		
5							
	π_s	0.028	0.094	0.251	0.533	0.095	
		(0.012)	(0.024)	(0.041)	(0.054)	(0.045)	
	α_s	2.298	24.893	41.335	59.375	87.706	
		(1.862)	(1.696)	(1.478)	(1.947)	(8.790)	
6							
	π_s	0.028	0.094	0.246	0.479	0.081	0.073
		(0.012)	(0.024)	(0.041)	(0.155)	(0.139)	(0.047)
	α_s	2.147	24.801	41.213	58.36	69.333	90.971
		(1.901)	(1.763)	(1.536)	(2.304)	(16.324)	(8.423)

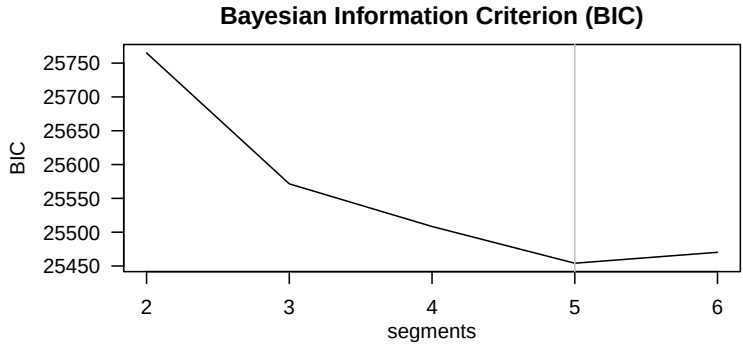


Figure 2.12: The Bayesian information criterion (BIC) of CMMs with $S = 2$ to 6 segments in the BART.

Table 2.9: Weighted scores per segment of game characteristics.

	Segment s					
	1	2	3	4	5	Total
π_s	0.03	0.09	0.25	0.53	0.09	
Average score ***	21.3	22.0	26.8	28.3	26.7	27.0
# pumps ***	30.8	28.6	38.6	45.5	49.4	42.1
# censored trials ***	7.8	6.3	10.1	13.4	15.8	12.0

A Wald test is performed to check for a significant difference between the segments. Three stars denotes $p < 0.01$.

Table 2.10: Regression coefficients with their standard errors. Within a categorical variable the sum of coefficients sum to zero and the continues variables IQ, Social Economic Status, and age are standardized.

	β-coefficients	(st error)
IQ	0.547	(0.616)
Social Economic Status	0.743	(0.811)
Age	1.275	(0.492)
Previous loss yes	-2.899	(0.420)
Previous loss no	2.899	(0.420)
Second previous loss yes	-1.163	(0.444)
Second previous loss no	1.163	(0.444)
Ethnicity		
Western	3.146	(2.866)
Non-Western	-1.714	(1.617)
Dutch	-1.432	(1.411)
ADHD group	0.046	(0.478)
Control group	-0.046	(0.478)

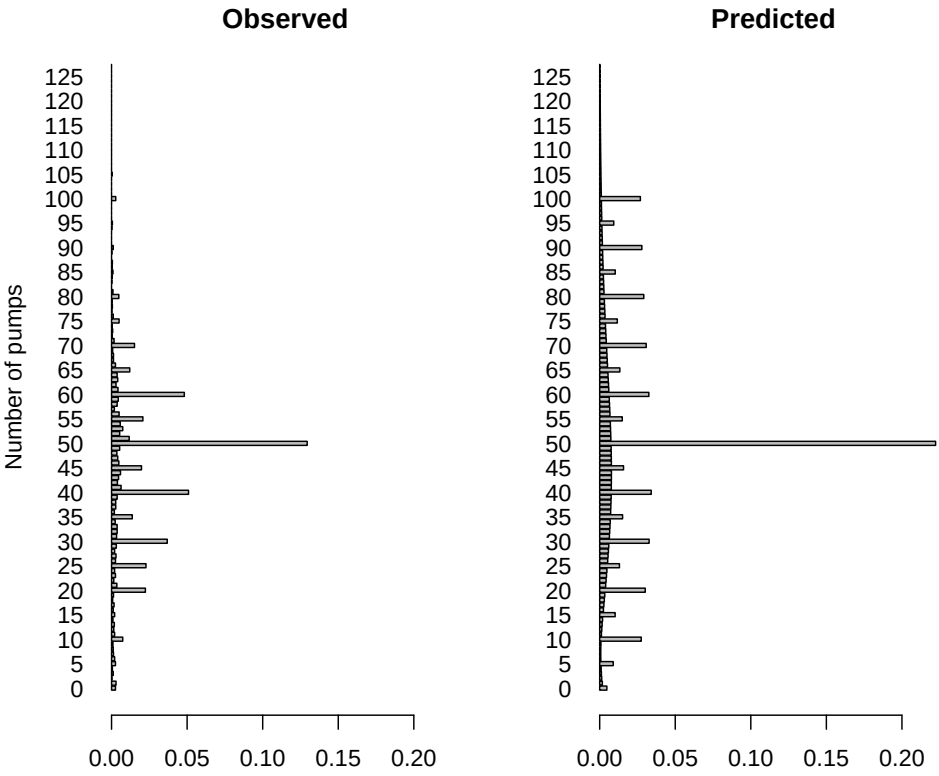


Figure 2.13: Distribution of the empirical (left panel) and predicted by the CMM (right panel) number of pumps for the uncensored observations in the BART.

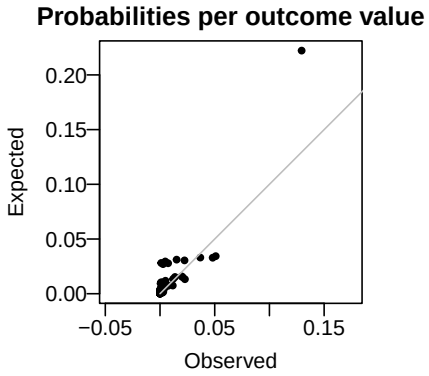


Figure 2.14: Scatterplot of the observed and expected probabilities per outcome value $\{0, 127\}$

2.E Censored Mixture Model With Segment Specific Effects for the Game Setting Variables

In addition to segment specific intercepts, the CMM also allows for segment specific effects for covariates. Here we present results of a CMM with four segments and segment specific effects for the game settings gain amount, loss amount, and number of loss cards. The linear combination in this model becomes

$$\eta_{its} = \alpha_s + \tilde{\boldsymbol{x}}'_{it}\tilde{\boldsymbol{\beta}}_s + \boldsymbol{x}^{*\prime}_{it}\boldsymbol{\beta}^*,$$

where $\tilde{\boldsymbol{x}}'_{it}$ contains the game settings gain amount, loss amount, and number of loss cards, $\tilde{\boldsymbol{\beta}}_s$ are the segment specific effects for these game settings, and $\boldsymbol{x}^{*\prime}_{it}$ contains the covariates excluding the game setting variables. Note that we did not include interaction terms between the game settings and sex in this analysis, because these effects are captured by the segment specific effects.

The segment specific effects for the game settings $\tilde{\boldsymbol{\beta}}_s$ are, next to the segment specific intercept α_s and segment probabilities π_s, presented in Section 2.5.5. The other results, presented in this supplementary material, are very similar to the results from the analysis with only segment specific intercepts, discussed in Section 3.4.

The regression coefficients from Table 2.11 have a linear interpretation, because our link function is close to the identity function, see Figure 2.4 in Section 2.4. Also, the numerical variables (age and IQ) are standardized to z-scores prior to the analysis and the categorical variables are constrained to have mean weights of the categories belonging to a single variable equal to zero, so that the intercept can be interpreted as the average score in the segment for a neutral child. Similar to the analysis with only segment specific intercepts, girls take more risk than boys and IQ is negatively associated with the number of cards turned over. Also, a household income between 2000 and 4000 euro's per month is related to lower risk taking levels. Children with a mother with a Dutch or Dutch Antilles ethnicity turn over fewer cards than the base

average and children with a mother with an African, Moroccan, or Turkish ethnicity turn over more cards than the base average. Furthermore, children with a mother who had no or primary education take more risk than children with a mother with a higher education. Moreover, the results from the previous round have a strong effect on observed behavior in the current round. In case a loss card was encountered in the previous or second previous round on average 1.6 and 1 cards less are turned over, respectively.

Again, the risk averse children in Segment 1 and the risk seekers in Segment 4 on average have more behavioral problems than the children in Segment 2 and 3. Particularly, the risk averse children are more anxious and aggressive. The risk seekers, on the other hand, have more attention problems and show more delinquent behavior than the children in other segments.

Last, we found a reasonable good correlation of 0.97, see Figure 2.15, with an RMSE of 8.5 and MAD of 5.5. Figures 2.16 and 2.17 display the distributions of the empirical and predicted number of cards turned over for the uncensored and censored observations, respectively. The Hellinger distance between the observed and predicted distribution of the uncensored cases is equal to 0.08. To guard against overfitting, we provide the same performance measures for the test set containing 1049 children. The out-of-sample RMSE is equal to 8.4, the MAD is equal to 5.4, and the correlation is equal to 0.97. The distributions of the empirical and predicted out-of-sample number of cards turned over for the uncensored and censored observations are presented in Figures 2.18 and 2.19, respectively. The out-of-sample Hellinger distance equals 0.08.

Table 2.11: Regression coefficients with their standard errors from the model with segment specific effects for the game setting variables. Within a categorical variable the sum of coefficients sum to zero and the continues variables age and IQ are standardized.

	β - coefficients (st error)	
Age	-0.105	(0.062)
Boy	-0.310	(0.072)
Girl	0.310	(0.072)
IQ	-0.350	(0.089)
Ethnicity mother		
Dutch	-0.686	(0.142)
Asian	-0.047	(0.231)
African	1.408	(0.349)
Moroccan	0.604	(0.301)
Dutch Antilles	-3.263	(0.349)
Surinamese	0.429	(0.277)
Turkish	1.155	(0.303)
Other Western	0.398	(0.247)
Education mother		
No or primary education	0.543	(0.193)
Secondary education	-0.206	(0.124)
Higher education	-0.337	(0.125)
Householdincome per month in euro's		
< 2000	0.133	(0.132)
$2000 - 4000$	-0.284	(0.100)
> 4000	0.151	(0.111)
Previous loss yes	-0.820	(0.039)
Previous loss no	0.820	(0.039)
Second previous loss yes	-0.489	(0.039)
Second previous loss no	0.489	(0.039)

Table 2.12: Weighted z-scores per segment of CBCL subscales scores and other CCT characteristics for the model with segment specific effects for the game settings.

	Segment s				
	1	2	3	4	Total
π_s	0.09	0.27	0.36	0.28	
CBCL subscales					
Internalizing *	0.08	-0.01	-0.05	0.05	0.00
Externalizing **	0.09	-0.04	-0.04	0.06	0.00
CBCL symptom scales					
Anxiety **	0.10	0.00	-0.05	0.04	0.00
Social withdrawl **	0.02	0.02	-0.06	0.05	0.00
Somatic complaints	0.06	-0.03	0.00	0.02	0.00
Social problems ***	0.11	-0.05	-0.06	0.09	0.00
Thought problems **	0.11	-0.02	-0.06	0.06	0.00
Attention problems ***	-0.01	-0.05	-0.06	0.13	0.00
Delinquent behavior **	0.00	-0.04	-0.03	0.08	0.00
Aggressive behavior **	0.12	-0.03	-0.04	0.05	0.00
Average score ***	-87.6	-120.8	-170.0	-228.1	-165.7
# cards turned over ***	4.9	7.5	9.9	11.7	9.3
# censored trials ***	5.7	8.4	11.3	14.3	10.8

A Wald test is performed to check for a significant difference between the segments. One star denotes $0.05 \leq p < 0.10$, two $0.01 \leq p < 0.05$, and three $p < 0.01$. The 223 children without a CBCL score measured at either six or nine years old were excluded.

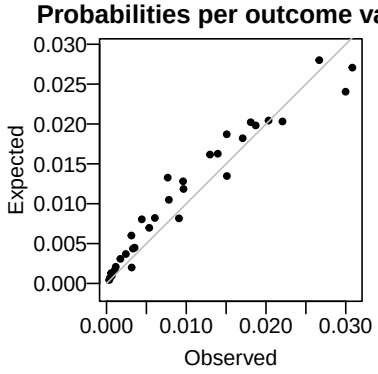


Figure 2.15: Scatterplot of the observed and expected probabilities per outcome value $\{0, 31\}$ for the model with segment specific effects for the game setting variables.

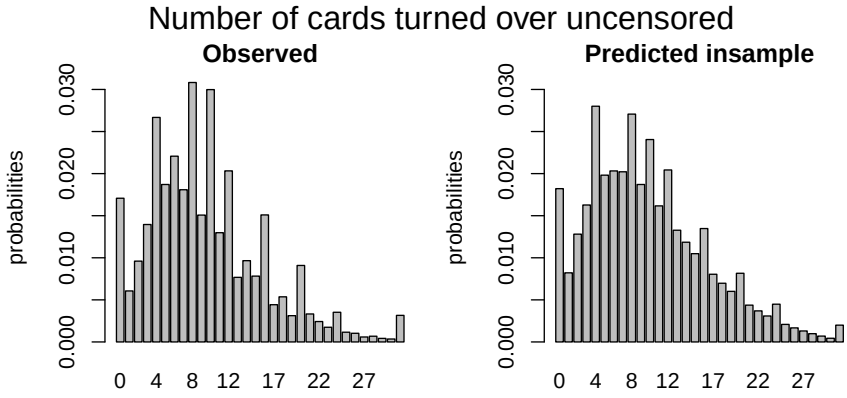


Figure 2.16: Distribution of the empirical (left panel) and predicted by the CMM with segment specific effects for the game settings (right panel) number of cards turned over for the uncensored observations in the training data.

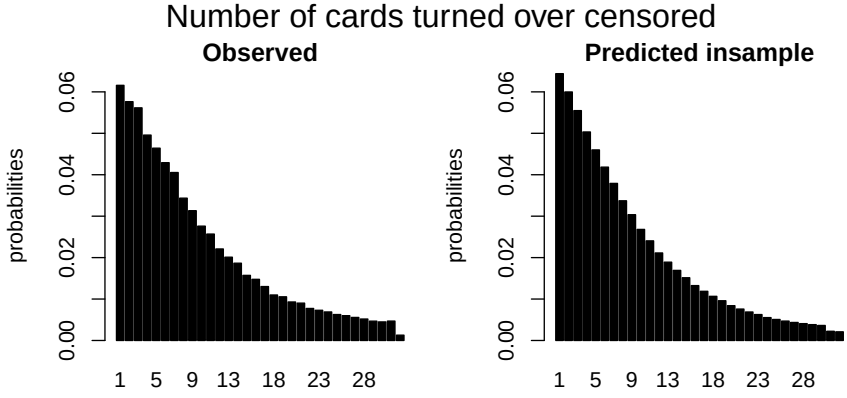


Figure 2.17: Distribution of the empirical (left panel) and predicted by the CMM with segment specific effects for the game settings (right panel) number of cards turned over corrected for the probability of being censored per card in the training data.

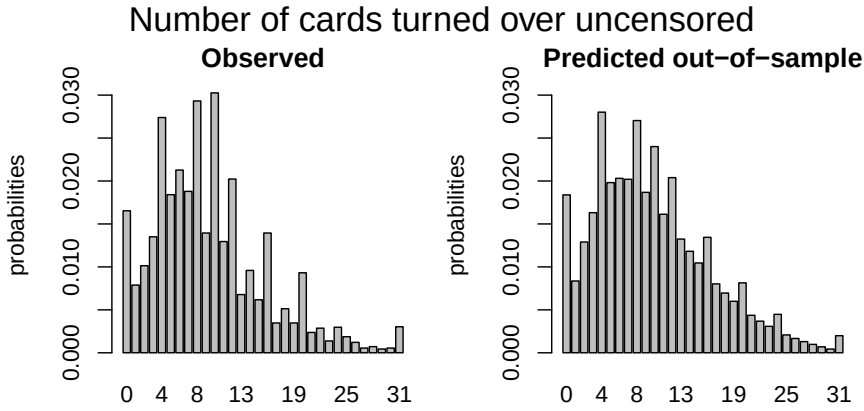


Figure 2.18: Distribution of the empirical (left panel) and predicted by the CMM with segment specific effects for the game settings (right panel) number of cards turned over for the uncensored observations in the test data.

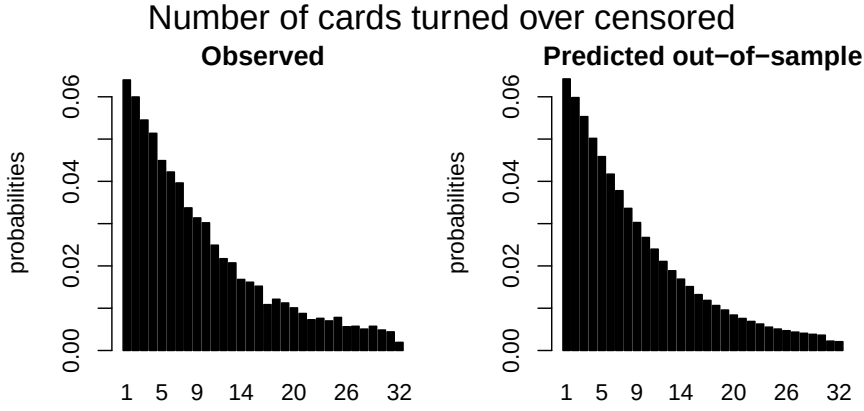


Figure 2.19: Distribution of the empirical (left panel) and predicted by the CMM with segment specific effects for the game settings (right panel) number of cards turned over corrected for the probability of being censored per card in the test data.

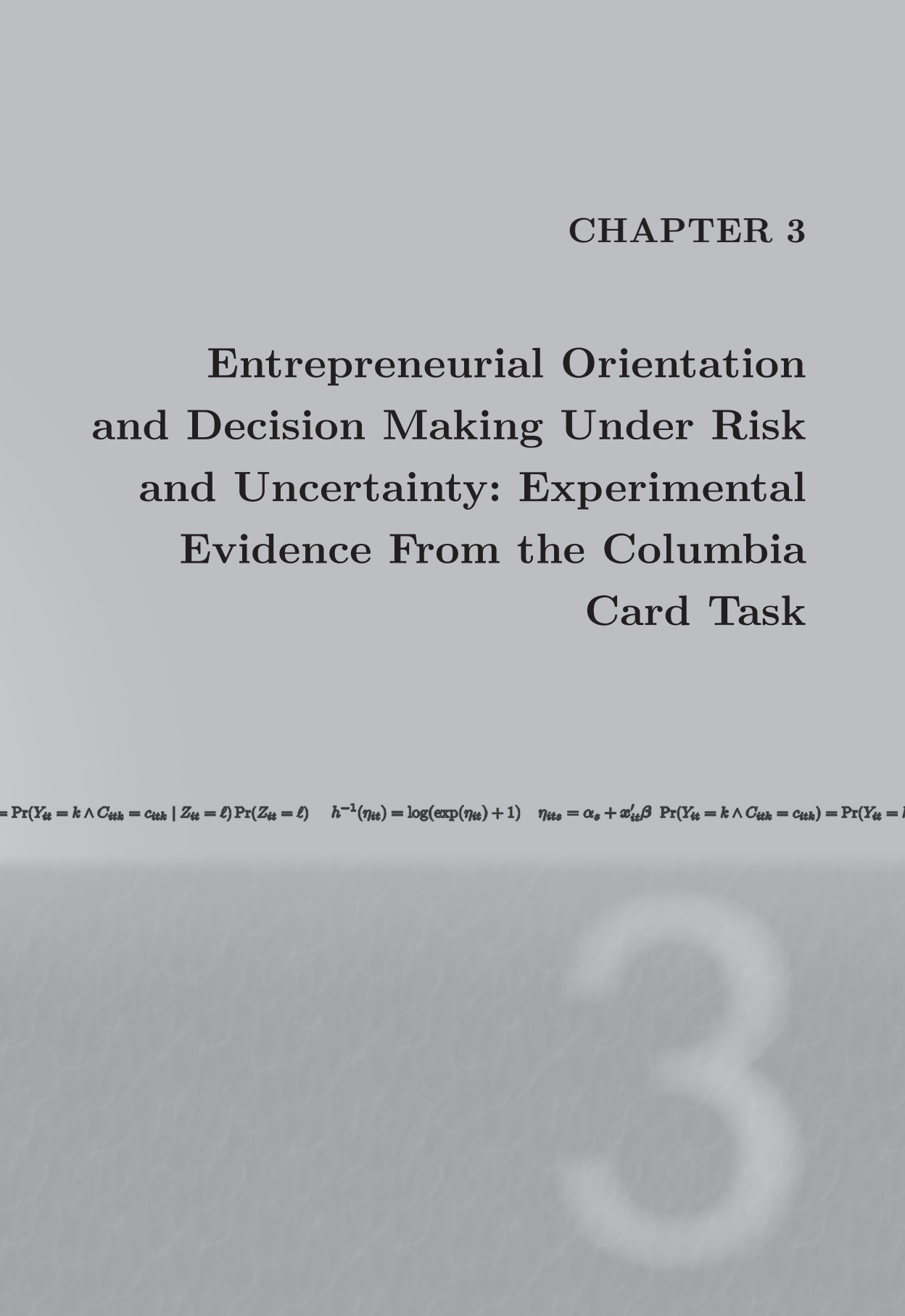
$$\Pr(Y_{it}=k\wedge C_{itk}=c_{itk})=\Pr(Y_{it}=k\wedge C_{itk}=c_{itk}\mid Z_{it}=\ell)\Pr(Z_{it}=\ell)\quad h^{-1}(\eta_{it})=\log(\exp(\eta_{it})+1)\quad \eta_{its}=\alpha_s+x_{it}'\beta\quad \Pr(Y_{it}=k\wedge C_{itk}=c_{itk})=$$

$$\left(\prod_{i=1}^{N}\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{it}w_{it}}\right)\qquad \left(\prod_{i=1}^{N}\sum_{s=1}^{S}\pi_s\prod_{t=1}^{T}\Theta_{z_{it},c_{it}w_{it},s}\right)\qquad \left(\prod_{i=1}^{N}\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{it}w_{it}}\right)$$

CHAPTER 3

Entrepreneurial Orientation and Decision Making Under Risk and Uncertainty: Experimental Evidence From the Columbia Card Task

$= \Pr(Y_{itt} = k \wedge C_{ittk} = c_{ittk} \mid Z_{itt} = \ell) \Pr(Z_{itt} = \ell) \quad h^{-1}(\eta_{itt}) = \log(\exp(\eta_{itt}) + 1) \quad \eta_{ittc} = \alpha_c + x'_{itt}\beta \quad \Pr(Y_{itt} = k \wedge C_{ittk} = c_{ittk}) = \Pr(Y_{itt} =$



Abstract

With risk and uncertainty being central concepts in entrepreneurship theories, it is surprising that recent 'lab-in-the-field' studies using lottery tasks question the uniqueness of an entrepreneur's willingness to accept risks. Driven by concerns about the ecological validity of lottery tasks, we conducted a lab experiment among 127 students in which we elicited decision making under risk and uncertainty using two incentivized versions of the warm Columbia Card Task. We find robust evidence that individual entrepreneurial orientation (IEO) is negatively related to decision making under risk and uncertainty. While the same relationship holds for the proactiveness and innovativeness subscales of IEO, the relationship is positive for the risk taking subscale. Moderation analyses show that heterogeneous sensitivity towards possible gains and losses explains the main relationship.

3.1 Introduction

Risk and uncertainty are core concepts in entrepreneurship theories (Cantillon, 1931; Kihlstrom and Laffont, 1979; Kirzner, 1973; Knight, 1921; Say, 1803), and a stream of studies emphasizes that behavioral preferences make that some individuals are more willing to accept the risks and uncertainties that come with entrepreneurship than others (Astebro et al., 2014; Rauch and Frese, 2007). The upsurge of the focus on behavioral preferences and the entrepreneurial personality finds amongst others grounds in the notion that engagement in entrepreneurship is partly heritable (Nicolaou et al., 2008; Van der Loos et al., 2013; Vladasel et al., 2021). Still, the empirical evidence about the relationships between risk and uncertainty preferences on the one hand and entrepreneurship on the other hand is rather mixed as a result of the use of non-representative samples, self-reports, and diverse study-contexts (for recent reviews, see Holm et al., 2013; Koudstaal et al., 2016).

To bypass these limitations, two recent studies conducted 'lab-in-the-field' experiments among entrepreneurs, managers and employees to derive risk preferences from actual choice behavior in (incentivized) lottery tasks (Holm et al., 2013; Koudstaal et al., 2016). In these experiments, participants are asked to choose between two options that both specify the chance of winning (or losing) a certain amount of money. In these designs, the chance of winning (or losing), the direction of the outcome (gain or loss), as well as the outcome magnitude is systematically varied across trials. Based upon observed behavior in these trials, one can describe the participants' behavior using expected utility theory (Von Neumann and Morgenstern, 1953) or even more comprehensively using (cumulative) prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Wakker, 2010). That is, the lottery design enables the description of individual behavior in terms of outcome sensitivity, aversion to loss, and probability weighting.

Both Holm et al. (2013) and Koudstaal et al. (2016) find that while entrepreneurs do differ in their self-reported preferences for taking risks and in how risk-seeking they perceive themselves to be, they do not differ from other people when it comes to actual choice behavior under risk and uncertainty (although the latter study finds some differences with regards to loss aversion). With risk preferences being of fundamental importance in most theories about entrepreneurship (Van Praag, 1999), a possible explanation for these surprising null-findings regarding the relationship between risk behavior and entrepreneurship could be the low ecological validity of the behavioral tasks used by Holm et al. (2013) and Koudstaal et al. (2016). Lottery tasks have been criticized for being artificial, possibly impacting the external or ecological validity of the experimental results (Pedroni et al., 2018; Schonberg et al., 2011). Importantly, the absence of immediate feedback in these lottery tasks makes that affective responses (senses of escalating tension and exhilaration) are not elicited. Affective responses are, however, often central to real-life risk taking (Schonberg et al., 2011). Hence, in addition to systematic trial-to-trial variation, an experimental risk taking task should ideally employ a dynamic decision situation that elicits affective engagement in individuals.

In the present study, we use such an experimental task, the Columbia Card Task

(CCT Dijkstra et al., 2022; Figner et al., 2009), to reassess the relationship between risk aversion and individual-level aspects of entrepreneurship. This reassessment is important, because the relationship between risk taking and entrepreneurship is a cornerstone in entrepreneurship theory, and as such the finding that risk taking behavior in lottery tasks is not different for entrepreneurs and non-entrepreneurs begs for an explanation (Holm et al., 2013; Koudstaal et al., 2016). In the CCT, participants have to turn virtual cards from a desk in a series of computer-based trials. Most of the cards are win cards. Turning these cards makes individuals earn points. However, a small number of cards are loss cards, and turning those makes participants lose points. In each trial, three pieces of information are supplied to participants: (1) the number of points they will gain when turning a win card, (2) the number of points they will lose when turning a loss card, and (3) the number of loss cards present in the trial. These three parameters are systematically varied over the trials and relate to the main components in cumulative prospect theory, namely outcome sensitivity, sensitivity to losses, and probability weighting. Thus, despite the CCT being developed in the field of developmental psychology, the task is also well-suited for use in economic research.

Different versions of the CCT exist. In the so-called hot CCT, individuals turn cards one by one – receiving feedback after every turn – until they voluntarily stop and cash the points or until they turn a loss card, at which point the specified loss amount is subtracted from the points earned and the trial forcedly stops. A disadvantage of the hot CCT is that for trials in which a loss card is turned, there is no information about the number of cards the participant intended to turn (given the information available to them). This issue is called statistical censoring. Censoring is not an issue in the cold version of the CCT, in which participants are asked to indicate at the start of every trial how many cards in total they would like to turn, after which an algorithm determines the trial's outcome. However, the absence of feedback is a marked disadvantage of the cold CCT in numerous research contexts, as it removes the affective engagement from the task. More recently, the two original CCT versions have been combined into a third version, the warm CCT (Huang et al., 2013). In this version, participants select the cards they would like to turn at the start of a trial (to avoid censoring) and then press a button that prompts the selected cards to turn (to provide feedback and hence to elicit affective engagement). Because of these attractive properties, we use the warm version of the CCT.

