Vocational Education and Dropout Rates in the Netherlands

P.M.M. Cox
Student nr: 3850048
Universiteit Utrecht

L.P. van der Tas
Student nr: 3873013
Universiteit Utrecht

Supervisor 1:
M. Dinkova MSc

Supervisor 2:
dr. Y.K. Grift

January 26, 2015

Abstract

This paper has been prepared as a BSc Thesis at Utrecht University, the Netherlands. First and foremost, it explores the psychological (Big Five personality traits), social and personal causes of dropout quantitatively through the use of multiple multivariate logistic regression models. Significant results have been found for age, Big Five's conscientiousness, overall learning intelligence, as well as some interaction effects with student origin. Furthermore, the economic consequences of dropout behaviour in general have been illustrated theoretically. We argue that dropout behaviour negatively influences economic prosperity, via human- and social capital theory.

Contents

Introduction

> *"Each year's class of dropouts will cost the country over \$200 billion during their lifetimes in lost earnings and unrealized tax revenue."*
>
> — James S. Catterall

Choosing a major and sticking with it is a relatively rare thing in the modern education sector. This applies to both higher (or academic) and vocational education, which is schooling based on occupation. Regardless of the motive, each year an enormous amount of students throw in the towel and decide to drop out of their current education. As this chapter's quote states, however, the macroeconomic implications are vast.

Dropout rates in both higher and vocational education in the Netherlands have been a rather hot topic over the last few years, leading to the Dutch government aiming to reduce dropout rates in vocational education below 5% in 2016. Dropout is defined as leaving education without acquiring a basic qualification, which is "a diploma at the level of senior general secondary education, pre-university education, or level-2 of senior secondary vocational education" [Statistics Netherlands, 2015]. During the academic year ending in 2013, this figure had already fallen to 5.7%, which is among the best in the European Union (EU) and low compared to the EU average of 13.7% [Dutch Government, 2014]. Hence, the Dutch government faces a challenge to close the final gap towards their goal. On top of that, with the baby-boom generation completely retiring in 2020, there will be sheer unsatisfiable demand for a young, talented workforce in many countries, including the Netherlands. In order to educate as many students as possible, it is of paramount importance that high dropout rates are counteracted effectively. Meanwhile the European Commission keeps a close eye, for they demand a smaller than 10% rate dropout rate as of 2020 for all member countries. Not only because it hampers economic growth, but also because students who drop out are likely to end up poor, unemployed and therefore simply costly for society, which further exacerbates the issue [European Commission, 2011].

Evidently, scientists are interested in motivations of students that decide to exit their current education. This has caused them to link dropout behaviour to personality characteristics, which are psychologically best explained by the Big Five personality traits. By these five criteria, every person can be 'rated' and hence accurately described. We will elaborate on this in the theoretical framework in chapter one.

This paper's main goal is to provide insight in the causes of dropout in Dutch vocational education. First and foremost, the quantitative relationship between personality characteristics and dropout rates will be explored. We ask ourselves: What are the most important personality characteristics that can explain dropout behaviour? The focus of this quantitative analysis will be the context of Dutch vocational education, rather than higher education or secondary school. This appears to be a niche in the existing literature and therefore merits further research. Several logistic regression models will be explored and estimated. Afterwards, we discuss the qualitative relationship between dropout rates in both vocational and higher education, and the supplementary economic implications for the Netherlands through human- and social capital theory. Thereby answering the question of how is an economy affected by dropout in vocational and higher education. Please refer to the conceptual framework in figure 1 for a visualisation of this research's design.

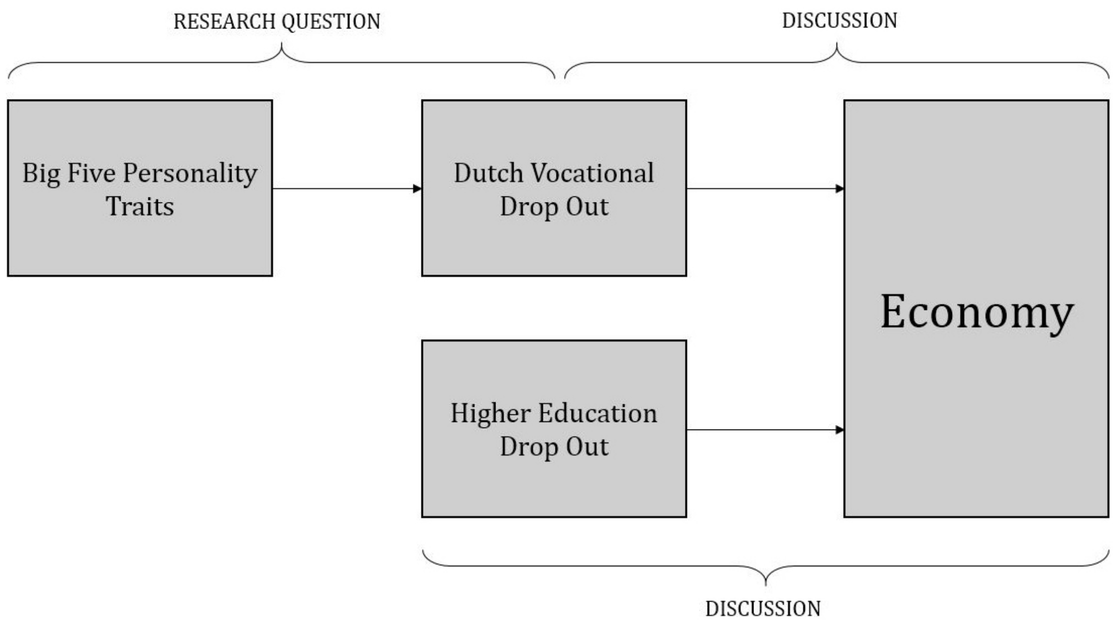

Figure 1: Conceptual Framework

The structure of this paper is as follows: Chapter one discusses the theoretical framework of our research. Afterwards, chapter two introduces and describes the data, as well as the methodology applied throughout the research. Chapter three will discuss the results we obtained from the analyses. Then, in chapter four, a theoretical discussion of the results and their implications in economic context is given. Finally, chapter five concludes the report and will summarise and evaluate our findings as well as the limitations encountered in the research process.

Chapter 1

Theoretical Framework

"The human mind is not a terribly logical or consistent place."

— Jim Butcher

This chapter will provide an overview and a description of the theories that function as the very building blocks and foundation of our research. We successively introduce the Big Five personality traits, a literature overview and human- and social capital theory.

1.1 The Big Five

Humans are social creatures. In their regular interaction with equals they display distinctive and characteristic patterns of thought, emotion and behaviour. This makes up their *personality*. Questioned to describe other people's personalities, one often uses rather narrow terms as outgoing, organised or stubborn. Personality psychologists throughout history have faced the challenge to reduce this virtually infinite set of personality traits to a much smaller subset of more comprehensible traits [Nolen-Hoeksema et al., 2009]. For our research, we have listed the most widely-used names and their definitions below, which are also used in popular textbooks for introductory psychology courses [Nolen-Hoeksema et al., 2009].

Big Five Personality Traits

- **Openness**: the degree to which one's basic orientation is focused outward;
- **Conscientiousness**: the degree to which one is organised, self disciplined and dependable;
- **Extraversion**: the degree to which one is energetic, assertive and social or talkative;

- **Agreeableness**: the degree to which one is compassionate, cooperative, helpful and well-tempered;
- **Neuroticism/emotional stability**: the degree to which one deals easily with emotional distress like anger, anxiety and depression.