In our student sample, we relate the participants' behavior in the experimental task to a well-known determinant of entrepreneurship, namely the participants' individual entrepreneurial orientation (IEO). Based on the original firm-level scale for Entrepreneurial Orientation (EO Lumpkin et al., 2009; Lumpkin and Dess, 1996), IEO (Bolton and Lane, 2012) measures the traits and attitudes that are inherent in the original EO scale at the individual level. Specifically, the scale comprises three subscales: risk taking (venturing into uncertain circumstances), innovativeness (engaging in creativity and experimentation), and proactiveness (identifying opportunities promptly). The importance of studying entrepreneurial orientation on an individual level has been demonstrated repeatedly. Bolton and Lane (2012) already show in a sample of students that IEO and its subscales positively associate with

intentions for entrepreneurship, a finding that is supported by Koe (2016) and Sahoo and Panda (2019). Hence, IEO is considered to be a determinant of actual entrepreneurial involvement. Furthermore, IEO has been related to competitive strategy (Lechner and Gudmundsson, 2014) and entrepreneurial performance (Bolton, 2012). Together, firm-level entrepreneurial orientation and individual entrepreneurial orientation are considered to be cornerstones in entrepreneurship research (Covin et al., 2020; Rauch et al., 2009). However, the two studies most directly related to the present study employ a measure for entrepreneurship related to actual occupational involvement in entrepreneurship (Holm et al., 2013; Koudstaal et al., 2016). We discuss the impact of this difference in Section 3.5, but note here that the use of a student sample comes with the advantage that in such a sample the reverse impact of occupational experiences on traits and attitudes is likely to be small (Bernoster et al., 2018).

We find robust evidence that students' IEO is related to decision making under risk in the warm CCT. Overall, IEO is negatively related to risk taking behavior in the CCT. The IEO subscales reveal the origin of this overall negative relationship: individuals scoring high on the risk subscale are more risk-seeking in the CCT, while those who score high on the innovativeness or proactiveness subscale exhibit more risk averse behavior. Moderation analyses show that heterogeneous sensitivity to possible gains and losses in a trial rather than sensitivity to success chances in the task explains the relationships. For a second series of experimental trials, we modified the warm CCT in such a way that information about the number of loss cards was not provided to the participants in order to reflect a situation of 'true' uncertainty (or ambiguity). The results in the modified experiment are similar to the results in the original CCT, but effect sizes are much smaller. Multiple robustness checks confirm the robustness of these findings. Hence, in contrast to Holm et al. (2013) and Koudstaal et al. (2016), who largely find that entrepreneurs do not differ from other people when it comes to decision making in lottery tasks, our results using an ecologically more valid behavioral task (the CCT) do provide evidence for a relationship between individual-level aspects of entrepreneurship and decision making under risk and uncertainty.

The remainder of our study is organized as follows. Section 3.2 discusses the experimental design of the warm CCT and our modifications of it, and Section 3.3 describes the data collection and statistical methodology employed in our study. Section 3.4 presents the empirical results, while Section 3.5 discusses the findings and concludes.

3.2 Experimental Design

In the warm version of the Columbia Card Task (Huang et al., 2013), participants see a virtual array of 32 (4×8) clickable cards that are positioned face down. Participants are informed that the majority of these cards are 'win cards', which make them earn points that can be redeemed for money at the end of the experiment. However, there are also some 'loss cards' for which points (and thus money) will be subtracted if turned. An example of a trial setup participants encounter is presented in Figure 1. While originally the win and loss cards have respectively yellow smiley faces and red

frowny faces (Figner et al., 2009; Huang et al., 2013), the present study adopts the design with a green plus sign and a red minus sign to reduce the possible impact of (abstracted) facial features on individuals' emotion and thereby decision making.

The points earned per turned win card and the points lost when turning a loss card are presented at the top of the screen. These parameters are systematically varied across trials. The win amount is either 10 or 30, and the loss amount is either 250 or 750. In addition, the number of loss cards in the trial is presented. This number is either 1 or 3. The three information parameters (win points, loss points, and number of loss cards) are orthogonally varied across trials by means of a full factorial design. This means that every possible combination of values (e.g., plus 10 points per win cards, minus 250 points when turning a loss card, and the presence of 3 loss cards) ensues. Running all these combinations once takes $2 \times 2 \times 2 = 8$ trials. For the present study, all combinations were run six times (in random order), resulting in $6 \times 8 = 48$ trials per participant.

In every trial, participants first select the cards they want to turn (or click 'no card' if they prefer to turn no cards). Then, they press the centrally positioned 'turn' button, after which the cards automatically turn in the order they were selected in, unless a loss card is encountered. When a loss card is turned, the trial forcedly stops and the specified loss points are subtracted from the already accumulated win points. The number of cards selected to be turned over is considered a measure for risk taking behavior, because the likelihood of experiencing a loss increases with each card that is turned over Figner et al., 2009. Because of the systematic variation of experimental parameters, the impact of the three main components of cumulative prospect theory on the decision making can be estimated: outcome sensitivity (10 vs. 30 points per win card), loss sensitivity (-250 vs. -750 points in case of a loss card), and probability weighting (1 vs. 3 loss cards in the trial).

With precise information about the parameters in a trial, study participants can in theory determine the number of cards they need to turn to optimize their expected payoff. This situation resembles the classic situation of decision making under risk (Knight, 1921). An important advantage of the CCT is its adjustability to reflect a situation of decision making under uncertainty. By masking information about the number of loss cards in each trial, study participants lack the necessary information to optimize their expected payoff. However, they are still exposed to possible gains and losses when turning cards. In the present study, we use another series of $6 \times 8 = 48$ trials to investigate the relationship between IEO and decision making under uncertainty using the same experimental stimuli as in the first 48 trails (10 vs. 30 points per win card and -250 vs. -750 points in case of a loss card). If participants asked the experimenter about the number of loss cards in these trials of the CCT (which was stated to be 'unknown' but was in fact still either 1 or 3), experimenters answered that this number was random and could be anything between 0 and 32.

3.3 Data and Methodology

The following section is concerned with the data collection process and the statistical methodology employed in this study. First, we discuss the participant recruitment

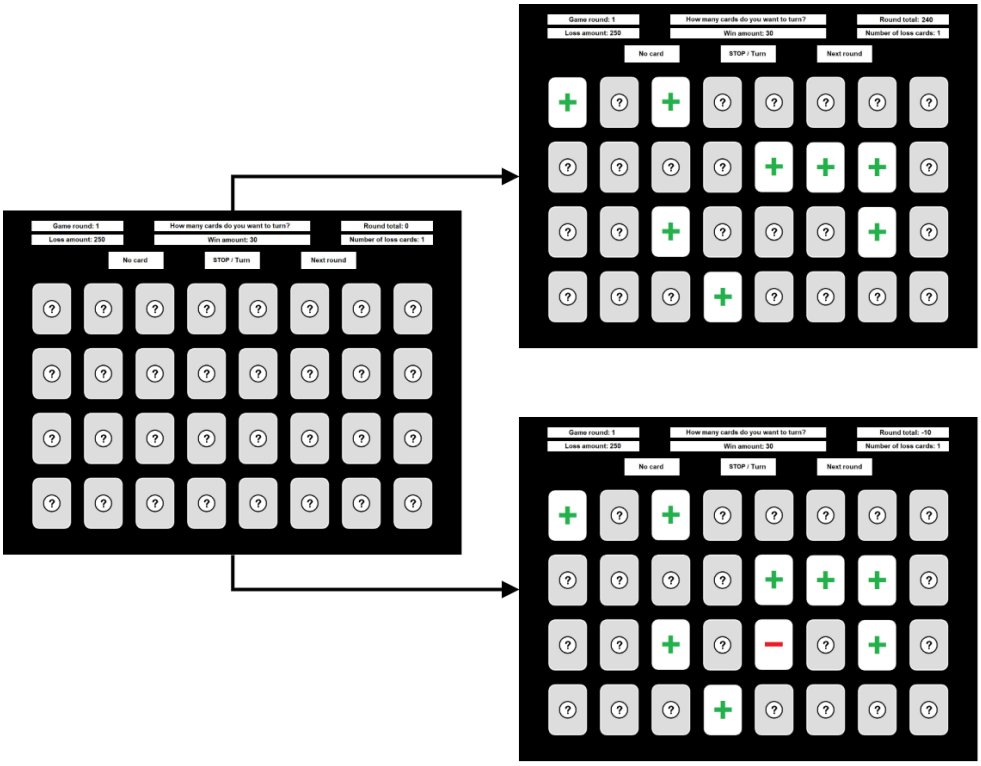


Figure 3.1: Experimental setup of the warm Columbia Card Task (CCT). The left screen shows the array of cards as presented to participants at the start of each trial. In this example, the win amount is 30 points, the loss amount 250 points, and only 1 loss card is present among all cards. At the start of the trial, the participant selects the cards they want to turn. They then press the STOP/Turn button to initiate the turnover of selected cards. The upper right screen shows a scenario in which they press the button after selecting 8 cards, and when no loss card is encountered. In this scenario, the participant earns $8 \times 30 = 240$ points. The lower right screen shows a scenario in which the participant presses the STOP/Turn button after selecting 9 cards, and in which the 9th card is a loss card. In this scenario, the earnings are $8 \times 30 - 250 = -10$ points.

and data collection procedure. Second, the variables are explained in detail. Last, we elaborate on the model employed in this study. The current model lies between the Censored Mixture Model (CMM) from Chapter 2 and a generalized linear model. That is, not all features of the CMM are used in this study. For instance, in the warm CCT censoring does not occur and we do not implement segment specific intercepts to model unobserved heterogeneity. The last part of this section, elaborates in detail the features of the CMM that are implemented in the current model and how the current model deviates from the CMM developed in Chapter 2.

3.3.1 Study Participants Recruitment and Data Collection Procedure

Participants voluntarily signed up for the experiment via one of three online participant systems running at the Dutch university the experiment was conducted at. One system was only accessible to psychology students; the other two were available to students from all disciplines. Before signing up, students were informed that the study comprised of an online questionnaire and a visit to the behavioral lab. In addition, psychology students were informed that they would receive course credit for their participation. Other students received a standard fee of 17.5 euros. The base fee offered was typical for the system via which participants registered themselves. In addition to the base fee, all students were informed they could earn up to 7.5 euros extra based on task performance.

After signing up, participants received an e-mail which confirmed their timeslot for the lab session. In addition, they were asked to not consume any alcohol, coffee, or energy drinks the day of their appointment in order to prevent these substances from impacting their (risk taking) behavior. The e-mail also contained a link to a secured online questionnaire, which participants were asked to fill out prior to their lab session and which included informed consent forms to be (digitally) signed. The full questionnaire (including demographic details and the surveys used in the present study) took 10 to 15 minutes to complete. Upon arrival in the behavioral lab, participants once more signed for informed consent, this time on paper forms.

Participants performed the experimental tasks while being seated in a comfortable chair in a dimly lit sound-attenuated room. The order in which the two tasks were presented was counterbalanced. Tasks were presented on a 19-inch computer monitor with a 1024×768 resolution from a viewing distance of approximately 130 cm. The tasks were programmed and presented using E-prime software (Psychology Software Tools, 2018). Before starting the task, participants received on-screen instructions, after which the experimenter confirmed that the participant had understood all of these. Each set of 48 trials (the original warm CCT and the modified warm CCT) lasted approximately 15 minutes. After finishing the lab session, participants were debriefed about the aims of the study.

3.3.2 Measures

The main outcome variable is the number of cards selected in a trial. Hence, for each individual there are 48 observations per CCT. As the CCT presents participants with 32 clickable cards, this variable ranges between 0 and 32. A secondary outcome variable is the overall performance in terms of points earned over the 48 trials of each of the two CCT versions.

The main explanatory variable is individual entrepreneurial orientation (IEO). The IEO questionnaire of Bolton and Lane (2012) consists of 10 items, of which three pertain to risk taking, four to innovativeness, and three to proactiveness. These dimensions of entrepreneurial orientation may vary independently from each other (Lumpkin and Dess, 1996) and are therefore relevant to examine separately, in addition to examining participants' total IEO score. All items can be found in Table 3.5 in Appendix 3.A. Participants answered the items on a 5-point Likert-scale (1 'strongly disagree' to 5 'strongly agree'), resulting in a total score that ranges from 10 to 50, with higher scores indicating stronger entrepreneurial orientation. In our analysis sample ($N = 127$), Cronbach's alpha is 0.70 for the full IEO construct, but only 0.64 for risk taking, 0.65 for innovativeness, and 0.42 for proactiveness. These relatively low reliabilities are likely partially driven by the small number of items per subscale, as Cronbach's alpha is typically lower when computed across fewer items (Pallant, 2007).

As control variables we include dummy variables for the experimental conditions in a specific trial: 10 vs. 30 points per win card, -250 vs. -750 points in case of a loss card, and 1 vs. 3 loss cards in the trial. Moreover, we include a dummy variable indicating whether study participants played the original or modified CCT first (to deal with sequence effects, we counterbalanced the order in which participants performed the tasks). Because of the experimental design and the relatively homogeneous sample of students from the sample Dutch university, we limited the further control variables to Sex (Female=1, Male=0) and Age (in years).

To facilitate easy interpretation of effect sizes in the regression models, the continues variables (i.e., IEO and its subscales and age) are standardized to have mean zero and standard deviation one in the analysis sample.

3.3.3 Statistical Methodology

To explain the number of cards selected in a trial in the CCT, we employ a model derived from the Censored Mixture Model (CMM), which was specifically designed for the CCT by Dijkstra et al. (2022). The CMM constitutes a generalized linear model with some additional features. Although the CMM is able to model unobserved heterogeneity through mixtures, we do not use that possibility here, because the number of participants N is relatively small. In addition, we do not use the ability of the CMM to accommodate censoring, because we exploit the warm CCT where censoring does not occur. Because of the nonnegative (0-32) and discrete (0,1,2, ...) nature of the dependent variable, the CMM assumes the dependent variable to result from a negative binomial distribution. The negative binomial distribution is preferred over the Poisson distribution, because it allows the mean and variance of

the distribution of the dependent variable to be different. In the negative binomial distribution, the mean is specified through an inverse link function. To be able to interpret the regression coefficients in a linear fashion, we specify the inverse link function as follows:

$$h^{-1}(\eta_{it}) = \log(\exp(\eta_{it}) + 1),$$

with η_{it} being the linear combination of the explanatory variables for individual i at trial t included in the $1 \times p$ vector $\boldsymbol{x}'_{it}$, where p is the number of variables. That is,

$$\eta_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta},$$

with α being the intercept and $\boldsymbol{\beta}$ the regression coefficients.

Importantly, the CMM offers flexibility regarding the probability assigned to certain outcomes. Because of the setup of the CCT, an array of 4×8 clickable cards, participants often choose multiples of four cards (e.g., one complete column or row, see Figure 3.2 for the distribution of the dependent variable in our samples). Within the CMM, these excesses are categorized and extra probability mass is assigned to these categories. The first category consists of the outcome zero and the second category contains the outcomes four, eight, ten, twelve, sixteen, and twenty.

Our main analyses constitute three different models which are similar in terms of the dependent variable (the number of selected cards) and control variables (Sex and Age). In the first model, the main explanatory variable is IEO. In the second model, the main explanatory variables are the subscales of IEO: Risk taking, Innovativeness, and Proactiveness. This model allows us to analyze which subcomponents of IEO are driving the results in the first model. In the third model, we extend the first model with interaction terms between IEO and the experimental parameters related to reward sensitivity (10 vs. 30 points per win card), loss sensitivity (-250 vs. -750 points in case of a loss card), and probability weighting (1 vs. 3 loss cards in the trial). Hence, the third model allows us to analyze whether heterogeneous sensitivity to possible gains, losses, or success chances explains the relationship between IEO and decision making under risk and uncertainty in the first model. We estimate these three models using data collected with the original CCT as well as using data collected with the modified CCT. Hence, in total we estimate six different models.

3.4 Results

This section provides the findings of this research. First, the data characteristics are presented. In the second part, we present the results obtained with the regressions on the number of cards selected, as discussed in subsection 3.3.3. The findings of the Risk sample and Uncertainty sample for all three models are compared. In addition, we perform a robustness check where participants who showed unconventional experimental behavior are excluded. The next analysis investigates the performance on the CCT. A linear regression is implemented and the average points earned is regressed on the control variables and IEO (Model 1) and its subscales (Model 2). Last, we explore the optimal strategy by maximizing the expected value and compare this to the behavior of the participants.

3.4.1 Descriptive Statistics

In total, 130 students participated in the study. Three of these participants did fill out the questionnaire (including the questions about IEO), but did not participate in the experiment (one for medical reasons and two did not show up for unknown reasons). These participants were excluded from all analyses. Because of incorrect software settings, three other participants were only able to complete 12 (out of 48) trials of the modified CCT. These participants were excluded from analyses on the modified CCT data. For the remainder of this article, the short-hand 'Risk sample' is used to refer to the data collected using the original warm CCT ($N = 127$) and 'Uncertainty sample' when referring to the data collected using the modified warm CCT ($N = 124$).

Table 3.1 contains the descriptive statistics of the two samples. Presumably because of the higher degree of uncertainty, the average number of selected cards is lower in the Uncertainty sample than in the Risk sample. Figure 3.2 provides a more detailed view on the distribution of the number of cards selected in the 2×48 trials. It is clear that in many trials not a single card is selected. Next to selecting zero cards, several other outcomes also seem particularly attractive: the outcomes four, eight, ten, twelve, sixteen, and twenty show clear peaks in both graphs. As explained before, these excesses can be explained by the 4×8 layout of the CCT. Creating a geometric pattern, such as complete rows or columns, corresponds to selecting a number of cards equal to a multiple of four.

The mean (unstandardized) values for IEO and its subscales are approximately in the middle of the respective scales. The slight differences between the descriptive statistics of the Risk sample and the Uncertainty sample are explained by the difference in sample size. The average age in both samples is 20.4 years, which corresponds approximately to third year bachelor students at the university where the experiments were conducted. A little more than half of the study participants is female.

Table 3.1: Descriptive statistics of the two analysis samples. Reported are the mean, standard deviation (SD), minimum value (Min), and maximum value (Max) for the Risk sample and Uncertainty sample. Note that IEO stands for individual entrepreneurial orientation.

	Risk sample ($N = 127$)				Uncertainty sample ($N = 124$)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Average #cards selected	7.79	6.69	0	32	6.69	6.06	0	32
IEO	35.94	4.92	22	48	35.98	4.91	22	48
IEO Risk taking	10.47	2.17	5	15	10.47	2.15	5	15
IEO Innovativeness	14.32	2.57	9	20	14.35	2.57	9	20
IEO Proactiveness	11.15	1.99	6	15	11.15	1.97	6	15
Age (in years)	20.41	2.19	17	29	20.40	2.20	17	29
Sex (female)	0.55	0.50	0	1	0.55	0.50	0	1

The pairwise correlations between the main variables in the analysis samples are shown in Table 3.2. The upper triangle reflects the correlations in the Risk sample,

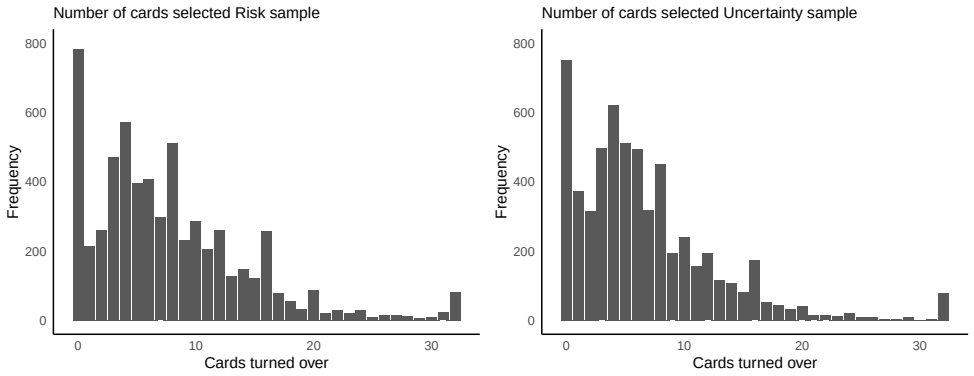


Figure 3.2: The number of cards selected per trial in the Risk sample (left panel) and Uncertainty sample (right panel).

Table 3.2: Correlation table of the analysis samples. The Risk sample is presented in the upper triangle and the Uncertainty sample in the lower triangle. Note that, individual entrepreneurial orientation is abbreviated to IEO and the stars denote the significance of the correlation coefficients, that is, one star denotes $0.05 \leq p < 0.10$, two $0.01 \leq p < 0.05$, and three $p < 0.01$.

	Risk sample						
Uncertainty sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Average #cards selected	-	−0.27***	−0.11	−0.22**	−0.27***	−0.02	0.03
(2) IEO	−0.02	-	0.75***	0.82***	0.59***	0.12	−0.18**
(3) IEO Risk taking	0.06	0.76***	-	0.47***	0.17*	0.10	−0.30***
(4) IEO Innovativeness	−0.03	0.82***	0.46***	-	0.22**	0.13	−0.11
(5) IEO Proactiveness	−0.08	0.60***	0.18**	0.23**	-	0.01	0.02
(6) Age	−0.05	0.13	0.09	0.13	0.04	-	0.08
(7) Sex (female)	0.13	−0.20**	−0.31***	−0.11	−0.00	0.10	-

and the lower triangle those in the Uncertainty sample. We find a significantly negative correlation between the average number of cards selected and IEO in the Risk sample. However, this correlation is insignificant in the Uncertainty sample. The table shows positive correlations across the subscales of the IEO in both samples, which is in line with the results of Bolton and Lane (2012).

3.4.2 Regression Results

Table 3.3 and 3.4 display the results of the regressions explaining the number of cards selected in a trial in the Risk sample and Uncertainty sample, respectively. Model 1 in both tables clearly shows that students with a high entrepreneurial orientation exhibit more risk averse behavior. That is, they select fewer cards when the probabilities of a loss card are known (Table 3.3) as well as in the situation where these probabilities are unknown (Table 3.4). Nevertheless, the effect of IEO is considerably smaller in

Table 3.4 than in Table 3.3.

Consistent with the results for the total IEO score, the effects of the IEO subscales (Model 2 in Table 3.3 and 3.4) are larger in magnitude in the Risk sample than in the Uncertainty sample. Reassuringly, students who indicate to be risk-seeking select more cards in the CCT than students who indicate to be less risk-seeking. The results also show that more innovative and/or proactive students take less risk than those who self-report to be less innovative and/or proactive. However, the result for innovativeness is not statistically significant in the Uncertainty sample.

In Model 3, we analyze whether heterogeneous sensitivity to possible gains, losses, or success chances explains the relationship between IEO and decision making under risk and uncertainty in Model 1. The moderation results show that the interaction effects of IEO with the gain amount and loss amount are statistically significant in both samples, but that the interaction with the number of loss cards is insignificant. Hence, in both experimental settings, entrepreneurially oriented students are more sensitive to the gain and loss amount than students who are less entrepreneurially oriented.

The effects of the control variables are quite consistent across the different experimental setups and models. The effect of the game variables (gain amount, loss amount, and number of loss cards) are all in the expected direction. More specifically, participants are most sensitive to the risk probability (in case it is known), followed by the loss amount and gain amount. Besides, the effects of the loss amount and gain amount are stronger in the Uncertainty game compared to the Risk game. This could be explained by the fact that in the Uncertainty condition participants have less information and will value the information that is available more. Furthermore, females take more risk than males and older students are more risk averse compared to younger students.

When we compare the regression results in the Risk sample and Uncertainty sample, we see that there is a clear distinction with respect to the effect of the number of loss cards in a trial. When the success chance is known (Risk sample), this factor has a significant effect on the number of cards selected (Table 3.3). With more loss cards in the game, participants select fewer cards. When the loss probability is unknown (Uncertainty sample), we do not find a significant effect of the number of loss cards in a trial (Table 3.4). This result follows naturally from the modified experimental design, because the number of loss cards cannot play a role in decision making when this information is not available to the decision maker.

A second clear difference pertains to the effect of task order. Regardless of which task version is played first, participants take less risk (i.e., they select fewer cards) in their second task. The absolute effect of task order is somewhat stronger in the Uncertainty sample (~2 cards, Table 3.4) than in the Risk sample (~1 card, Table 3.3). Playing more and more trials in the CCT increases affective engagement, making participants increasingly aware of the difficulty of decision making under risk and uncertainty. As a result, on average, they tend to reduce the number of cards selected throughout the trials. This reduction seems to carry over to the second task the participants engage in. The results show that those participants for whom the Uncertainty game is their second task (due to counterbalancing) select ~2 cards fewer

than participants for whom their first task is the Uncertainty game.

Table 3.3: Results of the regressions explaining the number of cards selected in a trial (Risk sample). Presented are the regression coefficients β_j with standard errors in parentheses. The reference categories are 10 for Gain amount, 250 for Loss amount, and 1 for #Loss cards. The continuous variables, IEO (individual entrepreneurial orientation), its subscales, and age are standardized to z-scores. The stars denote the significance of the coefficients, that is, one star denotes $0.05 \leq p < 0.10$, two $0.01 \leq p < 0.05$, and three $p < 0.01$.

	Model 1		Model 2		Model 3	
IEO	−0.838***	(0.062)			−0.914***	(0.168)
IEO Risk taking			0.336***	(0.086)		
IEO Innovativeness			−0.588***	(0.074)		
IEO Proactiveness			−0.746***	(0.076)		
IEO × Gain amount (30)					0.615***	(0.136)
IEO × Loss amount (750)					−0.329**	(0.146)
IEO × #Loss cards (3)					0.194	(0.154)
Gain amount (30)	2.559***	(0.137)	2.311***	(0.138)	2.193***	(0.141)
Loss amount (750)	−3.230***	(0.146)	−3.113***	(0.143)	−3.049***	(0.146)
#Loss cards (3)	−4.447***	(0.159)	−4.700***	(0.156)	−4.821***	(0.159)
Task order (uncertainty first)	−1.143***	(0.131)	−1.336***	(0.141)	−1.201***	(0.140)
Sex (female)	1.157***	(0.127)	0.956***	(0.146)	0.632***	(0.142)
Age	−0.145**	(0.064)	−0.270***	(0.065)	−0.232***	(0.064)
Intercept	9.888***	(0.190)	10.887***	(0.193)	11.050***	(0.196)
Likelihood value	17819.34		17642.25		17668.15	
N	127		127		127	
$N \times T$	6096		6096		6096	

Table 3.4: Results of the regressions explaining the number of cards selected in a trial (Uncertainty sample). Presented are the regression coefficients β_j with standard errors in parentheses. The reference categories are 10 for Gain amount, 250 for Loss amount, and 1 for #Loss cards. The continuous variables, IEO (individual entrepreneurial orientation), its subscales, and age are standardized to z-scores. The stars denote the significance of the coefficients, that is, one star denotes $0.05 \leq p < 0.10$, two $0.01 \leq p < 0.05$, and three $p < 0.01$.

	Model 1		Model 2		Model 3	
IEO	−0.124**	(0.061)			−0.030	(0.139)
IEO Risk taking			0.196**	(0.078)		
IEO Innovativeness			−0.087	(0.061)		
IEO Proactiveness			−0.208***	(0.063)		
IEO × Gain amount (30)					0.543***	(0.128)
IEO × Loss amount (750)					−0.411***	(0.134)
IEO × #Loss cards (3)					0.126	(0.113)
Gain amount (30)	3.303***	(0.132)	3.160***	(0.129)	3.278***	(0.131)
Loss amount (750)	−4.177***	(0.137)	−3.958***	(0.134)	−4.153***	(0.137)
#Loss cards (3)	−0.094	(0.115)	−0.024	(0.114)	−0.087	(0.115)
Task order (uncertainty first)	1.992***	(0.125)	1.953***	(0.126)	1.976***	(0.124)
Sex (female)	1.168***	(0.122)	1.178***	(0.127)	1.192***	(0.122)
Age	−0.258***	(0.063)	−0.270***	(0.063)	−0.243***	(0.063)
Intercept	5.537***	(0.162)	5.709***	(0.161)	5.520***	(0.162)
Likelihood value	16625.83		16686.88		16608.48	
N	124		124		124	
$N \times T$	5952		5952		5952	

Some participants showed unconventional behavior during the task. First, a few participants repeatedly selected 32 cards, a response that will by definition lead to encountering a loss card. Although participants may select 32 cards once or twice to explore the structure of the CCT, after two times they should realize that there are indeed loss cards present in the trials and that they are not being deceived. Second, some participants showed signs of fatigue towards the end of the task, indicated by them repeatedly selecting the 'no card' option (thus selecting 0 cards). Therefore, we performed a robustness check in which we excluded participants who selected 32 cards more than twice across the full 48 trials and/or who selected 0 cards five times or more in the last nine trials. The results of the regressions with these participants excluded are presented in Appendix 3.B as Table 3.6 (Risk sample, $N = 88$) and 3.7 (Uncertainty sample, $N = 96$). The main results are similar to the ones presented in Table 3.3 and 3.4. The only differences pertain to the control variables. In the Risk sample, males now seem to take more risk than females, and the coefficient for age is not significant anymore. This seems to be the result of dropping relatively many males choosing frequently the 'no card' option in the last few trials, and the smaller analysis sample. In the Uncertainty sample, the coefficient for age has also become insignificant. In addition, the interaction terms between IEO and the game settings in Model 3 have become insignificant and in Model 2 the subscale innovativeness is now significant at the one percent significance level. Overall, it seems that unconventional

behavior of participants in the CCT is not driving our main results.