The quest for the Big Five, as the traits would later be baptised, started already in 1932, when William McDougall wrote in the first issue of *Character and Personality* that "personality may with advantage be broadly analysed into five distinguishable but inseparable factors, namely, intellect, character, temperament, disposition and temper. Each of these is highly complex and comprises many variables" [McDougall, 1932] [p. 15]. Along the way, many colleagues of his have questioned, tested and reshaped his early-suggested factors [Digman, 1990]. Eventually evolving into what is now known as the Big Five personality traits, or five-factor model, where there is still obscurity regarding the names and interpretations of the factors: Different psychologists have varying opinions on the correct names of the traits. For example, some psychologists refer to emotional stability where others say neuroticism. Also, agreeableness and altruism are used interchangeably.

1.2 Literature Overview

The relation between students' personalities and dropout rates is an extensively studied topic of research. As we show in this section, this has been researched both directly and indirectly. The indirect approach makes use of academic motivation as a mediating factor. We will give a comprehensive overview of the major findings of earlier researches. Note that in these kind of researches, it is habitual to measure the Big Five on a one-to-five scale (Likert-type scale).

To start off, a Dutch study to the effect of personality characteristics on study outcome, concerning students in Universities of Applied Sciences, found that of the Big Five especially conscientiousness positively relates to study outcome [Van Bragt et al., 2011]. According to these researchers, students who score high on the conscientiousness scale, display good time management skills and organisation of their learning process. That, in result, makes them less likely to drop- or stop-out [Van Bragt et al., 2011]. Stop-out is a term for students who leave their studies on the current institution, but enroll in a new major at the same or another institution in the same, or the next academic year. Different, but more easily observable characteristics are studied in an Italian case study for the *Sapienza University* in Rome [Belloc et al., 2011]. It considers four possible outcomes for students: they drop out, change faculty, change institution or continue studying. Their results suggest positive relationships, i.e. dropout being more likely, for being male as well as being of Italian origin [Belloc et al., 2011]. Regarding personality characteristics, it is worth mentioning that a Dutch research, employing the same dataset as we do in this research, has found a negative (although very small) effect for age [Grift et al., 2009]. The higher a student's age, the less likely he or she is to leave education without a diploma.

That said, we turn to a different model that looks at students' updating of information [Stratton et al., 2005]. These researchers find that renewed information following from first-year academic performance is an important determinant of dropout. Besides, they state personal circumstances as delayed matriculation and parental and marital status can determine drop- or stop-out (temporary in nature) [Stratton et al., 2005]. On top of that, a review of Chinese technical and vocational education and training proves that dropouts are not more financially constrained across the board [Yao et al., 2013]. However, whether their parents or guardians were at home had statistically significant impact on their student's success and dropout rates at school [Yao et al., 2013]. Although those findings could be considered irrelevant, we want to emphasize the diversity in methodological approaches of other authors as well as display the manifold proven causes of dropout.

We now return to the core of the Big Five with a 1989 study focusing on early school dropout. This research found aggressive and low-performing students to be more likely to drop out [Cairns et al., 1989]. Since aggressiveness is a particularly low form of agreeableness, this would suggest that agreeableness also has a positive impact on dropout prevention. Furthermore, research has shown that the Big Five display significant effects on academic performance through three distinguishable motives: Avoidance, engagement and achievement [Komarraju and Karau, 2005]. With those intermediary motives, researchers found that students with a high degree of avoidance were more likely to score high on the neuroticism and extraversion scales. In contrast, high-scoring students on mainly openness but also extraversion showed higher levels of engagement with their studies. Finally, they found that students whose academic motivation was mainly based on achievement and outperforming others, were characterised by high levels of conscientiousness, neuroticism (i.e. low emotional stability) and openness.

Two of the same authors conducted another experiment four years later on students in American higher education, reinforcing the (rather consistent) belief that conscientiousness is a strong predictor of academic motivation. Although this research states nothing directly about dropout or retention, the findings are noteworthy. More intrinsically motivated students were found more open to new experiences and extraversion seemed to go hand in hand with extrinsic motivation. Also, students that scored low on the agreeableness scale, tend to be poorly motivated when it comes to studying. On top of all this, and contrary to intuition, the Big Five explain more variance in GPA (Grade Point Average) than the different academic motivation types do. In simpler terms: whether you are likely to succeed in college depends more on personality than on ability and motivation [Komarraju et al., 2009]. This mediating role of academic motivation is underlined by Hazrati-Viari et al., who studied more than 200 University students in order to discover the relation in Iran [Hazrati-Viari et al., 2012].