While the number of cards selected is the main outcome in the CCT, it is also important to analyze the performance on the CCT in terms of the total points earned, because the number of cards selected is the way through which participants try to maximize their compensation for taking part in the experiment. Table 3.8 (Risk sample) and 3.9 (Uncertainty sample) in Appendix 3.C present the results of linear regressions explaining the average number of points earned in a trial. In the Risk sample, only sex has a significant coefficient: the average points earned are somewhat higher for males than for females. In the Uncertainty sample, next to sex, also the task order and age have a significant effect on the performance in the CCT. In case the participant first played the Risk version, they perform better in the Uncertainty version. Furthermore, older participants perform better on the CCT. Interestingly, neither IEO (Model 1) nor its subscales (Model 2) significantly affect the performance in either of the two versions of the CCT. The absence of a relationship between IEO and the points earned in the CCT backs the customary interpretation of the number of cards selected (and not the points earned) as a proper proxy for risk preferences. After all, the number of cards a participant selects is a direct proxy of their risk preferences, whereas the eventual payoff is more distal and subject to other factors such as chance.

Furthermore, the optimal and risk neutral strategy can be computed by maximizing the expected value. Figure 3.3 in Appendix 3.D displays the curves of the expected values. The maximum expected values are presented in top left corners of the graphs and the corresponding optimal number of cards to select are displayed perpendicular to the maximum expected value at the line $y = 0$. In half of the cases, selecting zero cards is the optimal strategy, that is, the expected value is negative for every number of cards selected. Table 3.10 in Appendix 3.D presents the average points earned per game setting in the Risk sample, which is interesting to compare to the maximum expected values from Figure 3.3. In the cases where it is best to select zero cards, the average points earned are negative. In the other cases, the empirical average points earned are comparable to the maximum expected values, except for the most profitable game setting (gain amount 30, loss amount 250, and number of loss cards 1), where the average number of points earned (185.2) is almost half of the maximum expected value that could have been achieved (245.7).

3.5 Discussion and Conclusion

The results of our experiments show that the individual entrepreneurial orientation of the students in our sample is related to decision making under risk in the warm CCT. Overall, and somewhat surprisingly in light of the literature, IEO is negatively related to risk taking behavior in the CCT. The results for subscales of IEO are less surprising, namely that individuals who indicate to be more risk-seeking indeed take more risk and those who indicate to be innovative or proactive exhibit more risk averse behavior. The moderation analyses indicate that heterogeneous sensitivity to possible gains and losses rather than sensitivity to success chances explains the relationship between IEO and decision making under risk and uncertainty. We also

modified the warm CCT in such a way that the number of loss cards is unknown to the participants. This experimental setup reflects a situation of larger uncertainty than the conventional setup of the warm CCT. Overall, the results are similar in the modified CCT. The larger uncertainty however translates into smaller effect sizes of IEO and its subscales on decision making.

With risk aversion at the core of many theories about entrepreneurship (Van Praag, 1999), our main result regarding the negative relationship between IEO and decision making under risk and uncertainty is in need of an explanation. This explanation may lie in the multidimensionality of the IEO construct, because we do find that individuals self-reporting to be risk taking indeed take more risks in the CCT (both versions). Thus, not surprisingly, individuals reporting a high propensity for taking (entrepreneurial) risks actually do take more financial risks in a laboratory experiment than those who report to be less risk-seeking. Our results however suggest that the joint effect of proactiveness and innovativeness (which are negatively related to risk taking behavior in the CCT) are larger than the effect of the risk taking dimension of IEO. Being highly proactive may mean trying to reduce future uncertainties, resulting in a negative relationship between proactiveness and actual risky behavior. Similarly, more innovative people may also prefer to mitigate or actively manage risks, in order to have better chances of bringing their new products on the market. As such, in line with other studies (e.g., Craig et al., 2014; Dai et al., 2014; Kreiser and Davis, 2010), our results underscore the importance for future studies to appreciate heterogeneity in the entrepreneurship construct as the subscales of IEO relate differently to decision making under risk and uncertainty. The IEO subscales and other traits that have been shown to be relevant to entrepreneurship (such as autonomy and competitive aggressiveness Lumpkin and Dess, 1996) together are what separates those who are entrepreneurially oriented and may become entrepreneurs from those who are not entrepreneurially oriented and will gravitate more towards other types of employment.

Because of their experimental setup and identification strategy, it is most relevant to compare our findings to the results of the studies by Holm et al. (2013) and Koudstaal et al. (2016). These 'lab-in-the-field' studies, on the whole, find that individuals indicating to be risk averse do not exhibit particular risk averse behavior in lottery tasks. The explanation behind this somewhat surprising finding may lie in the ecological validity of lottery tasks, as discussed in the introduction. With the CCT, an experimental task known for its ecological validity, we do find a positive relationship between the self-reported risk taking dimension of IEO and the number of selected cards in the CCT. Our results therefore suggest that there might be more value in self-reports about risk preferences than concluded by Holm et al. (2013) and Koudstaal et al. (2016) based on their lottery experiments.

Nevertheless, the difference between our results and the results of Holm et al. (2013) and Koudstaal et al. (2016) may not only be explained by the different experimental paradigm used but also by the different entrepreneurship measures employed. The use of a student sample comes with the advantage that in it the reverse impact of occupational experiences on behavioral preferences is likely to be small (Bernoster et al., 2018), however it clearly comes with the disadvantage that IEO is not the same as

actual entrepreneurial behavior. When we on the one hand appreciate that students with high IEO exhibit intentions for entrepreneurship (e.g., Bolton and Lane, 2012), but that not all of them eventually become an entrepreneur, and on the other hand take into account that some individuals without particular strong entrepreneurial intentions during their academic studies may still engage in entrepreneurship later, it may be expected that the effect of IEO on decision making under risk and uncertainty amongst students is larger than the relationship between actual entrepreneurial involvement and risk taking behavior. Thus, the sample composition could possibly explain why our statistical evidence is relatively strong. Obviously, this conjecture could be tested in future studies by replicating the present study design with actual entrepreneurs and preferably several control groups, such as managers, employees, and also students.

Overall, our study provides clear evidence that entrepreneurial orientation is related to decision making under risk in the warm CCT, and that heterogeneous sensitivity to possible gains and losses rather than sensitivity to success chances explains this relationship. While risk aversion is central in many theories about entrepreneurship (e.g., Kihlstrom and Laffont, 1979), our findings specifically fit Kirzner's classic notion of the 'hunch' entrepreneurs have in situations of risk and uncertainty (Kirzner, 1973). That is, (prospective) entrepreneurs are sensitive to profitable opportunities by having a way of processing clues in the environment when taking calculated risks. When reducing the number of clues (as we do in the modified CCT), their behavior is much less different from others. As such, we show that the impact of behavioral preferences like risk and ambiguity aversion on actual behavior may depend on the information individuals can extract from their environment. Moreover, we conclude that the CCT provides a useful experimental setup to study the relationship between decision making under risk and uncertainty and entrepreneurial orientation – and, by extension, other aspects of entrepreneurship.

Appendix

3.A Individual Entrepreneurial Orientation Scale Items

Table 3.5: The 10 individual entrepreneurial orientation (IEO) scale items (Bolton and Lane, 2012). All items are answered on a 5-point Likert-scale (1 'strongly disagree' to 5 'strongly agree').

Subscale	**Items**
Risk Taking	I like to take bold action by venturing into the unknown
	I am willing to invest a lot of time and/or money on something that might yield a high return
	I tend to act 'boldly' in situations where risk is involved
Innovativeness	I often like to try new and unusual activities that are not typical but not necessarily risky
	In general, I prefer a strong emphasis in projects on unique, one-of-a-kind approaches rather than revisiting tried and true approaches used before
	I prefer to try my own unique way when learning new things rather than doing it like everyone else does
	I favor experimentation and original approaches to problem solving rather than using methods others generally use for solving their problems
Proactiveness	I usually act in anticipation of future problems, needs or changes
	I tend to plan ahead on projects
	I prefer to 'step-up' and get things going on projects rather than sit and wait for someone else to do it

3.B Robustness Check Unconventional Behavior

Table 3.6: Results of the regressions explaining the number of cards selected in a trial in the Risk sample (excluding individuals with unconventional experimental behavior). Presented are the regression coefficients β_j with standard errors in parentheses. The reference categories are 10 for Gain amount, 250 for Loss amount, and 1 for #Loss cards. The continuous variables, IEO (individual entrepreneurial orientation), its subscales, and age are standardized to z-scores. The stars denote the significance of the coefficients, that is, one star denotes $0.05 \leq p < 0.10$, two $0.01 \leq p < 0.05$, and three $p < 0.01$.

	Model 1		Model 2		Model 3	
IEO	−0.731***	(0.070)			−0.853***	(0.161)
IEO Risk taking			0.386***	(0.083)		
IEO Innovativeness			−0.629***	(0.068)		
IEO Proactiveness			−0.570***	(0.067)		
IEO × Gain amount (30)					0.583***	(0.134)
IEO × Loss amount (750)					−0.529***	(0.136)
IEO × #Loss cards (3)					0.279*	(0.149)
Gain amount (30)	1.829***	(0.141)	2.040***	(0.131)	1.748***	(0.141)
Loss amount (750)	−2.476***	(0.143)	−2.614***	(0.134)	−2.409***	(0.143)
#Loss cards (3)	−4.543***	(0.155)	−4.178***	(0.148)	−4.590***	(0.155)
Task order (uncertainty first)	−1.210***	(0.142)	−1.287***	(0.134)	−1.233***	(0.141)
Sex (female)	−0.516***	(0.155)	−0.088	(0.145)	−0.473***	(0.152)
Age	0.029	(0.069)	0.052	(0.068)	0.034	(0.069)
Intercept	11.258***	(0.202)	10.746***	(0.190)	11.274***	(0.202)
Likelihood value	11953.02		12037.86		11935.84	
N	88		88		88	
$N \times T$	4224		4224		4224	

Table 3.7: Results of the regressions explaining the number of cards selected in a trial in the Uncertainty sample (excluding individuals with unconventional experimental behavior). Presented are the regression coefficients β_j with standard errors in parentheses. The reference categories are 10 for Gain amount, 250 for Loss amount, and 1 for #Loss cards. The continuous variables, IEO (individual entrepreneurial orientation), its subscales, and age are standardized to z-scores. The stars denote the significance of the coefficients, that is, one star denotes $0.05 \leq p < 0.10$, two $0.01 \leq p < 0.05$, and three $p < 0.01$.

	Model 1		Model 2		Model 3	
IEO	−0.124**	(0.060)			−0.202	0.131
IEO Risk taking			0.184**	(0.074)		
IEO Innovativeness			−0.185***	(0.060)		
IEO Proactiveness			−0.113*	(0.061)		
IEO × Gain amount (30)					0.166	0.121
IEO × Loss amount (750)					−0.097	0.127
IEO × #Loss cards (3)					0.199*	0.112
Gain amount (30)	2.702***	(0.126)	2.702***	(0.126)	2.701***	0.126
Loss amount (750)	−3.614***	(0.131)	−3.591***	(0.131)	−3.608***	0.131
#Loss cards (3)	−0.111	(0.116)	−0.106	(0.116)	−0.115	0.116
Task order (uncertainty first)	1.842***	(0.121)	1.749***	(0.124)	1.842***	0.121
Sex (female)	0.732***	(0.124)	0.879***	(0.131)	0.733***	0.124
Age	−0.108	(0.066)	−0.128*	(0.067)	−0.106	0.066
Intercept	5.469***	(0.162)	5.414***	(0.163)	5.464***	0.162
Likelihood value	12547.41		12541.42		12544.39	
N	96		96		96	
$N \times T$	4,608		4,608		4,608	

3.C Results Average Number of Points Earned

Table 3.8: Results of the linear regressions explaining the average number of points earned in the 48 trials (Risk sample). Presented are the regression coefficients with standard errors in parentheses. Note that, the outcome variable, average number of points earned, and the continuous explanatory variables, IEO (individual entrepreneurial orientation), its subscales, and age are standardized to z-scores. The stars denote the significance of the coefficients, that is, one star denotes $0.05 \leq p < 0.10$, two $0.01 \leq p < 0.05$, and three $p < 0.01$.

	Model 1		Model 2	
IEO	0.038	(0.089)		
IEO Risk taking			−0.035	(0.104)
IEO Innovativeness			0.023	(0.100)
IEO Proactiveness			0.069	(0.090)
Task order (uncertainty first)	0.241	(0.174)	0.246	(0.176)
Sex (female)	−0.529***	(0.178)	−0.563***	(0.186)
Age	0.121	(0.087)	0.126	(0.088)
Intercept	0.174	(0.149)	0.19	(0.152)
R^2	0.09		0.10	
N	127		127	

Table 3.9: Results of the linear regressions explaining the average number of points earned in the 48 trials (Uncertainty sample). Presented are the regression coefficients with standard errors in parentheses. Note that, the outcome variable, average number of points earned, and the continuous explanatory variables, IEO (individual entrepreneurial orientation), its subscales, and age are standardized to z-scores. The stars denote the significance of the coefficients, that is, one star denotes $0.05 \leq p < 0.10$, two $0.01 \leq p < 0.05$, and three $p < 0.01$.

	Model 1		Model 2	
IEO	0.132	(0.087)		
IEO Risk taking			−0.008	(0.101)
IEO Innovativeness			0.124	(0.096)
IEO Proactiveness			0.054	(0.087)
Task order (uncertainty first)	−0.479***	(0.168)	−0.463***	(0.171)
Sex (female)	−0.484***	(0.173)	−0.514***	(0.180)
Age	0.158*	(0.085)	0.158*	(0.086)
Intercept	0.501***	(0.145)	0.510***	(0.147)
R^2	0.17		0.18	
N	124		124	

3.D Optimal Strategy

Table 3.10: Average points earned per game setting in the Risk sample and the maximum expected value when following the optimal strategy.

1 loss card		
	Loss amount	
Gain amount	250	750
Observed		
10	-5.1	-97.7
30	185.2	10.0
Optimal		
10	6.9	0
30	345.7	20.8

3 loss cards		
	Loss amount	
Gain amount	250	750
Observed		
10	-65.8	-150.2
30	11.9	-158.6
Optimal		
10	0	0
30	9.7	0

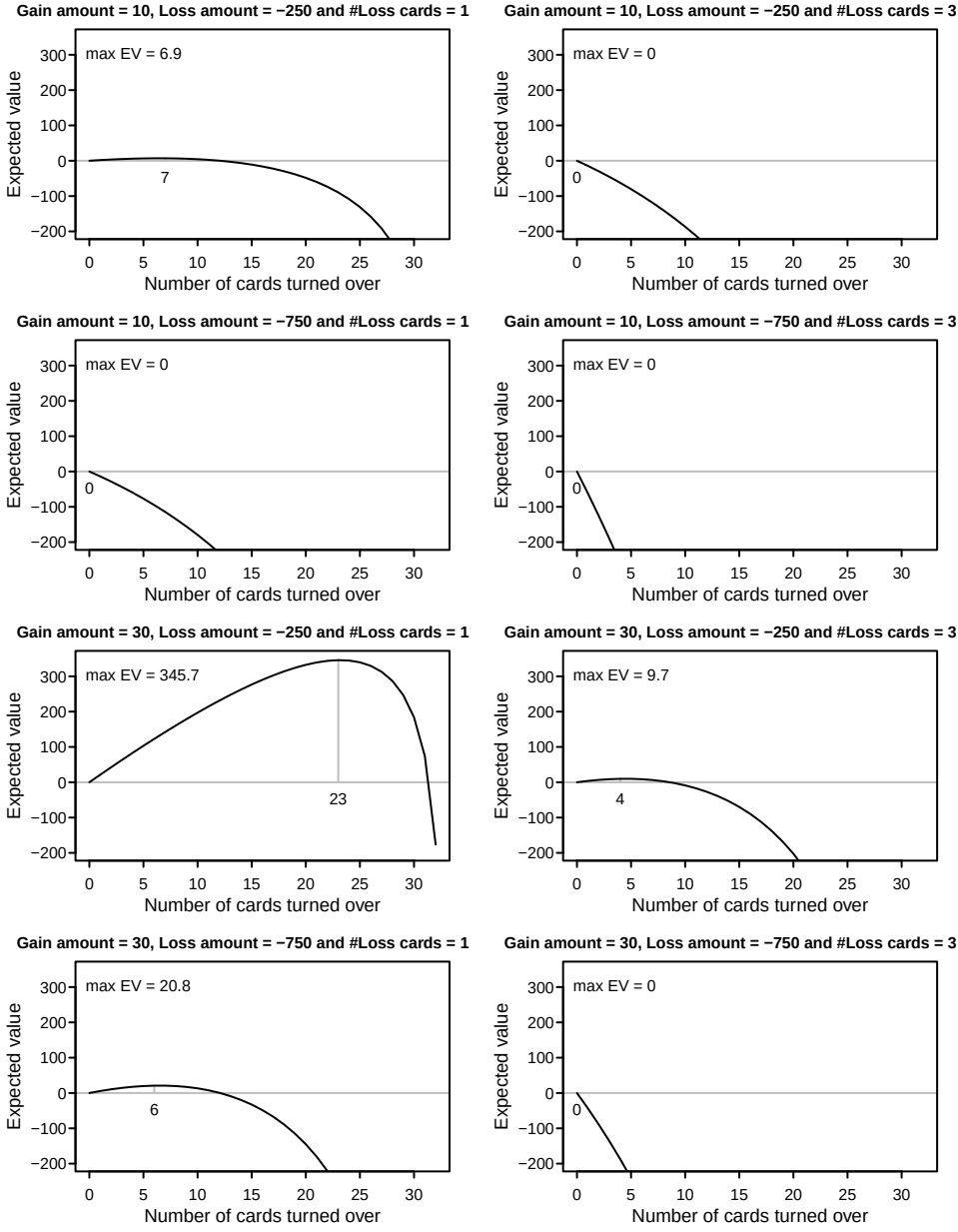


Figure 3.3: Curves of the expected values for all eight game settings. The maximum expected value is displayed in the top left corner of each graph (max EV) and the corresponding optimal number of cards is displayed perpendicular to the maximum expected value at the line $y = 0$.

PART II

Predicting Cognitive Ability Using Brain Morphology

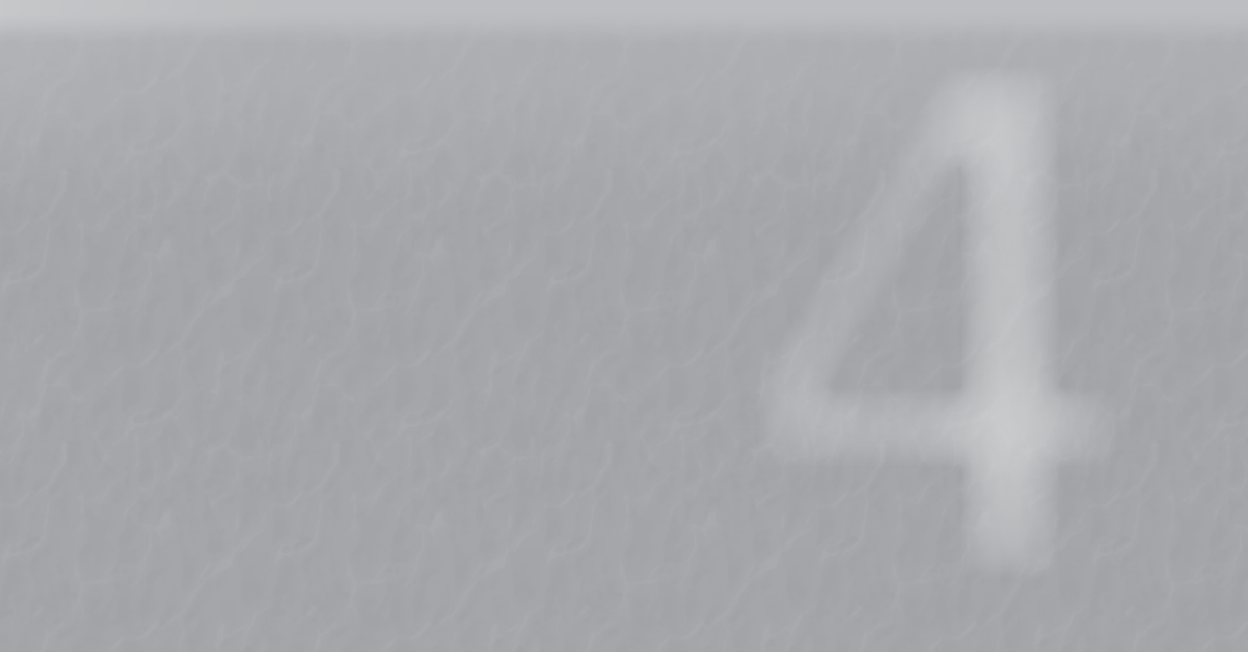
$$\Pr(Y_{it}=k\wedge C_{itk}=c_{itk})=\Pr(Y_{it}=k\wedge C_{itk}=c_{itk}\mid Z_{it}=\ell)\Pr(Z_{it}=\ell)\quad h^{-1}(\eta_{it})=\log(\exp(\eta_{it})+1)\quad \eta_{its}=\alpha_s+x_{it}'\beta\quad \Pr(Y_{it}=k\wedge C_{itk}=c_{itk})=$$

$$\left(\prod_{i=1}^{N}\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{it}w_{it}}\right)\qquad \left(\prod_{i=1}^{N}\sum_{s=1}^{S}\pi_s\prod_{t=1}^{T}\Theta_{z_{it},c_{it}w_{it},s}\right)\qquad \left(\prod_{i=1}^{N}\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{it}w_{it}}\right)$$

CHAPTER 4

Predicting Cognition of Children Through Cortical Surface Area

$= \Pr(Y_{it} = k \wedge C_{itk} = c_{itk} \mid Z_{it} = \ell) \Pr(Z_{it} = \ell) \quad h^{-1}(\eta_{it}) = \log(\exp(\eta_{it}) + 1) \quad \eta_{its} = \alpha_s + x'_{it}\beta \quad \Pr(Y_{it} = k \wedge C_{itk} = c_{itk}) = \Pr(Y_{it} =$



Abstract

Cognition is one of the most important functions of the brain. Many studies have shown that neuroimaging features, such as brain morphology, are predictive of cognitive ability. However, much of the previous work has focused on either parcellating the brain into anatomically divided regions, known as regions of interest (ROIs) or analyzing the brain in a vertex/voxel-wise framework, where every vertex is analyzed individually. The first approach has been criticized for the assumption that any underlying association with a phenotype is constrained by anatomical boundaries and for the loss of crucial information when summarizing across large areas of neural tissue. The shortcoming of the second approach is that effects of multiple vertices combined are ignored and that not all detailed information is used for predicting the phenotype. To overcome this criticism, we propose a twofold approach. First, we propose to use the whole brain, (i.e. almost 150,000 vertices per hemisphere), to release the assumption that a particular cognitive phenotype spatially correlates to the brain based on arbitrary anatomical parcellations. Second, we apply two machine learning techniques that capture all information in a multivariate way, using all 300,000 vertices at once. Better subset regression and adaptive lasso regression are compared to the frequently used vertex-wise approach. We use the exceptionally large data set of the ABCD Study, including $N > 10{,}000$ MRI scans, in particular the cortical surface area data. Data from the Generation R Study ($N > 3{,}000$ scans) is included to test the external validity. The cognitive task utilized is the Matrix Reasoning sub-test of the Wechsler Intelligence Test for Children-V (WISC-V), which serves as a proxy for nonverbal (performance) IQ.

4.1 Introduction

The rapid development of Magnetic Resonance Imaging (MRI) techniques has led to an explosive growth of cognitive neuroscience studies in the last twenty years. Many studies have shown the relation between neuroimaging features, such as brain morphology, and intellectual ability. For example, Fields (2008) investigates the relation between white matter volume in the brain and cognitive function, learning, and psychological disorders. Furthermore, features such as brain volume, cortical thickness, and gyrification have all been shown to be related to general intellectual ability (Reiss et al., 1996). Such associations between brain morphology and the performance on neuropsychological tests are called structure-function associations. Understanding this link between features of brain structure and neuropsychological tests at population level, is crucial for observing abnormal neurobiological processes in child psychopathology.

Multiple studies have shown the association between cortical thickness and cognitive ability. For instance, Karama et al. (2009) reports a positive association in healthy children (6-18 years) between cortical thickness and a general intelligence score based on the four subdomains Vocabulary, Similarities, Block Design, and Matrix Reasoning of the WASI (Wechsler Abbreviated Scale of Intelligence, (Wechsler, 1999)) in several regions of the brain. Also, Narr et al. (2007) finds clusters of positive associations between the Full Scale IQ and cortical thickness in the prefrontal and temporal lobe in both hemispheres. Similar associations are discovered by Menary et al. (2013). In addition, the authors examine the subtests of the WASI. They find a positive correlation between Performance IQ (subtests Block Design and Matrix Reasoning) and cortical thickness in the mid-superior temporal region in the right hemisphere. Negative correlations are revealed between the Vocabulary subtest and cortical thickness in the parahippocampal area and near the temporal-parietal junction in the left hemisphere and cortical thickness in the middle of the temporal lobe, inferior pariental lobe, superior occipital lobe and precuneus in the right hemisphere. Menary et al. (2013) divides the sample in a younger (9.0 to 16.45 years) and older (16.46 to 24.0 years) subsample to investigate the developmental patterns in these structure-function associations. The results obtained with the younger subsample are similar to the once obtained with the full sample. In the older subsample, only positive associations are found between a higher Performance IQ and greater cortical thickness.

Next to cortical thickness, cortical surface area is, although less frequently, associated with neuropsychological tests. For example, Vuoksimaa et al. (2016) demonstrates a positive relation between general cognitive ability and cortical surface area as well as cortical thickness. In addition, the authors find that the correlation between general cognitive ability and cortical surface area is stronger than the correlation with cortical thickness. Schnack et al. (2015) examines the development of cortical surface area in relation to IQ. Their sample contains over a thousand healthy individuals (9-60 years). They find that higher IQ is related to larger cortical surface area. In particular, this effect is strongest for children and decreases over their age. At the age of 60 this effect has vanished in the right hemisphere. Cognitive ability can be

distinguished in different subdomains. Colom et al. (2013) analyzes seven different intelligence factors and cognitive functions measured each by three different tests or tasks. They find that in the right hemisphere clusters of cortical surface area in the middle frontal gyrus correlate with fluid intelligence and working memory capacity. Furthermore, crystallized intelligence is associated with clusters of cortical surface area in the inferior frontal gyrus, again in the right hemisphere.

In our study, we focus on cortical surface area, because it is more stable over childhood than cortical thickness (Amlien et al., 2016; Duerden et al., 2020; Storsve et al., 2014). Much of the expansion of the cortex takes place during the prenatal period (Rakic, 1988). Therefore, even at an older age, cortical surface area serves as an attribute by which fetal environment and growth can be characterized.

Studying these structure-function associations, where brain morphology predicts cognitive performance, can provide insights in typical brain development. Once the baseline for typical brain development is established, we can examine atypical brain development related to cognitive deficit. Moreover, understanding the neurodevelopment at population level allows for the identification of abnormal neurobiological processes in child psychopathology.

Much of this previous work uses a parcellation of the brain into anatomically divided regions, known as regions of interest (ROIs), which are then analyzed in relation to a phenotype or exposure of interest. This approach certainly has several strengths, such as 'native space' and subject-specific accuracy, the use of ROIs has been criticized for the assumption that any underlying association with a phenotype is constrained by anatomical boundaries and for the potential loss of crucial information when aggregating across large areas of neural tissue. In addition, ROIs are often univariately correlated with the phenotype, which is most likely disadvantageous to their predictive power.

Another way to analyze MRI scans in relation to a phenotype is to univariately correlate vertices to the phenotype. It is needless to say that the predictive accuracy of a univariate approach is lower than the predictive accuracy in a multivariate approach. Analyzing vertices in a multivariate way (without prior vertex selection) is not possible with the simple regression techniques commonly used.

To overcome this criticism, we propose a twofold approach. First, we propose to use whole brain data to release the assumption that a particular cognitive phenotype must spatially correlate to the brain based on arbitrary anatomical parcellations. Second, we apply two machine learning techniques that elegantly capture all information in a multivariate way. In this way, we take into account all vertices at once.

Machine learning techniques have been widely applied, for instance, in the brain ageing literature: these techniques are applied to predict age using various brain metrics (Bellantuono et al., 2021; Peng et al., 2021). It is shown that the prediction error, that is, chronological age minus predicted age (called BrainAge), can serve as an indicator for brain diseases and risk of mortality. For example, Nenadić et al. (2017) shows that schizophrenia patients have a higher deviance in the BrainAge than healthy controls. Cole and Franke (2017) summarizes how BrainAge relates to several brain diseases, cognitive performance, and other age-associated health issues.

As BrainAge has shown promise as a useful neurobiological marker, we hypo-

thesize that cognitive ability within a similar construct may also be a fruitful target as a neurobiological marker. In the present study, we apply two machine learning techniques to predict nonverbal IQ, namely adaptive lasso regression (Zou, 2006) and better subset regression (Xiong, 2014), and we compare their predictive performance with that of the commonly used vertex-wise approach. Both machine learning techniques perform variable selection which allows, next to predicting IQ, to interpret the relevant brain regions, something most other machine learning techniques lack.

For this study, we use the exceptionally large data set of the Adolescent Brain Cognitive Development (ABCD) Study, including $N > 10{,}000$ MRI scans, in particular the cortical surface area data. Data from the Generation R Study ($N > 3{,}000$ scans) is included to test the external validity. The cognitive phenotype for both sets is nonverbal IQ measured with the Matrix Reasoning subtest of the Wechsler Intelligence Test for Children-V (WISC-V Wechsler, 2014).

Taken together, our contribution is that we apply machine learning techniques to an exceptionally large pre-adolescent data set in the area of cognitive ability. Importantly, Smith and Nichols (2018) already emphasized the need of large data sets in MRI studies. A large data set may identify subtle effects that are not significantly detectable in smaller samples. Several studies have looked at applying multivariate techniques to cognition. However, this literature suffers from three major flaws. First, it is uncommon to find papers analyzing a data set with $N > 250$ subjects. The use of such small data sets makes it difficult to significantly identify small effects for lack of power. Second, many studies are conducted within small clinical samples, which results in a lack of generalizability. Third, often a proper validation of the method is missing. Due to the small sample sizes, researchers are not able to set aside a test set to independently validate the method or even perform a proper cross validation. Combining the data from both the ABCD Study and Generation R Study solves these issues.

The remainder of this article is structured as follows. It starts with a discussion of the ABCD and Generation R data sets. In addition, the preprocessing of the images and the data imputation are explained. Next, the prediction methods vertex-wise linear regression, adaptive lasso regression, and better subset regression are elaborated. Subsequently, the results are presented and discussed. Last, we will give some guidance for future research.

4.2 Data

This study includes two cohort studies: the Adolescent Brain Cognitive Development (ABCD) Study (Garavan et al., 2018) and the Generation R Study (Kooijman et al., 2016). Data from the ABCD Study is collected at 21 research sites across the United States and almost 12,000 children of 9-10 years old participated the study. This study was designed to analyze the children's biological and behavioral development through adolescence into young adulthood. The Generation R Study is a large birth cohort study in Rotterdam, the Netherlands. The goal of this study is to investigate early environmental and genetic factors that influence children's growth, development, and health from birth onward. The Generation R Cohort includes almost 10,000 children

at birth of which almost 4,000 children are scanned using MRI at the age of nine.

The brain images acquired for this study are high-resolution, T1-weighted structural MRI scans. The images are processed according to the standard reconstruction using the FreeSurfer version 6.0 analysis suite (Fischl, 2012). Non-brain tissue is removed; voxel intensities are corrected for B1 field inhomogeneities; voxels are segmented into white matter, gray matter, and cerebal spinal fluid; and surface-based models of white matter and gray matter are generated. Cortical surface area is estimated at each point (vertex) along the cortical ribbon. Per hemisphere there are 163,842 vertices. After excluding vertices where cortical surface area cannot be measured (e.g., corpus callosum and lateral ventricles), 149,955 vertices in the left hemisphere and 149,926 vertices in the right hemisphere remain, that is, a total of p = 299,881 vertices. Our analyses are based on cortical surface area.

Cognition is collected at the same time as the MRI scan took place and is measured with the Matrix Reasoning subtest from the Wechsler Intelligence Test for Children-V (WISC-V Wechsler, 2014). The raw score is scaled to the normative standard score with mean ten and standard deviation three (Luciana et al., 2018).

In the ABCD Study there are 10,342 children that have a qualitatively sufficient MRI scan and completed the IQ test. The MRI scans are taken on different sites and different scanners are used. One site in New York (siteID = 22) was closed shortly after it opened. The 38 children scanned at this site were excluded from the analysis. Another eighteen children were excluded, because they have a missing value for the covariates sex (4 children) and/or ethnicity (16 children). Next to sex (58% females) and ethnicity (White = 53%, Black = 14%, Hispanic = 21%, Asian = 2%, and Other = 10%), age (9.9 years $\pm$ 0.6) is included as covariate and because of the use of different scanners we have to include a dummy variable for siteID. In addition, for every twin one and for every triplet two children are randomly excluded, that is, 1,034 children in total are excluded. The dataset of the ABCD Study is split in a training set (70%, N = 6,476) and a prior to analysis unseen test set (30%, N = 2,776).

The data from the Generation R Study is solely used to test the generalizability of the model. From the 3,892 children who completed the T1-weighted MRI scan, 693 children were excluded due to poor or insufficient data quality, incidental findings, or a different T1 acquisition. Another 570 children are excluded, because they have a missing value at IQ or ethnicity. In addition, 31 children are excluded, because for every twin we randomly exclude one child. That is, our Generation R sample consists of N = 2,598 children. Contrary to the ABCD Study, all MRI images from the Generation R Study are collected on the same scanner. Cognition of the children is measured at age thirteen with the WISC-V IQ test. For consistency, we use only the Matrix Reasoning subtest. Furthermore, we correct for the following confounding factors: age, sex, and ethnicity.

4.2.1 Data Imputation

Some subjects have missing values at vertices (i.e., 0.03% of the vertices have a missing value for at least one subject), due to problematic reconstruction of the

surface mesh caused by, for example, movement during the scan. Most vertices with missing values lie close to the corpus callosum (see Figure 4.4 in Appendix 4.A), a region known to be highly correlated with cognition (Danielsen et al., 2020). Therefore, simply excluding the vertices with missing values leads to lower prediction accuracy. Another well known method to deal with missing values is predictive mean matching (Landerman et al., 1997). This method searches for similar individuals and fills the missing value with the average of these similar individuals. Within MRI data, and other imaging data, neighboring vertices are highly correlated (generally above 0.95). Given this fact, we propose to interpolate missing values using k nearest neighbors. The nearest neighbors are found by computing the Euclidean distance from each vertex to all other vertices the. Subsequently, the missing value is replaced by the average of the k nearest vertices that have a value. More information about the imputation can be found in Appendix 4.A.

4.3 Methods

Before applying the prediction models, we first need to deal with the covariates. To control for the covariates, we residualize the variable of interest, IQ, by regressing out the effect of the covariates. That is, we continue with the residuals from the following regression:

$$\boldsymbol{w} = \boldsymbol{Z}\boldsymbol{\phi} + \boldsymbol{\epsilon},$$

where $\boldsymbol{w}$ is the vector with standardized IQ scores (with mean zero and standard deviation one), $\boldsymbol{Z}$ is a matrix containing the covariates age, sex, ethnicity, and siteID, $\boldsymbol{\phi}$ is a vector with regression coefficients, and $\boldsymbol{\epsilon}$ is the disturbance factor. Note that, the continuous covariate age is standardized to z-scores (with mean zero and standard deviation one) and that the categorical covariates, sex, ethnicity, and siteID, are dummy coded. In addition, $\boldsymbol{Z}$ contains a vector of ones to capture the intercept and reference groups of the categorical variables.

The residuals from this regression can be written as $\boldsymbol{e} = \boldsymbol{M}\boldsymbol{w}$, where $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$. The residuals are the part of IQ that is uncorrelated with the covariates, that is, $\boldsymbol{M}\boldsymbol{w}$ is $\boldsymbol{w}$ after the effects of the covariates have been partialled out. For the remainder of the paper, we define the $N \times 1$ vector $\boldsymbol{y}$ as $\boldsymbol{M}\boldsymbol{w}$ standardized to z-scores (with mean zero and standard deviation one) and the $N \times p$ matrix $\boldsymbol{X}$ as the vertices standardized to z-scores (with mean zero and standard deviation one). The out-of-sample equivalent of $\boldsymbol{y}$ denoted by $\tilde{\boldsymbol{y}}$ is created using the parameters obtained with the training set. For more details see Appendix 4.B.

The following subsections elaborate the three prediction methods vertex-wise linear regressions, adaptive lasso regression, and better subset regression.

4.3.1 Vertex-wise Linear Regressions

The common vertex-wise approach performs a linear regression per vertex, that is, almost 150,000 regressions per hemisphere. A widely used software package for this analysis is `QDECR` in `R` (Lamballais and Muetzel, 2021), which runs the regressions with

the vertices as dependent variable. The reason we follow this procedure is because we use the multiple testing correction that is native to the FreeSurfer library, which presumes the vertices to be the dependent variables. Note that this is contrary to the multivariate analyzes where we include all vertices simultaneously as predictor variables. However, also note that, in this special case the weights from the regression $\boldsymbol{x}_j$ on $\boldsymbol{y}$ are the same as the weights from the regression $\boldsymbol{y}$ on $\boldsymbol{x}_j$, where $\boldsymbol{x}_j$ is the j^{th} column of matrix $\boldsymbol{X}$, indicating the j^{th} vertex. This claim is legitimate, because in our case both $\boldsymbol{x}_j$ and $\boldsymbol{y}$ are vectors standardized to z-scores (with mean zero and standard deviation one). Thus, $\hat{\beta}_{\text{OLS},j} = (\boldsymbol{x}_j'\boldsymbol{x}_j)^{-1}\boldsymbol{x}_j'\boldsymbol{y} = (\boldsymbol{y}'\boldsymbol{y})^{-1}\boldsymbol{y}'\boldsymbol{x}_j$, because $\boldsymbol{x}_j'\boldsymbol{x}_j = \boldsymbol{y}'\boldsymbol{y} = n-1$ and $\boldsymbol{x}_j'\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{x}_j$. Hence, the predictions from this analysis can be computed as follows:

$$\hat{\boldsymbol{y}}_{\text{OLS},j} = \boldsymbol{x}_j\hat{\beta}_{\text{OLS},j} \quad \forall j = 1, ..., p,$$

where $\hat{\boldsymbol{y}}_{\text{OLS},j}$ indicates the vector with predictions from the analysis of the j^{th} vertex on $\boldsymbol{y}$.

4.3.1.1 Multiple Testing Correction

Since we perform around 150,000 regressions per hemisphere, a correction for multiple testing is needed. We adjust the p-values using the standard procedure in FreeSurfer (Greve and Fischl, 2018) that uses smoothed Gaussian Monte Carlo simulations. Tables with p-values are created for several critical values (often referred to as α) and smoothness levels. The smoothness level, indicated by global FWHM (fill-width/half-max), of an analysis is based on the correlation coefficient between the residuals from the regressions of the nearest neighbors. With the pair of the selected critical value and the obtained smoothness level the corrected p-values can be found. Next to this correction, we apply a Bonferroni correction to adjust for the fact that both hemispheres are analyzed. That is, the p-values are multiplied by two.

4.3.1.2 Weighted Sum Score

The vertex-wise linear regression is a univariate approach. Only one vertex at the time is considered. It is not valid to directly compare a univariate approach with the multivariate machine learning approaches, as the latter will always outperform the first. Therefore, we create a variable that comprises all (significant) vertices. Like in a polygenic risk score analysis (Choi et al., 2020), we create a weighted prediction over all univariate regressions, that is,

$$\boldsymbol{v}_{\text{WSS}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}^*_{\text{OLS}},$$

where $\hat{\boldsymbol{\beta}}^*_{\text{OLS}}$ has element $\hat{\boldsymbol{\beta}}_{\text{OLS},j}$ if this weight is significant in the vertex-wise approach and zero otherwise. Then, this new variable $\boldsymbol{v}_{\text{WSS}}$ is regressed on $\boldsymbol{y}$ and predictions of $\boldsymbol{y}$ are obtained.

The out-of-sample predictions are based on the imaging data from the test set, denoted by $\tilde{\boldsymbol{X}}$. Note that, $\tilde{\boldsymbol{X}}$ is standardized using the mean and standard deviation obtained from the training set. The created variable, $\tilde{\boldsymbol{v}}_{\text{WSS}} = \tilde{\boldsymbol{X}}\hat{\boldsymbol{\beta}}^*_{\text{OLS}}$, is multiplied

by the weight from the in-sample regression of $\boldsymbol{v}_{\text{WSS}}$ on $\boldsymbol{y}$. The resulting out-of-sample predictions are compared to the outcome variable in the test set, $\tilde{\boldsymbol{y}}$.

4.3.1.3 Average Brain Metric Score

Next to the weighted sum score, for similar reasons, we compute the average per individual over all variables with a significant weight in the vertex-wise linear regression, that is,

$$\boldsymbol{v}_{\text{ABM}} = 1/p^* \boldsymbol{X}\boldsymbol{\iota}^*$$

where p^* is the number of significant weights in the vertex-wise linear regression and $\boldsymbol{\iota}^*$ is a $p \times 1$ vector containing a one if the corresponding weight is significant in the vertex-wise linear regression and zero otherwise. Then, this new variable $\boldsymbol{v}_{\text{ABM}}$ is regressed on $\boldsymbol{y}$ and the estimated $\hat{\boldsymbol{\beta}}_{\text{ABM}}$-coefficient from this regression is used to make the in-sample and out-of-sample predictions.

4.3.2 Adaptive Lasso Regression

Adaptive lasso regression (Zou, 2006) is a variable selection and estimation technique, comparable to original lasso regression. The advantage of adaptive lasso regression over the original one is that it has oracle properties, meaning that the estimator is consistent in both variable selection and parameter estimation as N increases and thus for sufficiently large N will find the true nonzero weights. The difference between the original lasso regression and adaptive lasso regression is that for the latter the penalty term is weighted by a weight vector $\boldsymbol{h}$. To satisfy the necessary conditions for adaptive lasso to be an oracle procedure, this weight vector should be a function of a $\sqrt{N}$-consistent estimator of $\boldsymbol{\beta}$. An example of a $\sqrt{N}$-consistent estimator, is the ordinary least-squares (OLS) estimator obtained with the vertex-wise linear regressions (Section 4.3.1). The adaptive lasso estimates $\hat{\boldsymbol{\beta}}_{\text{lasso}}$ can be found by minimizing the function

$$L(\boldsymbol{\beta}_{\text{lasso}}) = \left\| \boldsymbol{y} - \sum_{j=1}^{p} \boldsymbol{x}_j \beta_{\text{lasso},j} \right\|^2 + \lambda \sum_{j=1}^{p} \hat{h}_j |\beta_{\text{lasso},j}|, \tag{4.1}$$

where $\hat{h}_j = \frac{1}{|\hat{\beta}_j^*|}$ with $\hat{\beta}_j^* = \hat{\beta}_{\text{OLS},j}$ the OLS estimator from the univariate regression of $\boldsymbol{x}_j$ on $\boldsymbol{y}$ and λ a tuning parameter that indicates the magnitude of the penalty. The optimal tuning parameter can be found by minimizing the prediction error in a k-fold cross validation procedure. Since the function (4.1) is convex the global minimum can be efficiently found and the issue of multiple local minima is absent.

4.3.2.1 Adaptive Lasso as Variable Selection Technique

Since adaptive lasso can also be used as a variable selection technique, one could perform a multivariate regression with the selected variables. We expect the in-sample explained variance of this multivariate regression to be slightly higher than

the in-sample explained variance of the adaptive lasso, because the weights from the multivariate regression are unbiased, unlike the ones from the adaptive lasso regression.

4.3.3 Better Subset Regression

In the setting with more variables than observations ($N < p$), variable selection can be a useful tool to analyze the data. Correctly identifying the subset of important variables can increase the prediction accuracy. Under the sparsity assumption, that is, only a small number of variables contribute to the outcome variable, a subset of variables that includes the true subset yields a smaller residual sum of squares than any subset of the same size that does not include the true subset.

Minimizing the residual sum of squares and identifying the best subset of m nonzero weights in $\boldsymbol{\beta}$ can be written as minimizing

$$f(\boldsymbol{\beta}) = ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||^2 \qquad \text{subject to } ||\boldsymbol{\beta}||_0 \leq m, \tag{4.2}$$

where $||\cdot||_0$ indicates the ℓ_0 norm, that is, the number of nonzero elements in the vector. For small p, this optimization problem can be solved by an exhaustive search over all possible subsets. For larger p this is a NP-hard combinatorial optimization problem and is infeasible to solve. Xiong (2014) proposes to search for a better subset instead of the best subset.

The iterative algorithm that finds the better subset, called orthogonalizing subset screening by Xiong, is based on majorization. The objective function $f(\boldsymbol{\beta})$ in (4.2) can be approximated by a simpler majorization function $g(\boldsymbol{\beta}, \boldsymbol{\beta}_0)$, which is always larger or equal to the objective function, $f(\boldsymbol{\beta}) \leq g(\boldsymbol{\beta}, \boldsymbol{\beta}_0)$, where $\boldsymbol{\beta}_0$ is a known vector of weights, for example, the estimates obtained from a previous iteration. When rewritting $f(\boldsymbol{\beta})$, we see that the difficult part of minimizing the objective function lies in $\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$, that is,

$$f(\boldsymbol{\beta}) = ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||^2 = \boldsymbol{y}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y}.$$

We search for a majorizing function that is quadratic in $\boldsymbol{\beta}$, but has the identity matrix $\boldsymbol{I}$ as its metric instead of $\boldsymbol{X}'\boldsymbol{X}$, that is, it should be of the form $\lambda\boldsymbol{\beta}'\boldsymbol{\beta} - 2\boldsymbol{\beta}'\boldsymbol{u} + c$.

To find $\boldsymbol{u}$, let λ be such that $\boldsymbol{X}'\boldsymbol{X} - \lambda\boldsymbol{I}$ is a negative semidefinite matrix. This can be satisfied by setting λ greater or equal to the largest eigenvalue of $\boldsymbol{X}'\boldsymbol{X}$, that is $\lambda \geq \lambda_{\max}$. Then $(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'(\boldsymbol{X}'\boldsymbol{X} - \lambda\boldsymbol{I})(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \leq 0$. From this inequality it follows that

$$\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \leq \lambda\boldsymbol{\beta}'\boldsymbol{\beta} - 2\lambda\boldsymbol{\beta}'(\boldsymbol{\beta}_0 - \lambda^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}_0) + \lambda\boldsymbol{\beta}_0'(\lambda\boldsymbol{I} - \boldsymbol{X}'\boldsymbol{X})\boldsymbol{\beta}_0.$$

Next, instead of minimizing $f(\boldsymbol{\beta}) = ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||^2$ we can minimize the simpler majorization function,

$$f(\boldsymbol{\beta}) = ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||^2 \leq \lambda||\boldsymbol{\beta} - \boldsymbol{u}||^2 + c_2 = g(\boldsymbol{\beta}, \boldsymbol{\beta}_0),$$

where $\boldsymbol{u} = \boldsymbol{\beta}_0 - \lambda^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}_0 + \lambda^{-1}\boldsymbol{X}'\boldsymbol{y}$ and $c_2 = \lambda\boldsymbol{\beta}_0'(\lambda\boldsymbol{I} - \boldsymbol{X}'\boldsymbol{X})\boldsymbol{\beta}_0 + \boldsymbol{y}'\boldsymbol{y}$. The only active part that can be minimized is $||\boldsymbol{\beta} - \boldsymbol{u}||^2 = \sum_j(\beta_j - u_j)^2$ where we are allowed

to choose only m of the β_j different from zero. Without any constraint, we would choose all $\beta_j = u_j$ so that $||\boldsymbol{\beta} - \boldsymbol{u}||^2 = 0$. To minimize under the constraint that only m of the β_j can be nonzero, it is optimal to choose those m weights, β_j to be nonzero for which $|u_j|$ is the largest. As a consequence, for these β_j we choose the update $\beta_j^{k+1} = u_j$ making their contribution $(\beta_j^* - u_j)^2 = 0$ and choose the remaining $p - m$ weights, $\beta_j^{k+1} = 0$ so that their contribution is $(\beta_j^* - u_j)^2 = (0 - u_j)^2 = u_j^2$. These settings ensure $g(\boldsymbol{\beta}^{k+1}, \boldsymbol{\beta}^k)$ to be minimal. Continue to update $\boldsymbol{\beta}$ until the decrease of $f(\boldsymbol{\beta}^{k+1})$ is smaller than a small value ϵ. Due to the monotonicity property the objective function does not increase after each iteration.

A disadvantage of the above procedure is its slow convergence. To speed up the algorithm we replace the update $\boldsymbol{\beta}^{k+1} = \boldsymbol{u}$ by the least-squares estimator to avoid redundant iterations in achieving the least-squares estimator. That is,

$$\boldsymbol{\beta}_M^{k+1} = (\boldsymbol{X}'_M \boldsymbol{X}_M)^+ \boldsymbol{X}'_M \boldsymbol{y}$$

where A^+ denotes the Moore-Penrose inverse of A, $\boldsymbol{\beta}_M^{k+1}$ contains the m elements of $\boldsymbol{\beta}^{k+1}$ corresponding to the m largest absolute values of $\boldsymbol{u}$, and $\boldsymbol{X}_M$ contains the m variables corresponding to the m largest absolute values of $\boldsymbol{u}$. Note that, $\boldsymbol{X}'_M \boldsymbol{X}_M$ is not necessarily of full rank, especially when columns in $\boldsymbol{X}$ are highly correlated, therefore, the Moore-Penrose inverse is used instead of the normal inverse. It is known that the least-squares estimator yields the least sum of squares. This implies that the method still retains the monotonicity property, so an update with the least-squares estimator results in at least an equal and possibly larger decrease in residual sum of squares after each iteration.

Another increase in convergence speed can be obtained by standardizing $\boldsymbol{X}$. This transformation ensures the largest eigenvalue does not get too large, with a slow algorithm as a consequence.

For every value of m a best subset exists. To identify the best number of variables selected m, k-fold cross validation is used.

4.4 Results

This section provides specific details, such as the computational settings and method specific results, on the different methods used, including the vertex-wise linear regressions, the adaptive lasso regression, and the better subset regression. The fifth subsection is devoted to the comparison of the performance of these models and the last subsection examines the external validity of the models in the Generation R Study. Next to the commonly used vertex-wise analysis, a multivariate regression model with the ROIs as predictors is used as benchmark. Since the data set is, prior to analysis, split into a training set and test set, we provide a measure of the model performance for both sets. This measure, the explained variance ($R^2 = \mathrm{cor}^2(\boldsymbol{y}, \hat{\boldsymbol{y}})$, where $\hat{y}$ is the vector with predicted outcomes) is computed in-sample using the predicted outcome obtained with the MRI data on which the models are trained. Subsequently, the models are validated in the prior to analysis unseen test set, resulting

in the out-of-sample explained variances. Note that, the test set is standardized using the means and standard deviations obtained from the training set.

4.4.1 Results Vertex-wise Linear Regression

The vertex-wise linear regression analysis is performed with the `QDECR` package in `R` (Lamballais and Muetzel, 2021). In the left hemisphere 145,586 regressions (out of 149,955, which is $> 97\%$) have a significant β-coefficient after correcting for multiple testing (i.e., corrected p-value < 0.05) and in the right hemisphere there are 142,463 regressions (out of 149,926, which is $> 95\%$) that have a significant β-coefficient after correcting for multiple testing (i.e., corrected p-value < 0.05). The explained variance, expressed in terms of the squared correlation between the outcome and predicted outcome (i.e., $R^2 = \text{cor}^2(\boldsymbol{y}, \hat{\boldsymbol{y}})$), of each regression with a significant β-coefficient are presented in the left panel of Figure 4.3 as a histogram. The out-of-sample R^2 are given in the middle panel of Figure 4.3.

Furthermore, the R^2 of the weighted sum score (WSS) and the average brain metric score (ABM) are presented as a dashed line in the left and middle panel of Figure 4.3 in-sample and out-of-sample, respectively.

4.4.2 Results Region of Interest Analysis

A common approach to analyze the relation between brain morphology and a phenotype is to partition the brain into anatomically divided regions (ROIs) and correlate each region to the phenotype. Since we are interested in predicting the phenotype rather than describing the associations with specific regions, we perform a multivariate linear regression with all 68 ROIs as predictors in one model and the residualized IQ variable as outcome variable. The R^2 obtained with this model is presented as a dashed line in the left panel (in-sample) and middleeee panel (out-of-sample) of Figure 4.3. Note that, for one subject there is no ROI data available, therefore we excluded this participant from this analysis.

In addition, we evaluated the correlation between residualized IQ and each region separately. After a Bonferroni correction for multiple testing (i.e., $p < 0.05/68$), we find that all regions in the left hemisphere are significant and all regions except for the paracentral and temporalpole region in the right hemisphere are significant.

4.4.3 Results Adaptive Lasso Regression

The adaptive lasso regression is performed with the `glmnet` package in `R` (Friedman et al., 2021). With $k = 5$-fold cross validation, the best value for the penalty term λ is identified. The function in `glmnet` computes its own λ sequence based on the specified length of the sequence and the ratio between the minimum value of λ (i.e., $\lambda_{\text{min}} = 0.01$ by default, because the number of observations is smaller than the number of variables) and the maximum value of λ (i.e., the smallest value of λ for which all coefficients are equal to zero). The sequence of λ is computed for the full model as well as for the five folds separately. Adapting λ for each fold leads to better

convergence. The default length of the sequence λ is 100. The mean squared error is used to identify the best value for λ.

The adaptive lasso regression selects 81 vertices covering the whole brain (i.e., 47 vertices in the left hemisphere and 34 vertices in the right hemisphere). These 81 vertices are also used in a multivariate regression. The left panel (in-sample) and middle panel (out-of-sample) of Figure 4.3 display the explained variances of both models as dashed lines. As expected, the in-sample explained variance is higher for the multivariate regression analysis. However, these results most likely suffer from overfitting, because out-of-sample the Adaptive Lasso regression performs better.

4.4.4 Results Better Subset Regression

The better subset regression requires start values for $\boldsymbol{\beta}$ as input. Appendix 4.C contains a simulation study to compare different start values. The possibilities investigated are random start values drawn from a uniform distribution, a vector with zeros, and a vector with the weights from the univariate regressions. In a data set with low or medium signal to noise ratio, the better subset regression performs slightly better when the start values for $\boldsymbol{\beta}$ are set to zero or come from the univariate regressions. Therefore, for our analysis we set the start values for $\boldsymbol{\beta}$ equal to zero.

The best number of variables selected, m, is identified by minimizing the residual sum of squares in a 5-fold cross validation. The grid of m contains fourteen values, $m \in \{20,\ 40,\ 60,\ 80,\ 100,\ 200,\ 300,\ 400,\ 500,\ 600,\ 700,\ 800,\ 900,\ 1000\}$. Figure 4.1 presents the residual sum of squares per number of variables selected m for the five folds separately (dotted lines) and for the average over the five folds (solid line). In addition, Figure 4.2 zooms in on the average residual sum of squares over the five folds. The lowest residual sum of squares is found at $m = 80$, which is similar to the 81 selected variables in the adaptive lasso regression. So, the better subset regression is again performed on the full training set with $m = 80$. Subsequently, the estimated model parameters from this model are used to validate the model in the test set. The results are presented in the left panel (in-sample) and middle panel (out-of-sample) of Figure 4.3 as dashed lines.

4.4.5 Comparison

The main results of this study are captured in Figure 4.3 which gives the results of our seven methods (vertex-wise linear regression (histogram), weighted sum score, average brain metric score, multivariate regression with ROIs, adaptive lasso regression, multivariate regression with selected variables from the adaptive lasso analysis, and better subset regression). The left panel of the figure presents the in-sample explained variance (R^2). The multivariate analysis with the selected variables by the adaptive lasso regressions has the highest explained variance. Furthermore, all multivariate analysis perform better than the single vertex-wise linear regressions.

The middle panel of Figure 4.3 displays the out-of-sample R^2. From this figure, it is clear that the adaptive lasso regression performs best, indicating that adaptive lasso suffers less from overfitting compared to the multivariate regression with selected variables from the adaptive lasso regression. Except the better subset regression, all

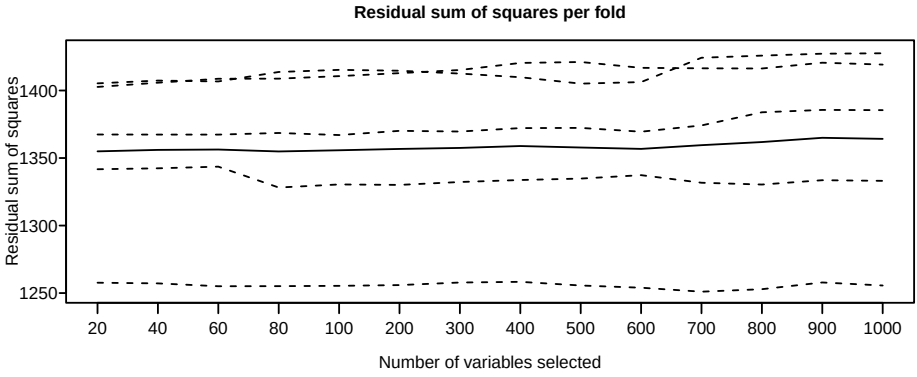


Figure 4.1: The residual sum of squares per number of variables selected m for all five folds (dotted lines) and averaged over the five folds (solid line).