Where Komarraju et al. [2009] used three types of motivation, the following study distinguishes seven. In 2010, 451 first-year college students were subject of another research with highly similar findings to the previous ones discussed. The degree to which students were extravert, agreeable, conscientious and open to new experiences positively related to how intrinsically motivated they were. As before, extrinsic motivation correlated positively with agreeableness, neuroticism, conscientiousness and extroversion [Clark and Schroth, 2010].

In summary, the Big Five factors conscientiousness and agreeableness appear to be the most consistent predictors of dropout behaviour. That is, we expect that the higher conscientiousness and agreeableness-levels, the less likely dropout is bound to happen. However, judging from the wide variety of methodological approaches in this literature overview, every case studied seems to be different. It requires a unique approach and the right tools for the job depending on the data at hand. Further on, in the methodology in chapter two, we will explain the tools we will be using for this paper.

1.3 Human Capital Theory

Even in the history of economics the principle of human capital was already considered by one of the most famous names in economics: Adam Smith. He proclaimed "all of the acquired and useful abilities of all of the inhabitants of a country as a part of capital" [Schultz, 1961]. From the moment a human being is born, he or she starts accruing a stock of knowledge and skills as a result of innate talent, or the experiences that specific person goes through. By entering an institution of (higher) learning, an individual can even further expand and accelerate this learning process. The entire accrued skill set of individuals, which eventually becomes vendible, is known as *human capital*; a very essential aspect of modern labour economics [Acemoglu and Autor, 2014]. With respect to our research, the theory described in this section provides the foundations on which we will build later on in (mainly) the discussion section of chapter four, when exploring the economic consequences of dropout.

Now, we will explain the characteristics of human capital by providing a simple mathematical framework of how it comes into existence, how it develops and how it influences the total output of an economy.

Human capital is a productive entity by nature as can be seen in the equation 1.1 displayed below: Human capital (h) multiplies the effects of labour (L) to, along with physical capital (K), increase total output (Y).

$$Y = F(K, hL) \tag{1.1}$$

In order for human capital to come into existence, it must first be produced. This production occurs as described in the previous section. The amount produced, and therefore the growth of human capital, is dependent on the level of investment ($I_{h,t}$) in human capital. This is shown in equation 1.2.

$$h_{t+1} = h_t - dh_t + I_{h,t} \tag{1.2}$$

Equation 1.3 explains that investment in human capital is negatively influenced by the amount of investment in physical capital ($I_{K,t}$) and the consumption (C), ceteris paribus on output (Y). This is a logical assumption because money invested in physical capital, or spent on consumption, cannot be invested in human capital; it is mutually exclusive.

Also, just like physical capital, human capital depreciates. The degree of depreciation is denoted by (d) in equation 1.2. Depreciation may seem obscure when associated with human capital. However, when considering the two kinds of depreciation of human capital, it makes sense; *technical* depreciation and *economic* depreciation [de Grip and van Loo, 2002]. The former refers to the actual loss in human capital by remaining idle (i.e. not practicing the accrued skills, which causes them to decay). Economic depreciation, on the other hand, refers to the loss of the degree to which an individual's skills can be marketed [Arrazola et al., 2005]. For instance, the skills of a manual scribe in a law room became more or less useless with the invention of the typewriter.

$$I_{h,t} = Y_t - C_t - I_{K,t} \tag{1.3}$$

Furthermore, a return is earned on human capital. This can be described by equation 1.4 where the wage rate w, associated with the level human capital, denotes the return. Following simple mercantile reasoning, this return is intuitive. It would not make sense to sink money in human capital investment if the eventual returns would not more than compensate the cost of the investment (plus interest).

$$Y = whL + rK \tag{1.4}$$

Finally, one must note that the time (t) is rival in use. Therefore, time invested in human capital cannot be used for e.g. exercising labour or producing physical capital [Weil, 2013]. Overall we see that, following from the presented equations, lower levels of human capital have a negative infuence on economic growth. Later on, we will introduce dropout in this reasoning as well.

1.4 Social Capital

To properly analyze the effects of human capital on dropout rates, we apply the functional effect of *social capital* in conjunction with the simple mathematical framework mentioned in the previous section. Social capital is a tool introduced by James S. Coleman. It is best delineated by its function: expediting certain (inter)actions between and among actors [Coleman, 1988].

Consider the following tangible example used by Coleman in his article: In the process of a sale in the diamond market, the selling party will transfer a bag of diamonds to the buying party to be freely examined in private. This practice seems risky, for the goods may be worth a fortune, but this form of transparency allows the market to function efficiently: one can determine the legitimacy of the goods pre-transaction. However, who is to say the selling merchant will not swap some of the best diamonds for less pristine ones after the inspection has been finished and the transaction has been committed to? This is where social capital comes into play: It facilitates an action (i.e. the sale) among the agents (i.e. the diamond merchants) by, in this case, providing social insurance of ensuring fair transaction. In this example, the social capital is derived from the close relationships among the diamond salesmen. For instance, consider New York, where the diamond market is mostly governed by the Jewish people. These people have strong community ties through e.g. religion, family, marriage and ethnicity. Should an agent commit a fraudulent transaction, it is frowned upon by the community and may cause the fraudulent merchant to be exiled from it, as well as from future business [Coleman, 1988].

Whereas physical capital is almost completely tangible, subsiding in material assets such as machinery and structures, human capital is already less tangible: It is embodied in the accrued knowledge and skills of an individual [Schultz, 1961]. Social capital is even more intangible, for it exists in the relationships among individuals [Coleman, 1988].

1.5 Dropout, Social- and Human Capital

Research done by Hoffer and Coleman on the effects of social capital on dropout rates of students at secondary school yields significant results. They prove that the lower the degree of social capital in the environment of the student, the more likely it is for that student to drop out of the schooling program they are currently enrolled in [Coleman and Hoffer, 1987]. However, for this result to make sense for our research, we must first define social capital in the context of our specific dropout research in this paper, as well as describe its link to human capital.

Social capital in the abovementioned setting is defined as the interaction between the student's parent(s) and the student in question, as well as the mere physical presence of the parent(s). The consistent presence of parents generally strengthens the bond between parents and child over time. Furthermore, through ongoing interaction, the student can be motivated towards better school performance. Moreover, regular parent-child interaction generally allows the parents to convey (part of) their human capital to their children over time [Coleman, 1988]. It must also be mentioned that the strengthening of human capital via social capital can also occur as a result of peer-to-peer influence. This will raise total human capital, as indicated by equation 1.5 (based on 1.2), with a multiplicative factor for social capital (φ) ceteris paribus on investment in and depreciation of human capital.

$$h_{t+1} = \varphi h_t - dh_t + I_{h,t} \tag{1.5}$$

Eventually, as a student accrues enough human capital, college perspectives may arise. The research proves that the perspective of qualifying for college further diminishes dropout perspectives [Coleman and Hoffer, 1987]. Also, once a student enters college, the accrual of his or her human capital will accelerate. This shall vastly boost the student's return on human capital, as shown by equations 1.6, 1.7 and 1.8 (based on equation 1.4): Once he or she acquires a degree, Θ will take a value of one and the wage rate will increase by the amount Δw. Subsequently, the total economic output will be boosted through the multiplicative positive effects of social- and human capital and the increase in wage rate, ceteris paribus on physical capital (K) and its interest (r).

$$Y = w(\varphi h)L + rK \tag{1.6}$$

$$w = w + \Theta(\Delta w) \tag{1.7}$$

$$\Theta = \{0 \cup 1\} \tag{1.8}$$

From the reasoning above, it can be concluded that the dropout rates of students are diminished by human capital. Human capital is, in turn, affected by the complex framework of social capital: Positive social capital will diminish dropout both directly and through human capital [Coleman and Hoffer, 1987]. Influence through social capital can occur both as a result of influence by the family of the student (mainly the parents), or through social capital outside the family (e.g. via a student's peers). This influence occurs through "obligations, expectations, trustworthiness of the social environment, information flow capability of the social structure and norms accompanied by sanctions" [Coleman, 1988].