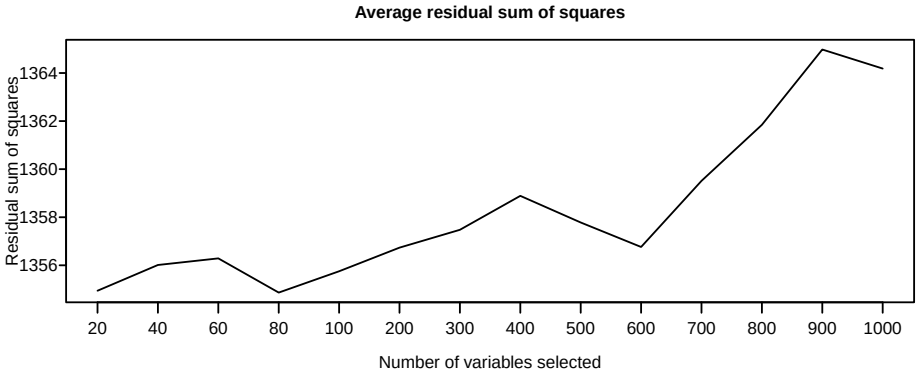


Figure 4.2: The averaged residual sum of squares per number of selected variables m. Note that, this graph zooms in on the solid line in Figure 4.1.

multivariate approaches perform better out-of-sample than the univariate vertex-wise linear regressions. Note that, the variable of interest, IQ, is residualized using the coefficients from the training set. Furthermore, the imaging data are standardized using the means and standard deviations from the training set.

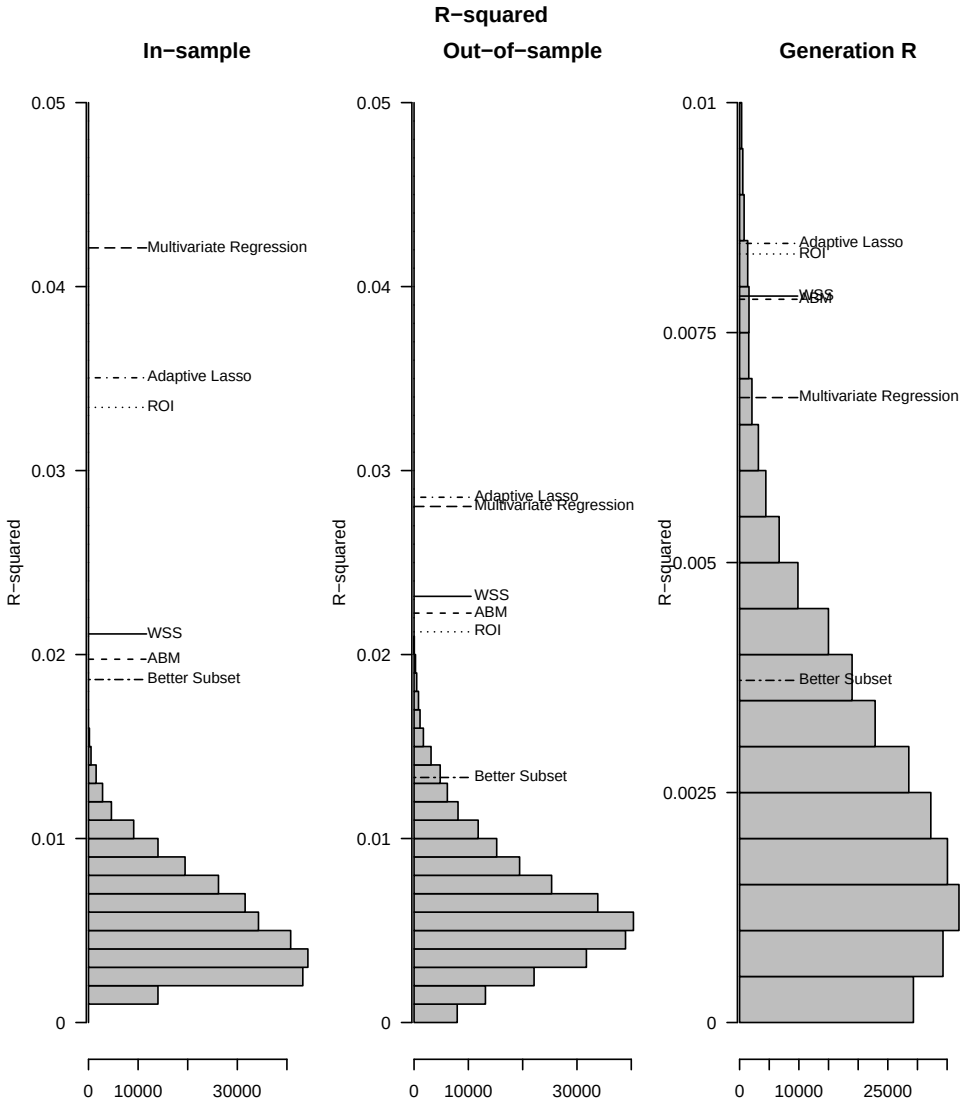


Figure 4.3: The explained variances (R^2) of the vertex-wise linear regression analysis (histogram), weighted sum score analysis (WSS), average brain metric score (ABM), multivariate regression with ROIs, adaptive lasso regression, multivariate regression with selected variables from adaptive lasso regression (multivariate regression), and better subset regression. Only the results from the significant vertex-wise linear regressions are presented in the histogram. The left panel shows the in-sample results, the middle panel the out-of-sample results, and the right panel the results obtained with the Generation R cohort. Note that the scale on the vertical axis of the right graph differs from the scale on the vertical axis of the other two graphs.

4.4.6 External Validity in Generation R Study

Next to the out-of-sample results discussed in the previous section, we also have an 'out-of-cohort' sample in which we can test the generalizability of the models. The model parameters obtained in the training set of the ABCD sample are used to compute the explained variance in the Generation R sample. Note that, the control variables, sex, age at MRI scan, age at IQ test, and ethnicity from the Generation R data set are used to residualize the variable of interest, IQ. Moreover, the standardization to z-scores of the continuous control variables (age at MRI scan and age at IQ test), the variable of interest, IQ, and the imaging data is done within the Generation R sample. The right panel of Figure 4.3 displays the explained variance of the seven models in the 'out-of-cohort' sample. The performance of the models in the current sample is rather poor and worse than the performances in the test set of the ABCD sample.

4.5 Discussion

This study compares two machine learning techniques, namely adaptive lasso regression and better subset regression, with the frequently used vertex-wise approach and a multivariate regression on ROIs in their performance to predict cognition of children through brain morphology. Brain images taken with MRI scanners in the ABCD Study are used to predict nonverbal IQ. The available N = 9,252 images are taken at 21 different sites across the United States, where different scanners are used. The siteID, age, sex, and ethnicity are used as covariates. The predictive performance of the models is expressed in the explained variance, computed as the squared correlation between the outcome vector and predicted outcome vector. For all techniques, the parameters are learned using a training set containing N = 6,476 individuals, after which they are validated in a prior to analysis unseen test set containing N = 2,776 individuals.

Figure 4.3 presents the results of our seven methods (vertex-wise linear regression (histogram), weighted sum score, average brain metric score, multivariate regression with ROIs, adaptive lasso regression, multivariate regression with selected variables from the adaptive lasso analysis, and better subset regression). The predictive performance is modest with the highest R^2 in-sample of 0.042 for the multivariate regression with selected variables from the adaptive lasso regression. Considering both the in-sample and out-of-sample performance, the adaptive lasso regression performs best. Moreover, most multivariate approaches outperform the univariate vertex-wise linear regressions, indicating that a whole brain approach is better than a vertex-wise analysis. In addition, we tested the models in a different cohort to examine the external validity. The 'out-of-cohort' results are rather poor, indicating that the results are not generalizable to different cohorts.

The modest explained variances both in-sample as well as out-of-sample could be explained by the fact that we focus on a subscale of IQ instead of Full Scale IQ. A single subtest contains more noise and is less accurate than a Full Scale IQ score. The accuracy of the predictions can likely be improved by examining the Full Scale

IQ.

Next to the subscale of IQ, the difficulty of combining two cohorts could, in particular, explain the poor performance in the Generation R cohort. In general, cohorts are inconsistent in scanner type, data quality, phenotypical measure, acquisition protocol et cetera (Bashyam et al., 2020). In addition, the underlying populations of the two cohorts might be different. For example, the ABCD data set (used to train the models) is collected in the United States and should be representative of the United States population. The Generation R data set (used to test the generalizability) is collected in Rotterdam, a city in the Netherlands and, therefore, represents the multi-ethnic population of Rotterdam. The populations in the United States and Rotterdam seem similar, at first glance. However, if we zoom in on, for example, the ethnicities, we see that Hispanics contribute for 21% and Afro-Americans for 14% to the ABCD data set. In the Generation R Study, the most common foreign ethnicities are Surinamese (7.1%), Turkish (5.9%), and Moroccan (4.3%).

Some subjects have missing values at vertices, due to problematic reconstruction of the surface mesh. We imputed these missing values by interpolation based on the values of the neighbors. The five nearest non-missing neighboring vertices are established and the average value of these vertices are used to impute the missing value. Most vertices with missing values lie close to the corpus callosum and, hence, close together, see 4.4 in Appendix 4.A. This could imply that the current vertex with missing value is surrounded by vertices with missing values and that the fifth nearest non-missing vertex lies far away. This could impact the accuracy of the imputed value, as the correlation between the vertices relates to the proximity of the five non-missing neighboring vertices. Figure 4.5 in Appendix 4.A displays a histogram with the number of vertices with missing value up to the 5^{th} non-missing nearest neighbor on a logarithmic scale. This graph shows that 59% of the vertices with missing values have their fifth non-missing neighbors within their 20 nearest neighbors and 95% missing values have their fifth non-missing neighbors within their 50 nearest neighbors. Although most non-missing neighboring vertices seem close enough, it could be worthwhile to do a sensitivity analysis without these missing vertices and compare the predictive performance of the models. Furthermore, a simulation study can investigate the imputation performance with different numbers of nearest neighbors. We believe that the imputation of (brain) imaging data is an important research field, but is out of the scope of this study.

Finally, the selected morphological brain metric in this study is cortical surface area. It is well known that other brain morphology metrics, like cortical thickness, also correlate with cognition (Narr et al., 2007). A combination of two or more metrics potentially increases the predictive performance. However, combining multiple morphological brain metrics on vertex level largely increases the number of variables, which probably leads to memory issues. Future work is required to develop methods and increase the efficiency of existing methods, that allow for analyzing high dimensional data within feasible computation time and taking into account the potential memory constraints.

Taken together, machine learning techniques are widely used in fields with an abundance of data and scarcity of theoretical starting points, like the field of neuroima-

ging. Using two machine learning techniques, namely adaptive lasso regression and better subset regression, we show that a multivariate whole brain approach performs better in predicting cognitive ability with brain morphology than the frequently used vertex-wise linear regression. Our study shows the promise of machine learning techniques as prediction methods for predicting phenotypes, like cognitive ability, with brain morphology.

Appendix

4.A Imputation Missing Values

Due to problematic reconstruction of the surface mesh caused by, for example, movement during the scan, some vertices have missing values. A total of 3,458,866 values are missing divided over 9,497 vertices. There are 39 vertices that have more than 8,500 missing values, with the largest number of missing values per vertex being 9,117. Most vertices with missing values lie close to the corpus callosum, see Figure 4.4.

Figure 4.4: Medial view of the left hemisphere (left) and right hemisphere (right) with the vertices with at least one missing value marked yellow.

We impute a missing value by taking the average over the $k = 5$ nearest neighbors that are not missing for that individual. Figure 4.5 presents the number of vertices with a missing value up to the fifth non-missing nearest neighbor on a logarithmic scale. 59% of the vertices with missing values have their fifth non-missing neighbor within their 20 nearest neighbors and 95% of the vertices with missing values have their fifth non-missing neighbors within their 50 nearest neighbors. The largest number of vertices with missing values up to the fifth non-missing nearest neighbor is 458.

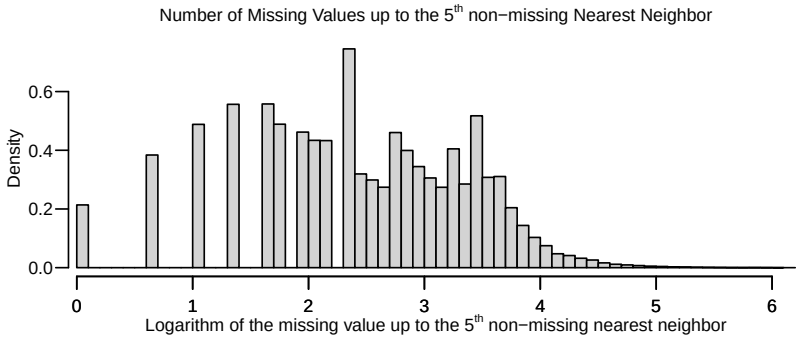


Figure 4.5: A histogram with the number of vertices with a missing value up to the 5^{th} non-missing nearest neighbor on a logarithmic scale.

4.B Out-of-Sample Testing

Similar as in the training set, we regress the effect of the covariates out of IQ and continue with the residuals. The residuals are computed as follows

$$\tilde{\boldsymbol{e}} = \tilde{\boldsymbol{w}} - \tilde{\boldsymbol{Z}}\hat{\phi},$$

where $\tilde{\boldsymbol{w}}$ is the vector containing IQ scores from the test set standardized with the mean and standard deviation obtained from the training set, $\tilde{\boldsymbol{Z}}$ is a matrix with the covariates age, sex, and siteID from the test set, and $\hat{\phi}$ is the weight obtained from the same regression in the training set. Thereafter, the residuals $\tilde{\boldsymbol{e}}$ are standardized with the mean and standard deviation from the residuals in the training set and denoted as $\tilde{\boldsymbol{y}}$. Also the imaging data in the test set is standardized with the corresponding means and standard deviations from the training set.

4.C Start Values Better Subset Regression - A Simulation Study

To evaluate how well the the better subset regression reconstructs the true non-zero weights under various conditions, we perform a simulation study. In particular, we focus on the start values for $\boldsymbol{\beta}$ and we vary the signal to noise ratio in the data set. Three different settings for the start values are analyzed, (i) all start values for $\boldsymbol{\beta}$ equal to zero, (ii) m start values for $\boldsymbol{\beta}$ equal to the coefficients of the m univariate regressions of $\boldsymbol{x}_j$ on $\boldsymbol{y}$ with the lowest sum of squared residuals (SSR) and the other start values for $\boldsymbol{\beta}$ equal to zero, and (iii) m start values for $\boldsymbol{\beta}$ drawn randomly from a uniform distribution. In addition, we apply better subset regression with these settings to different types of data sets varying the signal to noise ratio. In every iteration a data set with a high ($s = 5$), medium ($s = 1$), and low ($s = 0.1$) signal to noise ratio is generated. Furthermore, the data sets contain $N = 500$ observations and $p = 3000$ variables of which $m = 25$ variables are selected to compute the true $\boldsymbol{y}_{\text{true}}$. The elements of the predictor matrix $\boldsymbol{X}$ are standard normal distributed ($\boldsymbol{X} \sim N(\boldsymbol{0}, \boldsymbol{I})$) and thereafter standardized to z-scores and divided by $p^{0.5}$ to ensure the eigenvalues are not affected by the number of variables p. The true vector $\boldsymbol{\beta}$ contains $m = 25$ values drawn from the uniform distribution ($\boldsymbol{\beta}_{1:m} \sim \text{Unif}(-5, 5)$) and the remaining values are equal to zero. Finally, $\boldsymbol{y}$ is generated following the normal distribution with mean $\mu = \boldsymbol{X}\boldsymbol{\beta}$ and standard deviation $\sigma = (\text{var}(\boldsymbol{X}\boldsymbol{\beta})/s)^{0.5}$, with s the signal to noise ratio. The nine different settings (i.e., three different start settings times three signal to noise ratios) are evaluated based on the true variance recovered ($\text{cor}^2(\boldsymbol{y}_{\text{true}}, \hat{\boldsymbol{y}})$) and the proportion of correct non-zero weights.

Table 4.1 presents the results of the simulation study. For each setting 200 data sets are generated and the table displays the average statistics over the 200 repetitions. As excepted, the residual sum of squares (RSS), proportion of true variance recovered, and the proportion of correct non-zero weights recovered are highest for the cases with strong signal in the data ($s = 5$). Furthermore, for the data sets with a medium ($s = 1$) or low ($s = 0.1$) signal to noise ratio, we see that the better

subset regression performs slightly better when the start values for $\boldsymbol{\beta}$ are set to zero or come from the univariate regressions compared to start values randomly drawn from a uniform distribution. Therefore, for the analysis we set the start values for $\boldsymbol{\beta}$ equal to zero.

Table 4.1: The residual sum of squares (RSS), proportion of true variance recovered, and the proportion of non-zero weights recovered for all simulation settings. The standard errors are given in parentheses.

Start type	s	RSS		Proportion true variance		Proportion non-zero weights	
Zeros	0.1	259.69	(53.09)	0.09	(0.05)	0.11	(0.05)
Univariate regression	0.1	255.38	(49.64)	0.09	(0.05)	0.11	(0.05)
Random start values	0.1	299.45	(61.07)	0.06	(0.05)	0.06	(0.05)
Zeros	1	32.18	(6.30)	0.72	(0.06)	0.49	(0.07)
Univariate regression	1	33.13	(6.84)	0.71	(0.06)	0.48	(0.06)
Random start values	1	36.52	(9.05)	0.73	(0.09)	0.45	(0.08)
Zeros	5	7.83	(1.79)	0.93	(0.04)	0.64	(0.08)
Univariate regression	5	7.74	(1.69)	0.93	(0.03)	0.64	(0.08)
Random start values	5	7.68	(2.14)	0.94	(0.04)	0.68	(0.08)

$$\Pr(Y_{it}=k\wedge C_{itk}=c_{itk})=\Pr(Y_{it}=k\wedge C_{itk}=c_{itk}\mid Z_{it}=\ell)\Pr(Z_{it}=\ell)\quad h^{-1}(\eta_{it})=\log(\exp(\eta_{it})+1)\quad \eta_{its}=\alpha_s+x_{it}'\beta\quad \Pr(Y_{it}=k\wedge C_{itk}=c_{itk})=$$

$$\left(\prod_{i=1}^{N}\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{it}w_{it}}\right)\qquad \left(\prod_{i=1}^{N}\sum_{s=1}^{S}\pi_s\prod_{t=1}^{T}\Theta_{z_{it},c_{it}w_{it},s}\right)\qquad \left(\prod_{i=1}^{N}\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{it}w_{it}}\right)$$

CHAPTER 5

General Discussion

$= \Pr(Y_{tt} = k \wedge C_{tth} = c_{tth} \mid Z_{tt} = \ell) \Pr(Z_{tt} = \ell) \quad h^{-1}(\eta_{tt}) = \log(\exp(\eta_{tt}) + 1) \quad \eta_{ttg} = \alpha_g + x'_{itt}\beta \quad \Pr(Y_{tt} = k \wedge C_{tth} = c_{tth}) = \Pr(Y_{tt} =$

This thesis is about behavior and abilities and their investigation in a changing field. Data sets are rapidly growing in size, which enables the use of modern, more complex and flexible, statistical models, which most likely generates more reliable inferences and more accurate predictions. Larger samples may identify subtle effects that are not significantly detectable in smaller samples. Moreover, large sample sizes allow for a proper validation of the model by setting aside a large enough test set to independently validate the method.

The present thesis explores two examples of modern statistical methods employed in the field of behavioral and medical sciences. The thesis is split into two parts. In the first part, I develop the Censored Mixture Model (CMM) to analyze risk behavior and show two applications with the CMM. In the second part of this thesis, two machine learning approaches are compared with various simpler commonly used baseline approaches in their performance to predict cognitive ability using brain morphology.

Next, I will highlight a few considerations of the presented studies. First, I will elaborate on the considerations for the Columbia Card Task (CCT). Second, I will review the results of the neuroimaging study in Part II and give future directions. Last, I will discuss how the results of the present studies, and epidemiological studies in general, should be interpreted.

5.1 Part I: Considerations for the Columbia Card Task

In Chapter 2, we measure risk taking behavior of nine-year-old children. Risk taking behavior in children can be an early marker for risky behavior in later life, such as substance abuse. Further, as risky behavior in experimental tasks is a proxy for risk behavior in real-life (Lejuez et al., 2003; Pripfl et al., 2013), experimental tasks conducted at a young age can be used as a screening tool. A prevention scheme can be developed for children showing risky behavior in experimental tasks.

Risk taking behavior is often measured with a sequential risk task, in which participants are asked to repeat a certain action. To analyze these sequential risk tasks, the Censored Mixture Model (CMM) is developed in Chapter 2. In the present thesis, the experimental task used is the Columbia Card Task (CCT). To demonstrate that the CMM can analyze all sequential risk tasks, an application with the BART is included in Appendix 2.D of Chapter 2. The CCT is a computer-based card game where money can be won or lost by turning over cards. In the CCT, the gain amount (points earned by turning over a 'win' card), loss amount (points lost by turning over a 'loss' card), and number of loss cards are systematically varied per round. In this way the CCT allows to disentangle the effects of loss probability and both sensitivity to reward and punishment. Furthermore, a potential learning effect is minimal, because all parameters (i.e., gain amount, loss amount, and number of loss cards) are known.

Next to these strengths, the CCT also has some weaknesses. First, the CCT may be considered as an advanced risk task, since participants have to consider three

different game settings (i.e., gain amount, loss amount, and number of loss cards) during the decision-making process. The information provided may be distracting or confusing for the participants. Especially participants with lower cognitive abilities, including children who have not yet fully developed their cognitive capacity, may have a hard time understanding, processing, and applying this information. The training data used in Chapter 2 includes 69 children who turned over the 32^{nd} card (or 30^{th} in case of three loss cards), which is by definition a loss card (one child did this twice). In addition, 249 children encountered a loss card in every trial. Despite these 'questionable statistics', the CCT seems to be an adequate task to examine risk behavior of pre-adolescent children. The effects of the game parameters found in Chapter 2 are all in the expected direction, implying that, in general, the children understood the task. For the same reason, Van Duijvenvoorde et al. (2015) argues that children in the age category 8-11 years understand the CCT.

Second, there are no clear guidelines about how many trials a participant should receive. With the current settings (i.e., three game settings with each two values), the number of trials should be a multiple of eight to ensure that all combinations of game settings occur equally often. However, the question remains how many blocks of eight trials? One needs enough trials to accurately estimate the effect of the game settings, especially when one wants to include individual specific effects for the game setting parameters. However, too many trials and a fatigue effect may occur. In addition, conducting additional trials takes more time and, hence, comes with extra costs. In Chapter 3, we conduct 48 trials per version of the game (i.e., 48 trials of the Risk version and 48 trials of the Uncertainty version resulting in 96 trials in total). We perform a robustness check where we exclude participants which we believe to show signs of fatigue, because they repeatedly select the 'no card' option (thus selecting zero cards). The results in this robustness check are similar to the results obtained in the main analysis, implying that the potential fatigue effect does not drive the main results. Although fatigue may not occur in the student sample used in Chapter 3, I would argue that 48 trials for young children are too many, as they have a shorter attention span than students.

Third, three different versions of the CCT exist. The hot CCT, conducted in Chapter 2, measures affective decision making. In this version, individuals turn over cards one by one – receiving feedback after every turn – until they voluntarily stop (and convert their points into monetary reward) or until they turn over a loss card, at which point the specified loss amount is subtracted from the points earned and the trial inevitably stops. If a participant faces a loss card the game ends prematurely, the trial is censored and it is unknown how many cards a participant intended to turn over. The prevalence of censoring in the data set in Chapter 2 is 68%. Treating the censored observations as uncensored would lead to severe biases in the results. Hence, it is important to accommodate the censoring in the analysis. In the second version of the CCT, the so-called cold CCT, censoring is not an issue. Participants are asked to indicate at the start of every trial how many cards they would like to turn over. At the end of the game it is revealed whether a loss card was encountered or not. The feedback in the cold CCT comes after all trials have been played, that is, after all decisions have been made. The absence of feedback during the decision-making

process is a marked disadvantage of the cold CCT in numerous research contexts, as it removes impulsive behavior and affective engagement from the task. More recently, the two original CCT versions have been combined into a third version, the warm CCT (Huang et al., 2013), which is conducted in Chapter 3. In this version, participants select the cards they would like to turn at the start of a trial (to avoid censoring) and then press a button that prompts the selected cards to turn (to provide feedback and hence to elicit affective engagement). Feedback is provided at the end of each trial and potentially effects the behavior in the next trial. However, the excitement from turning over a win card, may lead to impulsively turning over yet another card in that trial, which would not have been selected without this feedback. This impulsive behavior within a trail is lost in the warm CCT compared to the hot CCT. Further investigation into the affective engagement elicited with the warm CCT is necessary to evaluate the warm CCT compared to the hot CCT.

Taken together, the Columbia Card Task is a valuable experimental task to measure risk taking behavior. However, one should take into account the complexity of the CCT when conducting the experiment. Moreover, one should consider the number of trials a participant receives to avoid a potential fatigue effect. Finally, multiple version of the CCT are available, which each elicits affective engagement at a different level. One should make a well-motivated decision about the conducted version.

5.2 Part II: Discussion on Neuroimaging Studies

Chapter 4 deals with predicting non-verbal IQ using brain morphology. Much work has been done in this field. However, most of this work is focused on a univariate analysis. We hypothesized that the predictive performance can be improved by using modern multivariate techniques, in particular machine learning techniques. In Chapter 4, two machine learning techniques, namely adaptive lasso and better subset regression, are compared to the univariate vertex-wise approach and a multivariate regression on aggregated regions of interest (ROIs). As expected, the multivariate approaches explain more variance in the outcome variable, non-verbal IQ, than the univariate approach. Moreover, adaptive lasso regression outperforms the multivariate regression on ROIs, indicating the promise of multivariate machine learning techniques.

Despite these interesting findings and the promise of machine learning techniques, some questions remain. To start with, it is unclear why the better subset regression does not outperform the univariate analyses and the multivariate regression on ROIs. Moreover, better subset regression seems to be sensitive to different subsamples. That is, the residual sum of squares obtained in the five folds in the cross-validation procedure vary substantially (see Figure 4.1 in Chapter 4). In addition, different variables are selected in these five folds (results not shown). The difference in selected variables is not surprising, as the variables are highly correlated. The correlated nature of the vertices may be an explanation for the low predictive accuracy of the better subset regression. better subset regression is mainly a variable selection technique and is focused on finding the important variables. In a setting where

variables are highly correlated (within brain imaging typically $\rho > 0.95$) the inverse of the sample covariance matrix (i.e., $(\boldsymbol{X}'\boldsymbol{X})^{-1}$) is not stable or does not exist. This may complicate the selection of the correct vertices (Hilafu and Yin, 2017; Zhu et al., 2021).

Furthermore, the out of sample performance in a different cohort is poor, implying that the results are not generalizable. In general, it is difficult to combine two cohorts, because they are inconsistent in scanner type, data quality, phenotypical measure, acquisition protocol et cetera (Bashyam et al., 2020). This makes it difficult to combine two cohorts in a training set or to use one of the cohorts to test the external validity. Next to these data collection inconsistencies, the underlying populations of the two cohorts might be different. For example, the ABCD data set (used to train the models) is collected in the United States and should be representative of the United States population. The Generation R data set (used to test the generalizability) is collected in Rotterdam, a city in the Netherlands and, therefore, represents the multi-ethnic population in Rotterdam. The populations in the United States and Rotterdam seem similar, at first glance. However, if we zoom in on, for example, the ethnicities, we see that Hispanics contribute for 21% and Afro-Americans for 14% to the ABCD data set. In the Generation R Study, the most common foreign ethnicities are Surinamese (7.1%), Turkish (5.9%), and Moroccan (4.3%).