Chapter 2

Methodology

"Art and science have their meeting point in method."

— Edward G. Bulwer-Lytton

With the theoretical framework in place, this chapter shall firstly describe the data and will then list the methods that will be used to conduct our research. Please note that all tables and figures in this paper are created by us (the authors) unless indicated otherwise, and are based on the dataset as described in the section below.

2.1 Data Description

The analysis in this paper has been performed using a dataset containing observations from one Dutch ROC: An abbreviation for "Regionaal Opleidings Centrum" or "Regional Education Centre" in English. The dataset contains 46,424 initial observations of 479 different variables. Each observation describes a specific student. The Big Five personality traits in the dataset were measured on a Likert-type scale running from one to five. In this manner, a one determines a low level of the corresponding variable, whereas a five is the highest possible value. The levels of the Big Five traits, as well as other student characterising variables, were determined not by the students themselves, but by professional psychologists on the basis of intake interviews [Grift et al., 2009]. Data have been collected throughout various years (1999 through 2008). Unfortunately, the dataset does not contain every variable that could be explanatory for student vocational dropout as discussed in the literature overview in the previous chapter, such as proxies for academic motivation and facts about where students' parents work and what their (approximate) earnings are. The latter could have served as an indicator for socioeconomic status.

In order to make the dataset more suitable for our analysis, we have dropped all the observations not containing any data on the Big Five personality traits. Unfortunately, only a very select cohort of students appear to have been tested for these traits. The observations which lack these data do not serve any purpose for this research, because the Big Five are quintessential in order to properly answer the research question. Moreover, we have dropped or disabled all confidential variables such as student names, personal reasons for dropout, etc. Including any of these variables will not improve the quality of the analysis, but excluding them does warrant the confidentiality of our research.

The trimming of the dataset has vastly reduced the number of observations to 499, due to the substantial amount of missing values in the Big Five variables. Table 2.1 outlines this. However, this data amount is still large enough to allow for analysis through e.g. OLS (Ordinary Least Squares) or other regression techniques. Not all of our Big Five data remains normally distributed for $\alpha = 5\%$ as can be seen in table 2.2. The variable *altrusme* (agreeableness) is not normally distributed, for it is statistically insignificant in the skewness and kurtosis test. However, this is not an issue for the analysis because, following the Central Limit Theorem, the variable will still be approximately normally distributed nonetheless [University Of Alabama, 2014].

Big Five Missing Entries

VARIABLES	Missing	Total	Percent Missing
zorgvuldig (*conscientiousness*)	45,634	46,424	98.30
altrusme (*agreeableness*)	45,925	46,424	98.93
extravert (*extraversion*)	45,634	46,424	98.30
stabiel (*emotional stability*)	45,634	46,424	98.30
open (*openness*)	45,634	46,424	98.30

Table 2.1: Overview of Missing Entries of Big Five Variables

Skewness/Kurtosis Tests for Normality

VARIABLES	Obs	Pr(Skewness)	Pr(Kurtosis)	*joint*⟶ *adj* $chi^2(2)$	⟵*joint* $Prob > chi^2$
zorgvuldig	499	0.1597	0.0000	20.23	0.0000
altrusme	499	0.0912	0.1623	4.81	0.0904
extravert	499	0.8475	0.0014	9.25	0.0098
stabiel	499	0.0382	0.0000	36.21	0.0000
open	499	0.7344	0.0002	12.81	0.0017