In summary, multivariate techniques show great promise in neuroimaging data, particularly in high dimensional data sets. Including all information simultaneously in the model leads to better (or at least equally good) predictions. Furthermore, the predictive accuracy depends on the data availability from a large and diverse population. By combining multiple (clinical) neuroimaging cohorts, such large and diverse data sets can be achieved. However, it is clear that site, scanner, or study related differences impact the data substantially and novel ways for addressing these issues must be considered. I anticipate future work will aim at minimizing the impact of image acquisition variability, resulting in improved out of sample predictive performance and generalizability of neuroimaging studies.

5.2.1 Future Directions in Neuroimaging

Over the past decade, the use of machine learning techniques to predict phenotypes using neuroimaging data has gained popularity. Although Chapter 4 contributes to improving the explained variance, it remains rather small. Because of the novelty of this research field, many potential directions could be deployed. Here, I will discuss a few that potentially contribute to improving the predictive performance. First and foremost, we should invest in increasing sample sizes. There are still articles published with sample sizes below $N < 250$. The external validity of such studies is low and the results, typically, do not reproduce (Marek et al., 2020). One way to increase study samples is by combining multiple neuroimaging data sets. However, novel statistical methods are needed to overcome the issues induced by variations in the data acquisition process. For example, recent work by Kia et al. (2021) presents the results of combining sixteen neuroimaging data sets and applying Hierarchical Bayesian Regression.

Second, in Chapter 4, we used only cortical surface area to predict cognitive ability. Next to the association with cortical surface area, also the associations between cortical thickness and cognitive ability (Ehrlich et al., 2012; Narr et al., 2007), gyrification and cognitive ability (Gregory et al., 2016; Li et al., 2021), and subcortical volumes and cognitive ability (Cox et al., 2019) are all well established. Combining these morphological brain characteristics could increase the predictive performance. However, combining multiple morphological brain metrics on vertex level largely increases the number of variables. Currently, we are already operating at the boundaries of what is feasible with our computing facilities (i.e., Snellius supercomputer operated by Surfsara, `www.surfsara.nl`). Including cortical thickness at vertex level would simply double the number of variables. Future work is required to develop methods and increase the efficiency of existing methods that allow for analyzing high dimensional data within feasible computation time and taking into account the potential memory constraints.

Next to the integration of multiple morphological brain metrics, another future direction could involve including multiple MRI modalities. Next to cortical and subcortical structures, we could also include features of the white matter fiber tracts, collected by diffusion tensor imaging (DTI) or data about brain activation collected by (resting state) functional MRI (fMRI). Both white matter microstructures (Ritchie et al., 2015) and resting state fMRI (Saxe et al., 2018) have proven to be related to cognitive ability. Furthermore, the review by Franke and Gaser (2019) illustrates the added value of combining multiple modalities in the model in the BrainAGE context. In particular, Brown et al. (2012) demonstrates the increased predictive accuracy of age in a multisite and multimodal framework applying a regularized multivariate non-linear regression.

In Chapter 4, we apply adaptive lasso regression and better subset regression. Several other machine learning techniques have been explored, but the predictive performance remains rather small (Jollans et al., 2019; Saha et al., 2021). As discussed above, in neuroimaging the variables are typically highly correlated. Potentially, different machine learning techniques are better in incorporating the correlated nature of the data. Many more machine learning and deep learning techniques exist and could be explored in further research.

5.3 Interpretation of Presented Results

The data used in Chapters 2 and 4 are embedded in extensive studies investigating public health and biological and behavioral development over a lifespan. The aim of these chapters is to model risk behavior and cognitive ability, respectively, on an aggregated level. I want to stress that the results from these chapters are not intended to make individualized predictions or inferences about individual level behavior or abilities, as these predictions and inferences may be unreliable as discussed below.

Epidemiologists aim to identify determinants (such as risk factors and causes) of health-related states and events on a population level, for the implementation of public health programs and health service planning. When interpreting epidemiological findings it is crucial to consider that epidemiological research concentrates on pub-

lic health at group level rather than at individual level. In general, epidemiological methods perform poorly in predicting individual-level risks (Smith, 2011).

A frequently used example is the one of an old person who reaches the age of 100 even though smoking daily and drinking alcohol regularly for almost their entire life[1]. Such chance events, getting old without any serious health issues although having had an unhealthy lifestyle, are very difficult to predict and are averaged out at population level (Smith, 2011). It has been suggested that epigenetics (i.e., the analysis of genes at population level) could be a solution to obtain more accurate individualized prediction. However, family studies show that not all differences between siblings can be accounted for by genes and shared environmental factors (Lichtenstein et al., 2000). Hence, there must be other factors that could explain why some people do and others do not develop a disease.

Non-shared environmental factors, like birth order, differential parental treatment, and peer groups, have been suggested as the missing link. However, studies have shown that also these measurable non-shared environmental factors do not fully account for the unexplained variance by genes and shared environmental factors (Plomin, 2011; Plomin and Daniels, 1987; Turkheimer and Waldron, 2000).

Smith (2011) argues that the remaining factors are not 'epidemiologically tractable'. Examples of such factors are stochastic events at cellular level and modifications of genes due to environmental interactions. I would like to endorse this statement by Smith and expand it to other fields. It is an illusion to be able to collect all relevant factors and come even close to an explained variance of 100% or a prediction error of zero, which is necessary for accurate individualized predictions.

First of all, it will be extremely expensive to search for all relevant factors. Data collection in itself is already quite expensive. Every additional variable that needs to be measured comes with extra costs. Hence, additional measurements have to be selected carefully. In addition, effect sizes in epidemiology are typically rather small compared to, for instance, economics and physics. That is, risk factors contribute only little to the probability of developing a disease, indicating that many factors have a small effect. Hence, it is questionable whether it is worth the investment, both in money and time, to search for all risk factors. Moreover, if it is unknown what the relevant factors are, it is unlikely that they are all included in the data collection.

Even if we are able to measure all relevant factors, due to the statistical concept of irreducible error, some randomness and inaccuracy in the predictions remain. This concept is equivalent to the epidemiologically intractable factors discussed by Smith (2011). The irreducible error is part of the prediction error that cannot be avoided, unless the variance of the error term is equal to zero (i.e., $\sigma_\epsilon = 0$). However, this is generally an invalid assumption.

Next to the irreducible error, the prediction error consists of the squared bias and variance of the predictions. Typically, the more complex a model is, the lower the squared bias, but the higher the variance and the larger the potential overfitting problem. This bias-variance trade off can be optimized using cross validation

[1] As an example Richard Overton died at the age of 112 even though he often smoked the cigar and drank whiskey, https://www.nytimes.com/2018/12/28/obituaries/richard-overton-dead.html.

in machine learning techniques. For example, shrinkage methods, such as the adaptive lasso in Chapter 4, typically have a smaller variance than the linear regression model, however this comes with the cost of some bias, as the parameter estimates are penalized. If the decrease of variance exceeds the increase of bias, it is worthwhile (see Chapter 7 of Hastie et al., 2009, for a discussion on the bias-variance trade off). Also, the variance can be further reduced by increasing the sample size. More data increases the data to noise ratio which reduces the variance of the model and results in more accurate out of sample predictions. Taken together, this advocates for increasing the sample sizes and turning more often to machine learning approaches which can easily include all available information in one model without the problem of overfitting.

$$\Pr(Y_{it}=k\wedge C_{itk}=c_{itk})=\Pr(Y_{it}=k\wedge C_{itk}=c_{itk}\mid Z_{it}=\ell)\Pr(Z_{it}=\ell)\quad h^{-1}(\eta_{it})=\log(\exp(\eta_{it})+1)\quad \eta_{its}=\alpha_s+x_{it}'\beta\quad \Pr(Y_{it}=k\wedge C_{itk}=c_{itk})=$$

$$\left(\prod_{i=1}^{N}\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{it}w_{it}}\right)\qquad \left(\prod_{i=1}^{N}\sum_{s=1}^{S}\pi_s\prod_{t=1}^{T}\Theta_{z_{it},c_{it}w_{it},s}\right)\qquad \left(\prod_{i=1}^{N}\prod_{t=1}^{T}\Omega_{y_{it}z_{it},c_{it}w_{it}}\right)$$

References

$= \Pr(Y_{itt} = k \wedge C_{ittk} = c_{ittk} \mid Z_{itt} = \ell) \Pr(Z_{itt} = \ell)$ $h^{-1}(\eta_{itt}) = \log(\exp(\eta_{itt}) + 1)$ $\eta_{ittc} = \alpha_c + x'_{itt}\beta$ $\Pr(Y_{itt} = k \wedge C_{ittk} = c_{ittk}) = \Pr(Y_{itt} =$

References

Amlien, I.K., A.M. Fjell, C.K. Tamnes, H. Grydeland, S.K. Krogsrud, T.A. Chaplin, M.G.P. Rosa and K.B. Walhovd (2016). "Organizing principles of human cortical development—thickness and area from 4 to 30 years: insights from comparative primate neuroanatomy". *Cerebral Cortex* 26.1, pp. 257–267.

Arbabshirani, M.R., S. Plis, J. Sui and V.D. Calhoun (2017). "Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls". *Neuroimage* 145, pp. 137–165.

Astebro, T., H. Herz, R. Nanda and R.A. Weber (2014). "Seeking the roots of entrepreneurship: Insights from behavioral economics". *Journal of Economic Perspectives* 28.3, pp. 49–70.

Baird, J.C., C. Lewis and D. Romer (1970). "Relative frequencies of numerical responses in ratio estimation". *Perception & Psychophysics* 8.5, pp. 358–362.

Bashyam, V.M., J. Doshi, G. Erus, D. Srinivasan, A. Abdulkadir, M. Habes, Y. Fan, C.L. Masters, P. Maruff, C. Zhuo et al. (2020). "Medical image harmonization using deep learning based canonical mapping: Toward robust and generalizable learning in imaging". *ArXiv preprint arXiv:2010.05355*.

Batty, G.D., I.J. Deary and L.S. Gottfredson (2007). "Premorbid (early life) IQ and later mortality risk: systematic review". *Annals of Epidemiology* 17.4, pp. 278–288.

Beauchamp, J.P., D. Cesarini and M. Johannesson (2017). "The psychometric and empirical properties of measures of risk preferences". *Journal of Risk and Uncertainty* 54.3, pp. 203–237.

Bechara, A., A.R. Damasio, H. Damasio and S.W. Anderson (1994). "Insensitivity to future consequences following damage to human prefrontal cortex". *Cognition* 50.1-3, pp. 7–15.

Bellantuono, L., L. Marzano, M. La Rocca, D. Duncan, A. Lombardi, T. Maggipinto, A. Monaco, S. Tangaro, N. Amoroso and R. Bellotti (2021). "Predicting brain age with complex networks: From adolescence to adulthood". *Neuroimage* 225, p. 117458.

Bernoster, I., C.A. Rietveld, A.R. Thurik and O. Torrès (2018). "Overconfidence, optimism and entrepreneurship". *Sustainability* 10.7, p. 2233.

Bolton, D.L. (2012). "Individual entrepreneurial orientation: Further investigation of a measurement instrument". *Academy of Entrepreneurship Journal* 18.1, p. 91.

Bolton, D.L. and M.D. Lane (2012). "Individual entrepreneurial orientation: Development of a measurement instrument". *Education+ Training*.

Brand, M., E. Fujiwara, S. Borsutzky, E. Kalbe, J. Kessler and H.J. Markowitsch (2005). "Decision-making deficits of Korsakoff patients in a new gambling task with explicit rules: Associations with executive functions." *Neuropsychology* 19.3, p. 267.

Brown, T.T., J.M. Kuperman, Y. Chung, M. Erhart, C. McCabe, D.J. Hagler Jr., V.K. Venkatraman, N. Akshoomoff, D.G. Amaral, C.S. Bloss et al. (2012). "Neuroanatomical assessment of biological maturity". *Current Biology* 22.18, pp. 1693–1698.

Cantillon, R. (1931). "Essay on the nature of trade in general (H. Higgs, Trans.)" *London: UK: Frank Cass & Company, Ltd.(Original work published in 1755)*.

Choi, S.W., T.S.-H. Mak and P.F. O'Reilly (2020). "Tutorial: a guide to performing polygenic risk score analyses". *Nature Protocols* 15.9, pp. 2759–2772.

Cochran, W.G. (1952). "The chi-square goodness-of-fit test". *Annals of Mathematical Statistics* 23.3, pp. 15–345.

Cole, J.H. and K. Franke (2017). "Predicting age using neuroimaging: innovative brain ageing biomarkers". *Trends in Neurosciences* 40.12, pp. 681–690.

Collins, L.M., S. Sussman, J.M. Rauch, C.W. Dent, C.A. Johnson, W.B. Hansen and B.R. Flay (1987). "Psychosocial Predictors of Young Adolescent Cigarette Smoking: A Sixteen-Month, Three-Wave Longitudinal Study 1". *Journal of Applied Social Psychology* 17.6, pp. 554–573.

Colom, R., M. Burgaleta, F.J. Román, S. Karama, J. Álvarez-Linera, F.J. Abad, K. Martınez, M.Á. Quiroga and R.J. Haier (2013). "Neuroanatomic overlap between intelligence and cognitive factors: morphometry methods provide support for the key role of the frontal lobes". *Neuroimage* 72, pp. 143–152.

Covin, J.G., J.P.C. Rigtering, M. Hughes, S. Kraus, C.-F. Cheng and R.B. Bouncken (2020). "Individual and team entrepreneurial orientation: Scale development and configurations for success". *Journal of Business Research* 112, pp. 1–12.

Cox, S.R., S.J. Ritchie, C. Fawns-Ritchie, E.M. Tucker-Drob and I.J. Deary (2019). "Structural brain imaging correlates of general intelligence in UK Biobank". *Intelligence* 76, p. 101376.

Craig, J.B., M. Pohjola, S. Kraus and S.H. Jensen (2014). "Exploring relationships among proactiveness, risk-taking and innovation output in family and non-family firms". *Creativity and Innovation Management* 23.2, pp. 199–210.

Dai, L., V. Maksimov, B.A. Gilbert and S.A. Fernhaber (2014). "Entrepreneurial orientation and international scope: The differential roles of innovativeness, proactiveness, and risk-taking". *Journal of Business Venturing* 29.4, pp. 511–524.

Danielsen, V.M., D. Vidal-Piñeiro, A.M. Mowinckel, D. Sederevicius, A.M. Fjell, K.B. Walhovd and R. Westerhausen (2020). "Lifespan trajectories of relative corpus callosum thickness: regional differences and cognitive relevance". *Cortex* 130, pp. 127–141.

Davies, G., A. Tenesa, A. Payton, J. Yang, S.E. Harris, D. Liewald, X. Ke, S. Le Hellard, A. Christoforou, M. Luciano et al. (2011). "Genome-wide association studies establish that human intelligence is highly heritable and polygenic". *Molecular Psychiatry* 16.10, pp. 996–1005.

Davis, O.S.P., L.M. Butcher, S.J. Docherty, E.L. Meaburn, C.J.C. Curtis, M.A. Simpson, L.C. Schalkwyk and R. Plomin (2010). "A three-stage genome-wide association study of general cognitive ability: hunting the small effects". *Behavior Genetics* 40.6, pp. 759–767.

Dekkers, T.J., A. Popma, E.J.S. Sonuga-Barke, H. Oldenhof, A. Bexkens, B.R.J. Jansen and H.M. Huizenga (2020). "Risk taking by adolescents with attention-deficit/hyperactivity disorder (ADHD): a behavioral and psychophysiological investigation of peer influence". *Journal of Abnormal Child Psychology* 48.9, pp. 1129–1141.

Dijkstra, N.F.S., H. Tiemeier, B. Figner and P.J.F. Groenen (2022). "A censored mixture model for modeling risk taking". *Psychometrika*, pp. 1–27.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp and G.G. Wagner (2011). "Individual risk attitudes: Measurement, determinants, and behavioral consequences". *Journal of the European Economic Association* 9.3, pp. 522–550.

Duerden, E.G., M.M. Chakravarty, J.P. Lerch and M.J. Taylor (2020). "Sex-based differences in cortical and subcortical development in 436 individuals aged 4–54 years". *Cerebral Cortex* 30.5, pp. 2854–2866.

Ehrlich, S., S. Brauns, A. Yendiki, B.-C. Ho, V. Calhoun, S.C. Schulz, R.L. Gollub and S.R. Sponheim (2012). "Associations of cortical thickness and cognition in patients with schizophrenia and healthy controls". *Schizophrenia Bulletin* 38.5, pp. 1050–1062.

Falk, A., A. Becker, T.J. Dohmen, B. Enke, D. Huffman and U. Sunde (2015). "The nature and predictive power of preferences: Global evidence".

Fields, R.D. (2008). "White matter in learning, cognition and psychiatric disorders". *Trends in Neurosciences* 31.7, pp. 361–370.

Figner, B., R.J. Mackinlay, F. Wilkening and E.U. Weber (2009). "Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35.3, p. 709.

Figner, B. and E.U. Weber (2011). "Who takes risks when and why? Determinants of risk taking". *Current Directions in Psychological Science* 20.4, pp. 211–216.

Fischl, B. (2012). "FreeSurfer". *Neuroimage* 62.2, pp. 774–781.

Franke, K. and C. Gaser (2019). "Ten years of BrainAGE as a neuroimaging biomarker of brain aging: what insights have we gained?" *Frontiers in Neurology*, p. 789.

Friedman, J., T. Hastie, R. Tibshirani, B. Narasimhan, K. Tay, N. Simon and J. Qian (2021). "Package 'glmnet'". *CRAN R Repositary*.

Garavan, H., H. Bartsch, K. Conway, A. Decastro, R.Z. Goldstein, S. Heeringa, T. Jernigan, A. Potter, W. Thompson and D. Zahs (2018). "Recruiting the ABCD sample: Design considerations and procedures". *Developmental Cognitive Neuroscience* 32, pp. 16–22.

Gottfredson, L.S. (1997). "Why g matters: The complexity of everyday life". *Intelligence* 24.1, pp. 79–132.

Gregory, M.D., J.S. Kippenhan, D. Dickinson, J. Carrasco, V.S. Mattay, D.R. Weinberger and K.F. Berman (2016). "Regional variations in brain gyrification are as-

sociated with general cognitive ability in humans". *Current Biology* 26.10, pp. 1301–1305.

Greve, D.N. and B. Fischl (2018). "False positive rates in surface-based anatomical analysis". *Neuroimage* 171, pp. 6–14.

Hastie, T., R. Tibshirani, J.H. Friedman and J.H. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.

Hilafu, H. and X. Yin (2017). "Sufficient dimension reduction and variable selection for large-p-small-n data with highly correlated predictors". *Journal of Computational and Graphical Statistics* 26.1, pp. 26–34.

Holm, H.J., S. Opper and V. Nee (2013). "Entrepreneurs under uncertainty: An economic experiment in China". *Management Science* 59.7, pp. 1671–1687.

Holper, L. and R.O. Murphy (2014). "Hemodynamic and affective correlates assessed during performance on the Columbia Card Task (CCT)". *Brain Imaging and Behavior* 8.4, pp. 517–530.

Holzmeister, F. and M. Stefan (2021). "The risk elicitation puzzle revisited: Across-methods (in) consistency?" *Experimental Economics* 24.2, pp. 593–616.

Huang, Y., S. Wood, D. Berger and Y. Hanoch (2013). "Risky choice in younger versus older adults: affective context matters." *Judgment & Decision Making* 8.2.

Ikram, M.A., G.G.O. Brusselle, S.D. Murad, C.M. van Duijn, O.H. Franco, A. Goedegebure, C.C.W. Klaver, T.E.C. Nijsten, R.P. Peeters, B.H. Stricker et al. (2017). "The Rotterdam Study: 2018 update on objectives, design and main results". *European Journal of Epidemiology* 32.9, pp. 807–850.

Jollans, L., R. Boyle, E. Artiges, T. Banaschewski, S. Desrivières, A. Grigis, J.-L. Martinot, T. Paus, M.N. Smolka, H. Walter et al. (2019). "Quantifying performance of machine learning methods for neuroimaging data". *NeuroImage* 199, pp. 351–365.

Junger, M. and M. van Kampen (2010). "Cognitive ability and self-control in relation to dietary habits, physical activity and bodyweight in adolescents". *International Journal of Behavioral Nutrition and Physical Activity* 7.1, pp. 1–12.

Kahneman, D. and A. Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk". *Econometrica* 47.2, pp. 263–291.

Karama, S., Y. Ad-Dab'bagh, R.J. Haier, I.J. Deary, O.C. Lyttelton, C. Lepage, A.C. Evans, Brain Development Cooperative Group et al. (2009). "Erratum to "Positive association between cognitive ability and cortical thickness in a representative US sample of healthy 6 to 18 year-olds"". *Intelligence* 37.4, pp. 432–442.

Kia, S.M., H. Huijsdens, S. Rutherford, R. Dinga, T. Wolfers, M. Mennes, O.A. Andreassen, L.T. Westlye, C.F. Beckmann and A.F. Marquand (2021). "Federated multi-site normative modeling using hierarchical Bayesian regression". *BioRxiv*.

Kihlstrom, R.E. and J.-J. Laffont (1979). "A general equilibrium entrepreneurial theory of firm formation based on risk aversion". *Journal of Political Economy* 87.4, pp. 719–748.

Kirzner, I.M. (1973). *Competition and entrepreneurship*. University of Chicago Press.

Klesges, R.C., M. Debon and J.W. Ray (1995). "Are self-reports of smoking rate biased? Evidence from the Second National Health and Nutrition Examination Survey". *Journal of Clinical Epidemiology* 48.10, pp. 1225–1233.

Kluwe-Schiavon, B., B. Sanvicente-Vieira, T.W. Viola, E. Veiga, V. Bortolotto and R. Grassi-Oliveira (2015). "Assessing affective and deliberative decision-making: Adaptation of the Columbia Card Task to Brazilian Portuguese". *The Spanish Journal of Psychology* 18, E89.

Knight, F.H. (1921). "Risk, uncertainty and profit". *Houghton Mifflin Company.*

Koe, W.-L. (2016). "The relationship between Individual Entrepreneurial Orientation (IEO) and entrepreneurial intention". *Journal of Global Entrepreneurship Research* 6.1, pp. 1–11.

Kooijman, M.N., C.J. Kruithof, C.M. van Duijn, L. Duijts, O.H. Franco, M.H. van IJzendoorn, J.C. de Jongste, C.C.W. Klaver, A. van der Lugt, J.P. Mackenbach et al. (2016). "The Generation R Study: design and cohort update 2017". *European Journal of Epidemiology* 31.12, pp. 1243–1264.

Koudstaal, M., R. Sloof and M. Van Praag (2016). "Risk, uncertainty, and entrepreneurship: Evidence from a lab-in-the-field experiment". *Management Science* 62.10, pp. 2897–2915.

Kreiser, P.M. and J. Davis (2010). "Entrepreneurial orientation and firm performance: The unique impact of innovativeness, proactiveness, and risk-taking". *Journal of Small Business & Entrepreneurship* 23.1, pp. 39–51.

Lamballais, S. and R.L. Muetzel (2021). "QDECR: a flexible, extensible vertex-wise analysis framework in R". *Frontiers in Neuroinformatics* 15.

Landerman, L.R., K.C. Land and C.F. Pieper (1997). "An empirical evaluation of the predictive mean matching method for imputing missing values". *Sociological Methods & Research* 26.1, pp. 3–33.

Lechner, C. and S.V. Gudmundsson (2014). "Entrepreneurial orientation, firm strategy and small firm performance". *International Small Business Journal* 32.1, pp. 36–60.

Lejuez, C.W., W.M. Aklin, H.A. Jones, J.B. Richards, D.R. Strong, C.W. Kahler and J.P. Read (2003). "The Balloon Analogue Risk Task (BART) differentiates smokers and nonsmokers." *Experimental and Clinical Psychopharmacology* 11.1, p. 26.

Lejuez, C.W., J.P. Read, C.W. Kahler, J.B. Richards, S.E. Ramsey, G.L. Stuart, D.R. Strong and R.A. Brown (2002). "Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART)." *Journal of Experimental Psychology: Applied* 8.2, p. 75.

Li, L., Y. Zuo and Y. Chen (2021). "Relationship between local gyrification index and age, intelligence quotient, symptom severity with Autism Spectrum Disorder: A large-scale MRI study". *Journal of Clinical Neuroscience* 91, pp. 193–199.

Lichtenstein, P., N.V. Holm, P.K. Verkasalo, A. Iliadou, J. Kaprio, M. Koskenvuo, E. Pukkala, A. Skytthe and K. Hemminki (2000). "Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland". *New England Journal of Medicine* 343.2, pp. 78–85.

Linnér, R.K., P. Biroli, E. Kong, S.F.W. Meddens, R. Wedow, M.A. Fontana, M. Lebreton, S.P. Tino, A. Abdellaoui, A.R. Hammerschlag et al. (2019). "Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million

individuals identify hundreds of loci and shared genetic influences". *Nature Genetics* 51.2, pp. 245–257.

Luciana, M., J.M. Bjork, B.J. Nagel, D.M. Barch, R. Gonzalez, S.J. Nixon and M.T. Banich (2018). "Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery". *Developmental Cognitive Neuroscience* 32, pp. 67–79.

Lumpkin, G.T., C.C. Cogliser and D.R. Schneider (2009). "Understanding and measuring autonomy: An entrepreneurial orientation perspective". *Entrepreneurship Theory and Practice* 33.1, pp. 47–69.

Lumpkin, G.T. and G.G. Dess (1996). "Clarifying the entrepreneurial orientation construct and linking it to performance". *Academy of Management Review* 21.1, pp. 135–172.

Marek, S., B. Tervo-Clemmens, F.J. Calabro, D.F. Montez, B.P. Kay, A.S. Hatoum, M.R. Donohue, W. Foran, R.L. Miller, E. Feczko et al. (2020). "Towards reproducible brain-wide association studies". *BioRxiv:2020.08.21.257758*.

Menary, K., P.F. Collins, J.N. Porter, R. Muetzel, E.A. Olson, V. Kumar, M. Steinbach, K.O. Lim and M. Luciana (2013). "Associations between cortical thickness and general intelligence in children, adolescents and young adults". *Intelligence* 41.5, pp. 597–606.

Messina, D., P. Borrelli, P. Russo, M. Salvatore and M. Aiello (2021). "Voxel-wise feature selection method for CNN binary classification of neuroimaging data". *Frontiers in Neuroscience* 15, p. 284.

Narr, K.L., R.P. Woods, P.M. Thompson, P. Szeszko, D. Robinson, T. Dimtcheva, M. Gurbani, A.W. Toga and R.M. Bilder (2007). "Relationships between IQ and regional cortical gray matter thickness in healthy adults". *Cerebral Cortex* 17.9, pp. 2163–2171.

Nenadić, I., M. Dietzek, K. Langbein, H. Sauer and C. Gaser (2017). "BrainAGE score indicates accelerated brain aging in schizophrenia, but not bipolar disorder". *Psychiatry Research: Neuroimaging* 266, pp. 86–89.

Nicolaou, N., S. Shane, L. Cherkas, J. Hunkin and T.D. Spector (2008). "Is the tendency to engage in entrepreneurship genetic?" *Management Science* 54.1, pp. 167–179.

Okbay, A., J.P. Beauchamp, M.A. Fontana, J.J. Lee, T.H. Pers, C.A. Rietveld, P. Turley, G.-B. Chen, V. Emilsson, S.F.W. Meddens et al. (2016). "Genome-wide association study identifies 74 loci associated with educational attainment". *Nature* 533.7604, pp. 539–542.

Pallant, J. (2007). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. McGraw Hill.

Pedroni, A., R. Frey, A. Bruhin, G. Dutilh, R. Hertwig and J. Rieskamp (2017). "The risk elicitation puzzle". *Nature Human Behaviour* 1.11, p. 803.

Pedroni, A., J. Rieskamp, T. Pachur, R. Frey, J.E. Westfall and B. Figner (2018). "A prospect theory account of the "hot" Columbia Card Task".

Peng, H., W. Gong, C.F. Beckmann, A. Vedaldi and S.M. Smith (2021). "Accurate brain age prediction with lightweight deep neural networks". *Medical Image Analysis* 68, p. 101871.

Penolazzi, B., P. Gremigni and P.M. Russo (2012). "Impulsivity and reward sensitivity differentially influence affective and deliberative risky decision making". *Personality and Individual Differences* 53.5, pp. 655–659.

Pleskac, T.J. (2008). "Decision making and learning while taking sequential risks." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34.1, p. 167.

Pleskac, T.J., T.S. Wallsten, P. Wang and C.W. Lejuez (2008). "Development of an automatic response mode to improve the clinical utility of sequential risk-taking tasks." *Experimental and Clinical Psychopharmacology* 16.6, p. 555.

Plomin, R. (2011). "Commentary: Why are children in the same family so different? Non-shared environment three decades later". *International Journal of Epidemiology* 40.3, pp. 582–592.

Plomin, R. and D. Daniels (1987). "Why are children in the same family so different from one another?" *Behavioral and Brain Sciences* 10.1, pp. 1–16.