Table 2.2: Skewness and Kurtosis Tests for Normality

Table 2.3 shows summary statistics for all of the Big Five variables in the dataset: *zorgvuldig* (conscientiousness), *altrusme* (agreeableness), *extravert* (extraversion), *stabiel* (emotional stability) and *open* (autonomy). The Likert-type scale in which all five Big Five variables have been measured is a notable detail here: Minimal values of one, maximum values of five and medians of three. On average in our sample, people were rated highest on *zorgvuldig* and lowest on *altrusme*. The standard errors fluctuate around one, with *altrusme* having the lowest standard deviation and *stabiel* the highest. In order to export most of the data tables and analysis results from the R programming environment to LaTeX, we have used the R package called *Stargazer* [Hlavac, 2014].

Big Five Summary Statistics

VARIABLES	mean	std dev	min	max	median	skewness	kurtosis
zorgvuldig	3.449	(1.097)	1	5	3	-0.153	2.321
altrusme	2.914	(0.845)	1	5	3	0.184	2.712
extravert	3.120	(1.082)	1	5	3	-0.0208	2.477
stabiel	3.315	(1.184)	1	5	3	-0.227	2.189
open	3.281	(1.030)	1	5	3	0.0367	2.408
Observations	499						

Table 2.3: Summary Statistics for Big Five Variables

In table 2.4, all relevant variables in the set are described. Roughly they fall into four categories: three binary variables, age, *specific* personality characteristics and *general* personality characteristics, i.e. the Big Five. Vrv_1 is the dropout rate of students in our sample. On average, a vast 62.7% dropped out at some point during the course of their education. It should be mentioned that the dropout rate in our dataset is technically *stopout*. It may be the case that students who quit their vocational education at the institution(s) providing our data enrolled at another institution for a different education track sometime. Nevertheless, compared to the Dutch national average of 5.7% this still is a huge difference. The data suggests that 41.5% of our sample is male, and the remaining 58.5% is female. Moreover, 24.6% of the students is of foreign origin, the others have Dutch roots. With respect to age, the sole continuous variable in our dataset, we see an average of 20.2 years old, with a minimum of 15 years old, and the oldest student enrolled in this educational institution was 62 years old. The variables that remain are specifications of certain personality characteristics. For example, the need for personal coaching and effort levels are documented. Refer to appendix A.1 for the summary statistics with translated variable names.

Dataset Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
vrv_1 (dropout=1)	499	0.627	0.484	0	1
geslachtBin (male=1)	499	0.415	0.493	0	1
allochtoonBin (foreign=1)	499	0.246	0.431	0	1
age_opleiding	499	20.152	7.817	15	62
capaciteitkunnen	499	3.168	0.885	1	5
leerintelligentie	499	3.269	0.992	1	5
numeriekeaanleg	499	3.126	1.048	1	5
verbaleaanleg	499	3.184	1.164	1	5
logischredeneren	499	3.253	0.974	1	5
redeneervermogen	499	3.060	1.105	1	5
evaluatievermogen	499	3.429	1.103	1	5
ruimtelijkinzicht	499	3.122	1.073	1	5
persoonlijkheidwillen	499	3.172	0.639	2	5
inzet	499	3.156	0.771	1	5
conscintieusordelijk	499	3.395	0.889	1	5
behoefteaanbegeleiding	499	2.735	0.735	1	5
hoopopsucces	499	3.255	1.065	1	5
angstvoorfalen	499	2.611	1.120	1	5
behoefteaanpersoonlijkestimuleri	499	2.908	1.070	1	5
enthousiasme	499	2.687	1.021	1	5
degelijkheid	499	3.455	0.965	1	5
stabiel	499	3.315	1.184	1	5
extravert	499	3.120	1.082	1	5
altrusme	499	2.914	0.845	1	5
zorgvuldig	499	3.449	1.097	1	5
open	499	3.281	1.030	1	5

Table 2.4: Dataset Summary Statistics

2.2 Method

This section will provide an overview and an explanation of the methods that will be used in our research.