Plomin, R. and J.C. DeFries (1998). "The genetics of cognitive abilities and disabilities". *Scientific American* 278.5, pp. 62–69.

Plomin, R., C.M.A. Haworth, E.L. Meaburn, T.S. Price, Wellcome Trust Case Control Consortium 2 and O.S.P. Davis (2013). "Common DNA markers can account for more than half of the genetic influence on cognitive abilities". *Psychological Science* 24.4, pp. 562–568.

Pripfl, J., R. Neumann, U. Köhler and C. Lamm (2013). "Effects of transcranial direct current stimulation on risky decision making are mediated by 'hot'and 'cold'decisions, personality, and hemisphere". *European Journal of Neuroscience* 38.12, pp. 3778–3785.

Psychology Software Tools (2018). "E-Prime". Version (2.0.10.356) [Computer software].

Rakic, P. (1988). "Specification of cerebral cortical areas". *Science* 241.4862, pp. 170–176.

Ranganath, R., D. Tran and D. Blei (2016). "Hierarchical variational models". *International Conference on Machine Learning*, pp. 324–333.

Rauch, A. and M. Frese (2007). "Let's put the person back into entrepreneurship research: A meta-analysis on the relationship between business owners' personality traits, business creation, and success". *European Journal of Work and Organizational Psychology* 16.4, pp. 353–385.

Rauch, A., J. Wiklund, G.T. Lumpkin and M. Frese (2009). "Entrepreneurial orientation and business performance: An assessment of past research and suggestions for the future". *Entrepreneurship Theory and Practice* 33.3, pp. 761–787.

Reiss, A.L., M.T. Abrams, H.S. Singer, J.L. Ross and M.B. Denckla (1996). "Brain development, gender and IQ in children: a volumetric imaging study". *Brain* 119.5, pp. 1763–1774.

Rietveld, C.A., D. Conley, N. Eriksson, T. Esko, S.E. Medland, A.A.E. Vinkhuyzen, J. Yang, J.D. Boardman, C.F. Chabris, C.T. Dawes et al. (2014). "Replicability and robustness of genome-wide-association studies for behavioral traits". *Psychological Science* 25.11, pp. 1975–1986.

Ritchie, S.J., M.E. Bastin, E.M. Tucker-Drob, S.M. Maniega, L.E. Engelhardt, S.R. Cox, N.A. Royle, A.J. Gow, J. Corley, A. Pattie et al. (2015). "Coupled changes in brain white matter microstructure and fluid intelligence in later life". *Journal of Neuroscience* 35.22, pp. 8672–8682.

Rogers, R.D., A.M. Owen, H.C. Middleton, E.J. Williams, J.D. Pickard, B.J. Sahakian and T.W. Robbins (1999). "Choosing between small, likely rewards and large, unlikely rewards activates inferior and orbital prefrontal cortex". *The Journal of Neuroscience* 19.20, pp. 9029–9038.

Saha, S., A. Pagnozzi, D. Bradford and J. Fripp (2021). "Predicting fluid intelligence in adolescence from structural MRI with deep learning methods". *Intelligence* 88, p. 101568.

Sahoo, S. and R.K. Panda (2019). "Exploring entrepreneurial orientation and intentions among technical university students: Role of contextual antecedents". *Education+ Training*.

Saxe, G.N., D. Calderone and L.J. Morales (2018). "Brain entropy and human intelligence: A resting-state fMRI study". *PloS One* 13.2, e0191582.

Say, J.-B. (1803). "A treatise on political economy or the production, distribution, and consumption of wealth".

Schnack, H.G., N.E.M. Van Haren, R.M. Brouwer, A. Evans, S. Durston, D.I. Boomsma, R.S. Kahn and H.E. Hulshoff Pol (2015). "Changes in thickness and surface area of the human cortex and their relationship with intelligence". *Cerebral Cortex* 25.6, pp. 1608–1617.

Schonberg, T., C.R. Fox and R.A. Poldrack (2011). "Mind the gap: Bridging economic and naturalistic risk-taking with cognitive neuroscience". *Trends in Cognitive Sciences* 15.1, pp. 11–19.

Seong, S.-B., C. Pae and H.-J. Park (2018). "Geometric convolutional neural network for analyzing surface-based neuroimaging data". *Frontiers in Neuroinformatics* 12, p. 42.

Slovic, P. (1966). "Risk-taking in children: Age and sex differences". *Child Development*, pp. 169–176.

Smith, G.D. (2011). "Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice". *International Journal of Epidemiology* 40.3, pp. 537–562.

Smith, S.M. and T.E. Nichols (2018). "Statistical challenges in "big data" human neuroimaging". *Neuron* 97.2, pp. 263–268.

Sniekers, S., S. Stringer, K. Watanabe, P.R. Jansen, J.R. Coleman, E. Krapohl, E. Taskesen, A.R. Hammerschlag, A. Okbay, D. Zabaneh et al. (2017). "Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence". *Nature Genetics* 49.7, pp. 1107–1112.

Sonnega, A., J.D. Faul, M.B. Ofstedal, K.M. Langa, J.W. Phillips and D.R. Weir (2014). "Cohort profile: the health and retirement study (HRS)". *International Journal of Epidemiology* 43.2, pp. 576–585.

Storsve, A.B., A.M. Fjell, C.K. Tamnes, L.T. Westlye, K. Overbye, H.W. Aasland and K.B. Walhovd (2014). "Differential longitudinal changes in cortical thickness,

surface area and volume across the adult life span: regions of accelerating and decelerating change". *Journal of Neuroscience* 34.25, pp. 8488–8498.

Sturmfels, P., S. Rutherford, M. Angstadt, M. Peterson, C. Sripada and J. Wiens (2018). "A domain guided cnn architecture for predicting age from structural brain images". *Machine Learning for Healthcare Conference*. PMLR, pp. 295–311.

Sudlow, C., J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray et al. (2015). "UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age". *PLoS Medicine* 12.3, e1001779.

Thaler, R.H. (2016). "Behavioral economics: Past, present, and future". *American Economic Review* 106.7, pp. 1577–1600.

Turkheimer, E. and M. Waldron (2000). "Nonshared environment: a theoretical, methodological, and quantitative review." *Psychological Bulletin* 126.1, p. 78.

Tversky, A. and D. Kahneman (1992). "Advances in prospect theory: Cumulative representation of uncertainty". *Journal of Risk and Uncertainty* 5.4, pp. 297–323.

Van der Loos, M.J.H.M., C.A. Rietveld, N. Eklund, P.D. Koellinger, F. Rivadeneira, G.R. Abecasis, G.A. Ankra-Badu, S.E. Baumeister, D.J. Benjamin, R. Biffar et al. (2013). "The molecular genetic architecture of self-employment". *PloS One* 8.4, e60542.

Van Duijvenvoorde, A.C.K., H.M. Huizenga, L.H. Somerville, M.R. Delgado, A. Powers, W.D. Weeda, B.J. Casey, E.U. Weber and B. Figner (2015). "Neural correlates of expected risks and returns in risky choice across development". *Journal of Neuroscience* 35.4, pp. 1549–1560.

Van Praag, C.M. (1999). "Some classic views on entrepreneurship". *De Economist* 147.3, pp. 311–335.

Vladasel, T., M.J. Lindquist, J. Sol and M. Van Praag (2021). "On the origins of entrepreneurship: Evidence from sibling correlations". *Journal of Business Venturing* 36.5, p. 106017.

Von Neumann, J. and O. Morgenstern (1953). *Theory of games and economic behavior*. Princeton University Press.

Vuoksimaa, E., M.S. Panizzon, C.-H. Chen, M. Fiecas, L.T. Eyler, C. Fennema-Notestine, D.J. Hagler Jr., C.E. Franz, A.J. Jak, M.J. Lyons et al. (2016). "Is bigger always better? The importance of cortical configuration with respect to cognitive ability". *Neuroimage* 129, pp. 356–366.

Wakker, P.P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge University Press.

Wallsten, T.S., T.J. Pleskac and C.W. Lejuez (2005). "Modeling behavior in a clinically diagnostic sequential risk-taking task." *Psychological Review* 112.4, p. 862.

Wechsler, D. (1999). "Manual for the Wechsler abbreviated intelligence scale (WASI)". *San Antonio, TX: The Psychological Corporation*.

Wechsler, D. (2014). "Wechsler intelligence scale for children–Fifth Edition (WISC-V)". *Bloomington, MN: Pearson*.

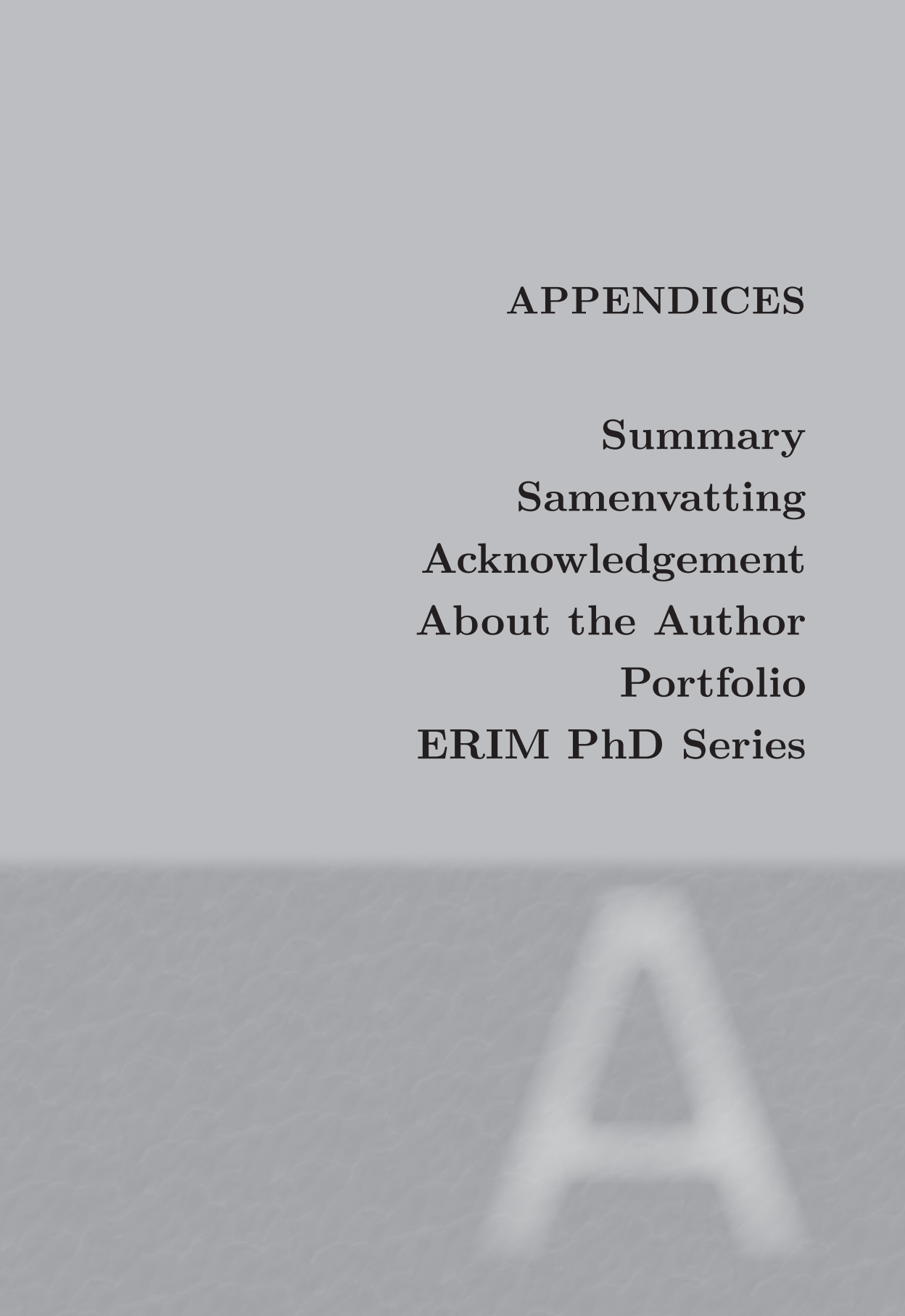
Weller, J.A., M.L. King, B. Figner and N.L. Denburg (2019). "Information use in risky decision making: Do age differences depend on affective context?" *Psychology and Aging* 34.7, p. 1005.

Xiong, S. (2014). "Better subset regression". *Biometrika* 101.1, pp. 71–84.

Zhu, W., C. Lévy-Leduc and N. Ternès (2021). "A variable selection approach for highly correlated predictors in high-dimensional genomic data". *Bioinformatics* 37.16, pp. 2238–2244.

Zou, H. (2006). "The adaptive lasso and its oracle properties". *Journal of the American Statistical Association* 101.476, pp. 1418–1429.

APPENDICES



Summary

Behavior and abilities are arguably the most fundamental concepts in life. Hence, these concepts and their causes and consequences, are studied in many disciplines, such as economics, psychology, sociology, and health sciences. Behavior is observable through someone's actions. Underlying a person's behavior are their beliefs and attitudes towards someone or something. Abilities are the personal qualities and skills needed to do something, in as far as they are innate traits and are not acquired through formal education.

The research of behaviors and abilities is quickly changing. In the last decade, more and more emphasis is put on data collection and the benefits of large data sets have become more evident. Hence, like in many other fields, the data sets are rapidly growing in size, both in terms of observations and variables. The increase in the sample size of studies enables the use of modern – often more complex and flexible – statistical methods. Furthermore, larger samples make it possible to identify subtle effects that are not detectable in smaller samples. In addition, large sample sizes allow for a proper validation of the model by setting aside a large enough test set to independently validate the method. This thesis employs modern statistical models in the field of fundamental economic behavior and medical sciences with the aim to generate more reliable inferences and more accurate predictions than would have been the case with established methods.

The dissertation is split into two parts. In Part I of this thesis, I develop the Censored Mixture Model (CMM) to analyze risk behavior and show two applications. The CMM is particularly developed to model risk behavior measured with a sequential risk task in which participants are asked to repeat a certain action, like inflating a balloon in the Balloon Analogue Risk Task (BART) or turning over cards in the Columbia Card Task (CCT). Although the CMM is applicable to all sequential risk tasks, in this dissertation the CCT is used to measure risk taking. The CCT is a computer-based card game in which participants are asked to turn over cards that are positioned face down. There are 32 cards divided in win cards and loss cards. By turning over a win card the participant earns points and by turning over a loss card they lose points and the current game round ends. At every step the participant has the choice between turning over another card and pressing the stop button to voluntarily stop this game round. The points earned by turning over a win card, the points lost by turning over a loss card, and the number of loss cards vary per round and are known to the participant. The number of cards turned over is considered a measure

for risk taking behavior. After playing all rounds, the participant is rewarded based on the overall performance on the CCT.

Modeling sequential risk tasks, such as the CCT, comes with a few challenges. The CMM is developed to solve three of these potential problems. *First*, by definition, most sequential risk tasks may randomly end prematurely, such as when the balloon pops in the BART or when a loss card is encountered in the CCT, leading to censored observations. Typically, the researcher is interested in the level of risk a participant is willing to take and the censoring obscures this. Treating the censored observations as uncensored would lead to biases in the results. The CMM accommodates for these censored observations by using the information that the participant intended to take more risk than the observed risk level. *Second*, the analyses are often based on the assumption of a smooth (normal) distribution of the residuals. However, certain outcomes are more attractive to participants than others. For example, participants tend to create geometric patterns when turning over cards in the CCT (for example, a complete row or column). These patterns lead to inflated values in the outcome distribution. By assigning additional probability mass to these outcomes in the likelihood function the CMM corrects for this behavior. *Third*, a priori unknown groups of participants can react differently to certain risk-levels. Similarly, individuals may have different tendencies towards rewards and/or losses. The CMM can model these unobserved individual tendencies by including segment specific intercepts and segment specific effects for covariates. *Finally*, for ease of interpretation, a link function is chosen such that the regression coefficients have a linear interpretation on the interval zero to infinity.

Chapter 2 describes the construction of the CMM and discusses all features of the model. In addition, the chapter includes an application with the CCT. The data in this chapter were collected as part of the Generation R Study. The CCT is one of the observational assessments that was conducted on nine-year-old children. A total of 3404 children participated and they played each sixteen rounds. Some interesting findings are that girls took more risk than boys and that children were mostly influenced by the loss probability (indicated by the number of loss cards) compared to the possible gains and losses when turning over cards. In addition, we found that children with risk averse or risk seeking behavior have higher levels of behavioral problems than children with intermediate risk behavior.

Chapter 3 contains an application of the CMM showing the relationship between individual entrepreneurial orientation (IEO) and both decision making under risk and uncertainty in a student sample. With a minor adjustment of the CCT, that is, the number of loss cards in a trial is unknown to the participant, decision making under risk as well as uncertainty can be elicited. We find robust evidence that IEO is negatively related to decision making under risk and uncertainty. In addition, we show that distinguishing the subscales of IEO is key to understanding this relationship. Individuals who indicate to be more risk-seeking indeed take more risk in the behavioral task, whereas those who indicate to be innovative or proactive exhibit more risk averse behavior. The results for the risk and uncertainty version are similar, however, we find smaller effect sizes of IEO and its subscales in the uncertainty version.

In Part II of this dissertation, I aim at predicting cognitive ability using certain structures in the brain. Cognition is one of the most important functions of the brain and many studies have associated intellectual ability to neuroimaging features, such as brain structures. Understanding the link between features of brain structure and neuropsychological tests at population level can provide insights in typical brain development. Once the baseline for typical brain development is established, we can examine atypical brain development related to cognitive deficit.

For Chapter 4, I had access to an exceptionally large pre-adolescent data set. The brain structures of more than 9000 nine-year-old children were acquired through magnetic resonance imaging (MRI) scans. In addition, a non-verbal IQ test was conducted on these children. I hypothesized that the predictive performance can be improved by using modern multivariate techniques, in particular machine learning techniques. In this chapter, two machine learning techniques, namely adaptive lasso and better subset regression, are compared to the univariate vertex-wise approach and a multivariate regression on aggregated regions of interest (ROIs). As expected, we find that the multivariate approaches explain more variance in the outcome variable, non-verbal IQ, than the univariate approach. Moreover, adaptive lasso regression outperforms the multivariate regression on ROIs, indicating the promise of multivariate machine learning techniques.

In summary, this dissertation is about behavior and abilities and their investigation in a changing field. Data sets are rapidly growing in size, which enables the use of modern, more complex and flexible, statistical models. I explore two examples of modern statistical methods employed in the field of behavioral and medical sciences. In the first part, I develop the Censored Mixture Model (CMM) to analyze risk behavior and show two applications with the CMM. In the second part, I show the promise of multivariate machine learning techniques as prediction methods for predicting cognitive ability using brain structures. Because of the increased sample sizes and the use of advanced statistical methods more reliable inferences and more accurate predictions can be generated.

Samenvatting

Gedrag en capaciteiten zijn misschien wel de meest bepalende kenmerken van een persoon. Daarom worden deze concepten en hun oorzaken en gevolgen bestudeerd in vele vakgebieden, zoals economie, psychologie, sociologie en gezondheidswetenschappen. Gedrag is waarneembaar door iemands acties. Aan het gedrag van een persoon ligt zijn/haar overtuiging en houding ten opzichte van iets of iemand ten grondslag. Capaciteiten zijn iemands persoonlijke kwaliteiten en vaardigheden die nodig zijn om iets te bereiken. Capaciteiten zijn aangeboren eigenschappen en worden niet verworven door middel van training.

Het onderzoek naar gedrag en capaciteiten verandert snel. In de laatste tien jaar is steeds meer nadruk gelegd op dataverzameling en zijn de voordelen van grote datasets duidelijker geworden. Als gevolg hiervan groeien de datasets, net als in veel andere onderzoeksvelden, snel in omvang, zowel wat betreft het aantal observaties als het aantal variabelen. De toename van de steekproefgrootte in studies maakt het gebruik van moderne – vaak complexere en flexibelere – statistische methoden mogelijk. Bovendien maken grotere steekproeven het mogelijk om subtiele effecten te identificeren die niet detecteerbaar zijn in kleinere steekproeven. Daarnaast maken grote steekproeven het mogelijk een voldoende grote testset opzij te zetten om daarmee het model onafhankelijk te valideren. Dit proefschrift maakt gebruik van moderne statistische modellen op het gebied van fundamenteel economisch gedrag en gezondheidswetenschappen om zo betrouwbaardere conclusies te trekken en nauwkeurigere voorspellingen te maken dan het geval zou zijn met gevestigde methoden.

Het proefschrift is opgesplitst in twee delen. In deel I van dit proefschrift ontwikkel ik het Censored Mixture Model (CMM) om risicogedrag te analyseren en laat ik twee toepassingen zien. Het CMM is in het bijzonder ontwikkeld voor het modelleren van sequentiële risicotaken waarbij deelnemers wordt gevraagd een bepaalde actie te herhalen, zoals het opblazen van een ballon in de Balloon Analogue Risk Task (BART) of het omdraaien van kaarten in de Columbia Card Task (CCT). Hoewel het CMM toepasbaar is op alle sequentiële risicotaken wordt in dit proefschrift de CCT gebruikt om risicogedrag te meten. De CCT is een digitaal kaartspel waarbij deelnemers wordt gevraagd kaarten om te draaien die met de afbeelding naar beneden liggen. Er zijn 32 kaarten, verdeeld in winkaarten en verlieskaarten. Door een winkaart om te draaien verdient de deelnemer punten en door een verlieskaart om te draaien verliest hij/zij punten en eindigt de ronde. Op ieder moment gedurende het spel heeft de deelnemer de keuze tussen het omdraaien van een volgende kaart, met de

kans dat dit een verlieskaart is, en het vrijwillig stoppen van de huidige speelronde door op de stopknop te klikken. De punten die verdiend kunnen worden met het omdraaien van een winkaart, de punten die verloren gaan door het omdraaien van een verlieskaart en het aantal verlieskaarten variëren per ronde en zijn bekend bij de deelnemer. Het aantal omgedraaide kaarten wordt beschouwd als een maatstaf voor risicogedrag. Na het spelen van alle rondes wordt de deelnemer beloond op basis van de algehele prestatie op de CCT.

Het modelleren van sequentiële risicotaken zoals de CCT brengt enkele uitdagingen met zich mee. Het CMM is ontwikkeld om drie van deze uitdagingen het hoofd te bieden. Een eerste probleem dat zich mogelijk voordoet in sequentiële risicotaken is dat wanneer een ronde voortijdig eindigt de observatie voor die ronde gecensureerd is. Dit gebeurt bijvoorbeeld wanneer de ballon in de BART knapt of wanneer een verlieskaart wordt omgedraaid in de CCT. Doorgaans is de onderzoeker geïnteresseerd in het risico dat een deelnemer bereid is te nemen. De gecensureerde observaties verhinderen dit. Het analyseren van de gecensureerde observaties alsof ze ongecensureerd zijn, kan leiden tot vertekende resultaten. Het CMM corrigeert deze gecensureerde waarnemingen door gebruik te maken van het feit dat de deelnemer van plan was meer risico te nemen dan het waargenomen risiconiveau. Een tweede probleem is dat bestaande analyses vaak aannemen dat de residuen van de scores een effen (normale) verdeling hebben. Echter, in de praktijk zijn bepaalde uitkomsten aantrekkelijker voor deelnemers dan andere. Deelnemers hebben bijvoorbeeld de neiging om geometrische patronen te creëren bij het omdraaien van kaarten in de CCT (bijvoorbeeld een volledige rij of kolom). Deze patronen leiden tot pieken in de uitkomstverdeling. Door extra kansmassa toe te kennen aan deze uitkomsten in de likelihoodfunctie corrigeert het CMM voor dit gedrag. Ten derde, a priori onbekende groepen deelnemers kunnen verschillend reageren op bepaalde risiconiveaus. Evenzo kunnen personen verschillend reageren op eventuele beloningen en/of verliezen. Het CMM kan deze niet-waargenomen individuele voorkeuren modelleren door middel van segment specifieke intercepten en segment specifieke effecten voor covariaten. Tot slot, ter vereenvoudiging van de interpretatie, is een linkfunctie gekozen zodat de regressiecoëfficiënten een lineaire interpretatie hebben op het interval nul tot oneindig.

In hoofdstuk 2 wordt de constructie van het CMM beschreven en worden de kenmerken van het model toegelicht. Daarnaast bevat het hoofdstuk een toepassing met de CCT. De data in dit hoofdstuk zijn verzameld als onderdeel van de Generation R Studie. De CCT is een van de onderzoeken die is uitgevoerd bij negenjarige kinderen. In totaal deden 3404 kinderen mee en speelden ze ieder zestien rondes. Enkele interessante bevindingen zijn dat meisjes meer risico namen dan jongens en dat kinderen meer werden beïnvloed door de kans op verlies (aangeduid met het aantal verlieskaarten) dan door de mogelijke winsten en verliezen bij het omdraaien van kaarten. Daarnaast vonden we dat kinderen met risicomijdend of risicozoekend gedrag meer gedragsproblemen hebben dan kinderen die gemiddeld risicogedrag vertonen.

Hoofdstuk 3 bevat een toepassing van de CMM die de relatie laat zien tussen individuele ondernemersoriëntatie (IEO) en zowel besluitvorming in een risicovolle situatie als in een onzekere situatie in een steekproef onder studenten. Met een

kleine aanpassing van de CCT, namelijk het verhullen van het aantal verlieskaarten in een ronde, kan zowel besluitvorming in een risicovolle situatie als in een onzekere situatie worden gemeten. We vinden robuust bewijs dat IEO negatief gecorreleerd is met besluitvorming in een risicovolle en in een onzekere situatie. Bovendien laat het onderzoek zien dat het essentieel is om onderscheid te maken in de subschalen van IEO om zo deze relatie beter te begrijpen. Individuen die aangeven meer risicozoekend te zijn nemen inderdaad meer risico in de CCT, terwijl degenen die aangeven innovatief of proactief te zijn juist meer risicomijdend gedrag vertonen. De resultaten voor beide varianten van de CCT (dat is de variant waarbij het risico bekend is en de variant waarbij het risico onzeker is) zijn vergelijkbaar, maar we vinden kleinere effecten van IEO en de subschalen in de variant waarbij het risico onzeker is.

In deel II van dit proefschrift richt ik mij op het voorspellen van cognitie met behulp van bepaalde hersenstructuren. Cognitie is een van de belangrijkste functies van de hersenen en veel studies hebben relaties aangetoond tussen intellect en neurologische kenmerken, zoals hersenstructuren. De relatie tussen hersenstructuren en neuropsychologische testen, zoals een IQ test, op populatieniveau kan inzicht verschaffen in de standaard hersenontwikkeling. Wanneer die standaard is vastgesteld, kunnen we atypische hersenontwikkeling die verband houdt met een verminderd cognitief vermogen onderzoeken.

Voor Hoofdstuk 4 had ik de beschikking over een uitzonderlijk grote dataset. De hersenstructuren van meer dan 9000 negenjarige kinderen zijn vastgelegd middels MRI (magnetic resonance imaging) scans. Daarnaast is bij deze kinderen een non-verbale IQ-test afgenomen. Mijn hypothese was dat de voorspellingen verbeterd konden worden door het gebruik van moderne multivariate technieken, in het bijzonder machine learning technieken. In dit hoofdstuk worden twee machine learning technieken, namelijk adaptive lasso en better subset regression, vergeleken met de univariate vertex-wise methode en een multivariate regressie op geaggregeerde regions of interest (ROIs). Zoals verwacht vinden we dat de multivariate methoden meer variantie verklaren in de uitkomstvariabele, non-verbaal IQ, dan de univariate methode. Bovendien presteert adaptive lasso beter dan de multivariate regressie op ROIs, wat de toegevoegde waarde van multivariate machine learning technieken benadrukt.

In het kort gaat dit proefschrift over gedrag en capaciteiten en de onderzoeken hiernaar in dit snel veranderende veld. Datasets groeien snel in omvang, wat het gebruik van moderne, complexere en flexibelere statistische modellen mogelijk maakt. Ik bekijk twee voorbeelden van moderne statistische methoden die worden gebruikt op het gebied van gedrags- en gezondheidswetenschappen. In het eerste deel ontwikkel ik het Censored Mixture Model (CMM) om risicogedrag te analyseren en toon ik twee toepassingen met het CMM. In het tweede deel laat ik de toegevoegde waarde zien van multivariate machine learning technieken als voorspellingsmethoden voor het voorspellen van cognitieve vaardigheden op basis van hersenstructuren. Door de grotere steekproeven en het gebruik van geavanceerde statistische methoden kunnen betrouwbaardere conclusies worden getrokken en nauwkeurigere voorspellingen worden gedaan.

Acknowledgement

The Generation R Study is conducted by the Erasmus Medical Center in close collaboration with the Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal Health Service Rotterdam area, the Rotterdam Homecare Foundation, and the Stichting Trombosedienst and Artsenlaboratorium Rijnmond (STAR-MDC). We gratefully acknowledge the contribution of all participating children and their parents, general practitioners, hospitals, midwives, and pharmacies in Rotterdam. The general design of the Generation R Study is made possible by the Erasmus Medical Center Rotterdam, the Erasmus University Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw), the Netherlands Organization for Scientific Research (NWO), the Ministry of Health, Welfare and Sport.

Part of the data used in Chapter 4 was obtained from the Adolescent Brain Cognitive Development (ABCD) Study (`https://abcdstudy.org`), held in the NIMH Data Archive (NDA). The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U24DA041147, U01DA041093, and U01DA041025. A full list of supporters is available at `https://abcdstudy.org/federal-partners.html`. Chapter 4 reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.

The work in this thesis was supported by the Research Excellence Initiative (REI) of the Erasmus University Rotterdam grant awarded to Patrick Groenen, Henning Tiemeier, Roy Thurik, and Ingmar Franken, project number 265.403. Supercomputing resources were supported by the NWO Physical Science Division (Exacte Wetenschapen) and SURFsara (Cartesius/Snellius and Lisa compute cluster, `www.surfsara.nl`). Financial support for printing this thesis was provided by the Erasmus Research Institute of Management (ERIM) and the Generation R Study.

About the Author

Nienke Dijkstra was born in 1993 in Capelle aan den IJssel, The Netherlands. After completing VWO at the Erasmus College (Zoetermeer), she started a Bachelor of Science (BSc) in Econometrics and Operations Research at the Erasmus University Rotterdam in 2011. During the bachelor, Nienke studied at the Auckland University of Technology in New Zealand for five months. In 2017, Nienke obtained the degree of Master of Science (MSc) in Econometrics and Management Science at the Erasmus School of Economics, Erasmus University Rotterdam. Thereafter, Nienke started as a PhD candidate under the supervision of Professor Patrick Groenen, Professor Henning Tiemeier, and Professor Roy Thurik. She carried out her research within the Department of Applied Economics at the Erasmus School of Economics as a member of the Erasmus Research Institute of Management (ERIM). In addition, Nienke was affiliated to the Department of Child and Adolescent Psychiatry/Psychology and the Generation R Study Group at the Erasmus Medical Center in Rotterdam.

Nienke's research focuses on statistical modeling approaches applied to economic behavior and medical sciences. Her work has been published in the peer-reviewed journal Psychometrika. In addition, she has presented her work, amongst others, at international conferences on Computational Statistics and on Data Science, Statistics, and Visualisation, and at international meetings of the Psychometric Society and the Society for Epidemiologic Research. Nienke will continue her work as researcher at SEO Economic Research in Amsterdam, The Netherlands.

Portfolio

Publication

A Censored Mixture Model for Modeling Risk Taking. N.F.S. Dijkstra, H. Tiemeier, B. Figner, and P.J.F. Groenen, 2022, *Psychometrika*, pp. 1–27.

Manuscript under Review

Entrepreneurial Orientation and Decision Making Under Risk and Uncertainty: Experimental Evidence From the Columbia Card Task.
N.F.S. Dijkstra, K. de Groot, and C.A. Rietveld.

Manusript in Preparation

Predicting Cognition of Children Through Cortical Surface Area.
N.F.S. Dijkstra, P.J.F. Groenen, H. Tiemeier, and R. Muetzel.

Conferences Attended

- Research Meeting Organization, Strategy and Entrepreneurship (Erasmus University Rotterdam, Netherlands in 2020)
- Classification Methods in the Social and Behavioral Sciences (Tilburg University, Netherlands in 2019)
- International Meeting of Psychometric Society (Santiago, Chile in 2019)
- Society for Epidemiologic Research (Minneapolis, USA in 2019)
- Compstat (Iasi, Roemenia, 2018)
- Data Science Statistics and Visualization (Vienna, Austria in 2018)
- Econometric Institute PhD Conference (Erasmus University Rotterdam, Netherlands in 2018)

- Research Meeting Generation R (Erasmus Medical Center, Netherlands in 2018)

PhD Courses and Certificates

- Advanced Econometrics II (*Tinbergen Institute*)
- Advanced Microeconometrics (*Tinbergen Institute*)
- Behavioral Decision Theory (*Erasmus Research Institute of Management*)
- Causal Mediation Analysis (*Netherlands Institute for Health Sciences*)
- Research Methodology and Measurements (*Erasmus Research Institute of Management*)
- Cambridge Certificate of Proficiency in English (*Cambridge ESOL examinations*)
- Publishing Strategy (*Erasmus Research Institute of Management*)
- Scientific Integrity (*Erasmus Research Institute of Management*)
- Spatial Econometrics (*Essex Summer School*)

Teaching Activities

- Master thesis supervision (2019, 2020)
- Bachelor thesis supervision (2018, 2019, 2020)
- Data Science and Predictive Machine Learning (2018, 2019)
- Introduction to Data Science and Machine Learning (2017, 2018)

Other Activities

- Board member of Young ESE Program

The ERIM PhD Series

The ERIM PhD Series contains PhD dissertations in the field of Research in Management defended at Erasmus University Rotterdam and supervised by senior researchers affiliated to the Erasmus Research Institute of Management (ERIM). Dissertations in the ERIM PhD Series are available in full text through: `https://pure.eur.nl`.

ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam (EUR).

Dissertations in the last four years

Abdelwahed, A., *Optimizing Sustainable Transit Bus Networks in Smart Cities*, Supervisors: Prof. W. Ketter, Dr. P. van den Berg & Dr. T. Brandt, EPS-2022-549-LIS

Ahmadi, S., *A motivational perspective to decision-making and behavior in organizations*, Supervisors: Prof. J.J.P. Jansen & Dr. T.J.M. Mom, EPS-2019-477-S&E

Albuquerque de Sousa, J.A., *International stock markets: Essays on the determinants and consequences of financial market development*, Supervisors: Prof. M.A. van Dijk & Prof. P.A.G. van Bergeijk, EPS-2019-465-F&A

Alves, R.A.T.R.M, *Information Transmission in Finance: Essays on Commodity Markets, Sustainable Investing, and Social Networks*, Supervisors: Prof. M.A. van Dijk & Dr. M. Szymanowska, EPS-2021-532-LIS

Anantavrasilp, S., *Essays on Ownership Structures, Corporate Finance Policies and Financial Reporting Decisions*, Supervisors: Prof. A. de Jong & Prof. P.G.J. Roosenboom, EPS-2021-516-F&E

Arampatzi, E., *Subjective Well-Being in Times of Crises: Evidence on the Wider Impact of Economic Crises and Turmoil on Subjective Well-Being*, Supervisors: Prof. H.R. Commandeur, Prof. F. van Oort & Dr. M.J. Burger, EPS-2018-459-S&E

Arslan, A.M., *Operational Strategies for On-demand Delivery Services*, Supervisors: Prof. R.A. Zuidwijk & Dr. N.A.H. Agatz, EPS-2019-481-LIS

Aydin Gökgöz, Z., *Mobile Consumers and Applications: Essays on Mobile Marketing*, Supervisors: Prof. G.H. van Bruggen & Dr. B. Ataman, EPS-2021-519-MKT

Azadeh, K., *Robotized Warehouses: Design and Performance Analysis*, Supervisors: Prof. dr. ir M.B.M. de Koster & Prof. D. Roy, EPS-2021-515-LIS

Avci, E., *Surveillance of Complex Auction Markets: a Market Policy Analytics Approach*, Supervisors: Prof. W. Ketter, Prof. H.W.G.M. van Heck & Prof. D.W. Bunn, EPS-2018-426-LIS

Balen, T.H. van, *Challenges of Early Stage Entrepreneurs: the Roles of Vision Communication and Team Membership Change*, Supervisors: Prof. J.C.M. van den Ende & Dr. M. Tarakci, EPS-2019-468-LIS

Bansraj, S.C., *The Principles of Private Equity: Ownership and Acquisitions*, Supervisors: Prof. J.T.J Smit & Dr. V. Volosovych, EPS-2020-507-F&A

Bavato, D., *With New Eyes: The recognition of novelty and novel ideas*, Supervisors: Prof. D.A. Stam & Dr. S. Tasselli, EPS-2020-500-LIS

Bernoster, I., *Essays at the Intersection of Psychology, Biology, and Entrepreneurship*, Supervisors: Prof. A.R. Thurik, Prof. I.H.A. Franken & Prof. P.J.F. Groenen, EPS-2018-463-S&E

Blagoeva, R.R., *The Hard Power Of Soft Power: A behavioral strategy perspective on how power, reputation, and status affect firms*, Supervisors: Prof. J.J.P. Jansen & Prof. T.J.M. Mom, EPS-2020-495-S&E

Breugem, T., *Crew Planning at Netherlands Railways: Improving Fairness, Attractiveness, and Efficiency*, Supervisors: Prof. D. Huisman & Dr. T.A.B. Dollevoet, EPS-2020-494-LIS

Bunderen, L. van, *Tug-of-War: Why and when teams get embroiled in power struggles*, Supervisors: Prof. D.L. van Knippenberg & Dr. L. Greer, EPS-2018-446-ORG

Burg, G.J.J. van den, *Algorithms for Multiclass Classification and Regularized Regression*, Supervisors: Prof. P.J.F. Groenen & Dr. A. Alfons, EPS-2018-442-MKT

Chan, H.Y., *Decoding the consumer's brain: Neural representations of consumer experience*, Supervisors: Prof. A. Smidts & Dr. M.A.S. Boksem, EPS-2019-493-MKT

Couwenberg, L., *Context dependent valuation: A neuroscientific perspective on consumer decision-making*, Supervisors: Prof. A. Smit, Prof. A.G. Sanfrey & Dr. M.A.S. Boksem, EPS-2020-505-MKT

Dalmeijer, K., *Time Window Assignment in Distribution Networks*, Supervisors: Prof. A.P.M. Wagelmans & Dr. R. Spliet, EPS-2019-486-LIS

Dolgova, E., *On Getting Along and Getting Ahead: How Personality Affects Social Network Dynamics*, Supervisors: Prof. P.P.M.A.R. Heugens & Prof. M.C. Schippers, EPS-2019-455-S&E

Fasaei, H., *Changing the Narrative: The Behavioral Effects of Social Evaluations on the Decision Making of Organizations*, Supervisors: Prof. J.J.P. Jansen, Prof. T.J.M. Mom & Dr. M.P. Tempelaar, EPS-2020-492-S&E

Eijlers, E., *Emotional Experience and Advertising Effectiveness: on the use of EEG in marketing*, Supervisors: Prof. A. Smidts & Prof. M.A.S. Boksem, EPS-2019-487-MKT

El Nayal, O.S.A.N., *Firms and the State: An Examination of Corporate Political Activity and the Business-Government Interface*, Supervisors: Prof. J. van Oosterhout & Dr. M. van Essen, EPS-2018-469-S&E

Frick, T.W., *The Implications of Advertising Personalization for Firms, Consumer, and Ad Platfroms*, Supervisors: Prof. T. Li & Prof. H.W.G.M. van Heck, EPS-2018-452-LIS

Fu, G., *Agency Problems in the Mutual Fund Industry*, Supervisors: Prof. M.J.C.M. Verbeek & Dr. E. Genc, EPS-2022-526-F&A

Fytraki, A.T., *Behavioral Effects in Consumer Evaluations of Recommendation Systems*, Supervisors: Prof. B.G.C. Dellaert & Prof. T. Li, EPS-2018-427-MKT

Gai, J., *Contextualized Consumers: Theories and Evidence on Consumer Ethics, Product Recommendations, and Self-Control*, Supervisors: Prof. S. Puntoni & Prof. S.T.L. Sweldens, EPS-2020-498-MKT

Ghazizadeh, P., *Empirical Studies on the Role of Financial Information in Asset and Capital Markets*, Supervisors: Prof. A. de Jong & Prof. E. Peek, EPS-2019-470-F&A

Giessen, M. van der, *Co-creating Safety and Security: Essays on bridging disparate needs and requirements to foster safety and security*, Supervisors: Prof. G. Jacobs, Prof. J.P. Cornelissen & Prof. P.S. Bayerl, EPS-2022-542-ORG

Hanselaar, R.M., *Raising Capital: On pricing, liquidity and incentives*, Supervisors: Prof. M.A. van Dijk & Prof. P.G.J. Roosenboom, EPS-2018-429-F&A

Harms, J.A., *Essays on the Behavioral Economics of Social Preferences and Bounded Rationality*, Supervisors: Prof. H.R. Commandeur & Dr. K.E.H. Maas, EPS-2018-457-S&E

Hartleb, J., *Public Transport and Passengers: Optimization Models that Consider Travel Demand*, Supervisors: Prof. D. Huisman, Prof. M. Friedrich & Dr. M.E. Schmidt, EPS-2021-535-LIS

Hendriks, G., *Multinational Enterprises and Limits to International Growth: Links between Domestic and Foreign Activities in a Firm's Portfolio*, Supervisors: Prof. P.P.M.A.R. Heugens & Dr. A.H.L. Slangen, EPS-2019-464-S&E

Hengelaar, G.A., *The Proactive Incumbent: Holy grail or hidden gem? Investigating whether the Dutch electricity sector can overcome the incumbent's curse and lead the sustainability transition*, Supervisors: Prof. R.J.M. van Tulder & Dr. K. Dittrich, EPS-2018-438-ORG

Hoogervorst, R., *Improving the Scheduling and Rescheduling of Rolling Stock: Solution Methods and Extensions*, Supervisors: Prof. D. Huisman & Dr. T.A.B. Dollevoet, EPS-2021-534-LIS

Jia, F., *The Value of Happiness in Entrepreneurship*, Supervisors: Prof. D.L. van Knippenberg & Dr. Y. Zhang, EPS-2019-479-ORG

Kampen, S. van, *The Cross-sectional and Time-series Dynamics of Corporate Finance: Empirical evidence from financially constrained firms*, Supervisors: Prof. L. Norden & Prof. P.G.J. Roosenboom, EPS-2018-440-F&A

Karali, E., *Investigating Routines and Dynamic Capabilities for Change and Innovation*, Supervisors: Prof. H.W. Volberda, Prof. H.R. Commandeur & Dr. J.S. Sidhu, EPS-2018-454-S&E

Kerkkamp, R.B.O., *Optimisation Models for Supply Chain Coordination under Information Asymmetry*, Supervisors: Prof. A.P.M. Wagelmans & Dr. W. van den Heuvel, EPS-2018-462-LIS

Kim, T.Y., *Data-driven Warehouse Management in Global Supply Chains*, Supervisors: Prof. R. Dekker & Dr. C. Heij, EPS-2018-449-LIS

Kishore Bhoopalam, A., *Truck Platooning: Planning and Behaviour*, Supervisors: Prof. R.A. Zuidwijk & Dr. N.A.H. Agatz, EPS-2021-540-LIS

Klitsie, E.J., *Strategic Renewal in Institutional Contexts: The paradox of embedded agency*, Supervisors: Prof. H.W. Volberda & Dr. S. Ansari, EPS-2018-444-S&E

Koolen, D., *Market Risks and Strategies in Power Systems Integrating Renewable Energy*, Supervisors: Prof. W. Ketter & Prof. R. Huisman, EPS-2019-467-LIS

Kong, L., *Essays on Financial Coordination*, Promotor: Prof. M.J.C.M. Verbeek, Dr. D.G.J. Bongaerts & Dr. M.A. van Achter. EPS-2019-433-F&A

Koritarov, V.D., *The Integration of Crisis Communication and Regulatory Focus: Deconstructing and Optimizing the Corporate Message*, Supervisors: Prof. C.B.M. van Riel, Dr. G.A.J.M. Berens & Prof. P. Desmet, EPS-2021-522-ORG

Korman, B., *Leader-Subordinate Relations: The Good, the Bad and the Paradoxical*, Supervisors: S.R. Giessner & Prof. C. Tröster, EPS-2021-511-ORG

Kyosev, G.S., *Essays on Factor Investing*, Supervisors: Prof. M.J.C.M. Verbeek & Dr. J.J. Huij, EPS-2019-474-F&A

Lamballais Tessensohn, T., *Optimizing the Performance of Robotic Mobile Fulfillment Systems*, Supervisors: Prof. M.B.M de Koster, Prof. R. Dekker & Dr. D. Roy, EPS-2019-411-LIS

Legierse, W., *The Timing and Pricing of Initial Public Offerings: Evidence from the Low Countries*, Supervisors: Prof. A. de Jong & Prof. P.G.J. Roosenboom, EPS-2022-543-F&A

Leung, W.L., *How Technology Shapes Consumption: Implications for Identity and Judgement*, Supervisors: Prof. S. Puntoni & Dr. G. Paolacci, EPS-2019-485-MKT

Li, Wei., *Competition in the Retail Market of Consumer Packaged Goods*, Supervisors: Prof. D.Fok & Prof. Ph.H.B.F. Franses, EPS-2021-503-MKT

Li, X., *Dynamic Decision Making under Supply Chain Competition*, Supervisors: Prof. M.B.M de Koster, Prof. R. Dekker & Prof. R. Zuidwijk, EPS-2018-466-LIS

Lieshout, R. van, *Integration, Decentralization and Self-Organization Towards Better Public Transport*, Supervisors: Prof. D. Huisman, Dr. P.C. Bouman & Dr.ir. J.M. van den Akker, EPS-2022-547-LIS

Maas, A.J.J., *Organizations and their external context: Impressions across time and space*, Supervisors: Prof. P.P.M.A.R. Heugens & Prof. T.H. Reus, EPS-2019-478-S&E

Maira, E., *Consumers and Producers*, Supervisors: Prof. S. Puntoni & Prof. C. Fuchs, EPS-2018-439-MKT

Manouchehrabadi, B., *Information, Communication and Organizational Behavior*, Supervisors: Prof. G.W.J. Hendrikse & Dr. O.H. Swank, EPS-2020-502-ORG

Matawlie, N., *Through Mind and Behaviour to Financial Decisions*, Supervisors: Prof. J.T.J. Smit & Prof. P. Verwijmeren, EPS-2020-501-F&A

Mirzaei, M., *Advanced Storage and Retrieval Policies in Automated Warehouses*, Supervisors: Prof. M.B.M. de Koster & Dr. N. Zaerpour, EPS-2020-490-LIS

Nair, K.P., *Strengthening Corporate Leadership Research: The relevance of biological explanations*, Supervisors: Prof. J. van Oosterhout & Prof. P.P.M.A.R. Heugens, EPS-2019-480-S&E

Nikulina, A., *Interorganizational Governance in Projects: Contracts and collaboration as alignment mechanisms*, Supervisors: Prof. J.Y.F. Wynstra & Prof. L. Volker, EPS-2021-523-LIS

Novales Uriarte, A., *Thriving with Digitized Products: How Firms Leverage their Generative Capacity via Experimentation, Learning, and Collaboration*, Supervisors: Prof. H.W.G.M. van Heck & Prof. M.Mocker, EPS-2022-544-LIS

Nullmeier, F.M.E., *Effective contracting of uncertain performance outcomes: Allocating responsibility for performance outcomes to align goals across supply chain actors*, Supervisors: Prof. J.Y.F. Wynstra & Prof. E.M. van Raaij, EPS-2019-484-LIS

Paul, J., *Online Grocery Operations in Omni-channel Retailing: Opportunities and Challenges*, Supervisors: Prof. M.B.M de Koster & Dr. N.A.H. Agatz, EPS-2022-541-LIS

Petruchenya, A., *Essays on Cooperatives: Emergence, Retained Earnings, and Market Shares*, Supervisors: Prof. G.W.J. Hendriks & Dr. Y. Zhang, EPS-2018-447-ORG

Pocchiari, M., *Managing Successful and Resilient Shared-Interest Communities: The Role of Digitization Technologies and Disruptive Events*, Supervisors: Prof. G.H. van Bruggen & Dr. J.M.T. Roos, EPS-2022-552-MKT

Polinder, G.J., *New Models and Applications for Railway Timetabling*, Supervisors: Prof. D. Huisman & Dr. M.E. Schmidt, EPS-2020-514-LIS

Qian, Z., *Time-Varying Integration and Portfolio Choices in the European Capital Markets*, Supervisors: Prof. W.F.C. Verschoor, Prof. R.C.J. Zwinkels & Prof. M.A. Pieterse-Bloem, EPS-2020-488-F&A

Ramezan Zadeh, M.T., *How Firms Cope with Digital Revolution: Essays on Managerial and Organizational Cognition*, Supervisors: Prof. H.W. Volberda & Prof. J.P. Cornelissen, EPS-2021-508-S&E

Ratara, C., *Behavioural and Neural Evidence for Processes Underlying Biases in Decision-Making*, Supervisors: Prof. A. Smidts & Dr. M.A.S. Boksem, EPS-2022-548-MKT

Reh, S.G., *A Temporal Perspective on Social Comparisons in Organizations*, Supervisors: Prof. S.R. Giessner, Prof. N. van Quaquebeke & Dr. C. Troster, EPS-2018-471-ORG

Riessen, B. van, *Optimal Transportation Plans and Portfolios for Synchromodal Container Networks*, Supervisors: Prof. R. Dekker & Prof. R.R. Negenborn, EPS-2018-448-LIS

Romochkina, I.V., *When Interests Collide: Understanding and modeling interests alignment using fair pricing in the context of interorganizational information systems*, Supervisors: Prof. R.A. Zuidwijk & Prof. P.J. van Baalen, EPS-2020-451-LIS

Schneidmüller, T., *Engaging with Emerging Technologies: Socio-cognitive foundations of incumbent response*, Supervisors: Prof. H. Volberda & Dr. S.M. Ansari, EPS-2020-509-S&E

Schouten, K.I.M., *Semantics-driven Aspect-based Sentiment Analysis*, Supervisors: Prof. F.M.G. de Jong, Prof. R. Dekker & Dr. F. Frasincar, EPS-2018-453-LIS

Sihag, V., *The Effectiveness of Organizational Controls: A meta-analytic review and an investigation in NPD outsourcing*, Supervisors: Prof. J.C.M. van den Ende & Dr. S.A. Rijsdijk, EPS-2019-476-LIS

Sluga, A., *Hour of Judgment: On judgment, decision making, and problem solving under accountability*, Supervisors: Prof. F.G.H. Hartmann & Dr. M.A.S. Boksem, EPS-2021-520-F&A

Slob, E., *Integrating Genetics into Economics*, Supervisors: Prof. A.R. Thurik, Prof. P.J.F. Groenen & Dr. C.A. Rietveld, EPS-2021-517-S&E

Smolka, K.M., *Essays on Entrepreneurial Cognition, Institution Building and Industry Emergence*, Supervisors: P.P.M.A.R. Heugens & Prof. J.P. Cornelissen, EPS-2019-483-S&E

Stirnkorb, S., *Changes in the Information Landscape and Capital Market Communication*, Supervisors: Prof. E. Peek & Prof. M. van Rinsum, EPS-2021-536-F&A

Stuppy, A., *Essays on Product Quality*, Supervisors: Prof. S.M.J. van Osselaer & Dr. N.L. Mead, EPS-2018-461-MKT

Suurmond, R., *In Pursuit of Supplier Knowledge: Leveraging capabilities and dividing responsibilities in product and service contexts*, Supervisors: Prof. J.Y.F Wynstra & Prof. J. Dul, EPS-2018-475-LIS

Tierean, S.H., *Mind the Gap: The role of psychic distance and supplier's reputation in international buyer-supplier relationships*, Supervisors: Prof. C.B.M. van Riel, Prof. G. van Bruggen & Dr. G.A.J.M. Berens, EPS-2022-551-ORG

Toxopeus, H.S., *Financing sustainable innovation: From a principal-agent to a collective action perspective*, Supervisors: Prof. H.R. Commandeur & Dr. K.E.H. Maas, EPS-2019-458-S&E

Tuijn, M. *Target the untargeted: essays in unconventional disclosures and policies*, Supervisors: Prof. E. Peek & Prof. E.M. Roelofsen, EPS-2020-499-F&A

Turturea, R., *Overcoming Resource Constraints: The Role of Creative Resourcing and Equity Crowdfunding in Financing Entrepreneurial Ventures*, Supervisors: Prof. P.P.M.A.R. Heugens, Prof. J.J.P. Jansen & Dr. I. Verheuil, EPS-2019-472-S&E

Valboni, R., *Building Organizational (Dis-)Abilities: The impact of learning on the performance of mergers and acquisitions*, Supervisors: Prof. T.H. Reus & Dr. A.H.L. Slangen, EPS-2020-407-S&E

Visser, T.R., *Vehicle Routing and Time Slot Management in Online Retailing*, Supervisors: Prof. A.P.M. Wagelmans & Dr. R. Spliet, EPS-2019-482-LIS

Vongswasdi, P., *Accelerating Leadership Development: An evidence-based perspective*, Supervisors: Prof. D. van Dierendonck & Dr. H.L. Leroy, EPS-2020-512-ORG

Waltré, E., *Leading for Performance in Adversity: Managing Failure, Negative Emotions, and Self-Threats*, Supervisors: Prof. D.L. van Knippenberg & Dr. H.M.S. Dietz, EPS-2022-513-ORG

Wang, R., *Those Who Move Stock Prices*, Supervisors: Prof. P. Verwijmeren & Prof. S. van Bekkum, EPS-2019-491-F&A

Wiegmann, P.M., *Setting the Stage for Innovation: Balancing Diverse Interests through Standardisation*, Supervisors: Prof. H.J. de Vries & Prof. K. Blind, EPS-2019-473-LIS

Wijaya, H.R., *Praise the Lord!: Infusing Values and Emotions into Neo-Institutional Theory*, Supervisors: Prof. P.P.M.A.R. Heugens & Prof. J.P. Cornelissen, EPS-2019-450-S&E

Williams, A.N., *Make Our Planet Great Again: A Systems Perspective of Corporate Sustainability*, Supervisors: Prof. G.M. Whiteman & Dr. S. Kennedy, EPS-2018-456-ORG

Witte, C.T., *Bloody Business: Multinational investment in an increasingly conflict-afflicted world*, Supervisors: Prof. H.P.G. Pennings, Prof. H.R. Commandeur & Dr. M.J. Burger, EPS-2018-443-S&E

Wu, J., *A Configural Approach to Understanding Voice Behavior in Teams*, Supervisors: Prof. D.L. van Knippenberg & Prof. S.R. Giessner, EPS-2020-510-ORG

Yalcin, G., *Consumers in the Age of AI: Understanding Reactions Towards Algorithms and Humans in Marketing Research*, Supervisors: Prof. S. Puntoni & Dr. A. Klesse, EPS-2022-550-MKT

Ye, Q.C., *Multi-objective Optimization Methods for Allocation and Prediction*, Supervisors: Prof. R. Dekker & Dr. Y. Zhang, EPS-2019-460-LIS

Zhang, Q., *Financing and Regulatory Frictions in Mergers and Acquisitions*, Supervisors: Prof. P.G.J. Roosenboom & Prof. A. de Jong, EPS-2018-428-F&A

Zhu, S., *Spare Parts Demand Forecasting and Inventory Management: Contributions to Intermittent Demand Forecasting, Installed Base Information and Shutdown Maintenance*, Supervisors: Prof. R. Dekker & Dr. W.L. van Jaarsveld, EPS-2021-538-LIS

Zon, M. van, *Cost Allocation in Collaborative Transportation*, Supervisors: Prof. A.P.M. Wagelmans, Dr. R. Spliet & Dr. W. van den Heuvel, EPS-2021-530-LIS

Behavior and abilities are arguably the most fundamental concepts in life. Hence, these concepts and their causes and consequences, are studied in many disciplines, such as economics, psychology, sociology, and health sciences. The research of behaviors and abilities is quickly changing: data sets are rapidly growing in size, both in terms of observations and variables. The increase in the sample size of studies, repeated measurements and complex multi-level designs necessitates the use of modern – often more complex and flexible – statistical methods. This thesis employs modern statistical models in the field of fundamental economic behavior and medical sciences with the aim to generate more reliable inferences and more accurate predictions than would have been the case with established methods. In the present thesis I will give two examples. In the first part of the thesis, I develop the Censored Mixture Model (CMM) to analyze risk behavior and show two applications. In the second part, I compare two machine learning approaches with various simpler commonly used baseline approaches in their performance to predict cognitive ability using brain morphology.

ERiM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are the Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focused on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

ERiM

ERIM PhD Series
Research in Management

Erasmus University Rotterdam (EUR)
Erasmus Research Institute of Management
Mandeville (T) Building
Burgemeester Oudlaan 50
3062 PA Rotterdam, The Netherlands

P.O. Box 1738
3000 DR Rotterdam, The Netherlands
T +31 10 408 1182
E info@erim.eur.nl
W www.erim.eur.nl