2.2.1 Exploratory Data Analysis

First off, to provide deeper knowledge of the data itself and of the patterns in which it is structured, we will perform *EDA* (Exploratory Data Analysis). Through the use of various statistical visualisation techniques the data of the Big Five will be explored. Moreover, we will explore other major contributing variables suggested by the theory in the previous chapter, such as gender, age and foreign origin.

2.2.2 Logistic Regression Analysis

In order to draw causal conclusion from our data, logistic regression analysis will be applied. The dependent variable modelled is the dichotomous factor dropout rate, and the explanatory variables are the Big Five personality traits: extraversion, agreeableness, conscientiousness, emotional stability and autonomy. Additional control variables will be added and are necessary to overcome omitted variable bias and to pursue the investigation of theoretical relations.

Logistic regression is the best choice for the estimation of these models, as we aim to explain a binary dependent variable. All variables, except for age, are discrete in nature: Every one of the Big Five variables is ordinal and the remaining control variables are dichotomous. Were we to use Ordinary Least Squares (OLS) to estimate the models, we would violate the assumption of linearity [Zeigler-Hill, 2013]. Logistic regression tackles this problem by applying a logarithmic transformation on the outcome variable as shown in equation 2.1, assuming Bernouilli probability distribution. This allows for the modelling of a non-linear relationship in a linear way.

The observant reader will wonder why we have chosen a logit model over a probit model. Logit converts the variables to a logarithmic scale (equation 2.1), whereas probit leaves the data normally distributed. As a result, logit allows for the observation of odds ratios and probit does not. The estimation results of both logit and probit models are mainly similar for large samples, however when used on a smaller sample distinct differences in estimates can be observed. The choice between probit and logit should be based on the type of interpretation desired, or based on maximum likelihood. For our research, we consider logit the most suitable. There is no shocking difference in likelihood between either model, however logit does provide the benefit of odds ratios. This allows us to make statements about likelihood even before interpreting the models' marginal effects.

$$\pi_i = logit^{-1}(\alpha + \beta X) = \frac{e^{\alpha+\beta X}}{1 + e^{\alpha+\beta X}} \;\therefore\; \alpha + \beta X = logit(\pi_i) = ln\frac{\pi_i}{1-\pi_i} \tag{2.1}$$

Chapter 3

Analysis Results

"Act, not for the results, but for the action. See, not through your beliefs, but through the eyes. Think before you do and do because it is right."

— Debasish Mridha

In this chapter all results from the methods in the previous chapter will be listed and discussed. The R programming language has been used to perform all analysis, visualisation and data exploration in this paper [R Development Core Team, 2008].

3.1 Exploratory Data Analysis

In order to get a more firm grasp on the data at hand (being the Big Five characteristics and important control variables gender and foreign origin) and the way in which it is structured and potentially (cor)related, we have applied EDA through the use of a multitude of numerical and graphical statistical visualisation methods. These techniques vary from simple boxplots to multilateral scatterplots, featuring dynamic colouring, jittering and alpha blending. Dynamic colouring is the process of applying a specific color to observations of a certain observed variable. Jittering is a technique to scatter observations around a single point in order to visualise the quantity of observations on that point; this would not be visible if all were plotted on the same position. Building on the former technique, alpha blending makes plotted points partly transparent in order to counter overplotting. The more points overlap each other, the more opaque the points will become. Opaqueness therefore signals vast observation density. All visualisations have been obtained using the *R programming language* in conjunction with the *ggplot2* package [Wickham, 2009].

3.1.1 Exploring Vocational Dropout

The following subsection will explore vocational dropout, the paramount variable of this paper. There is a grand total of 499 students in our sample. As shown in the previous sections of this chapter they can be categorised in two subgroups based on our control variables: gender and foreign origin. Therefore, to keeps things clear and organised, the distribution of students in our sample is visualised in figure 3.1. This visualisation displays the absolute number of students per subgroup, the percentage relative to the total amount of students in the sample and whether or not they drop out of vocational education.

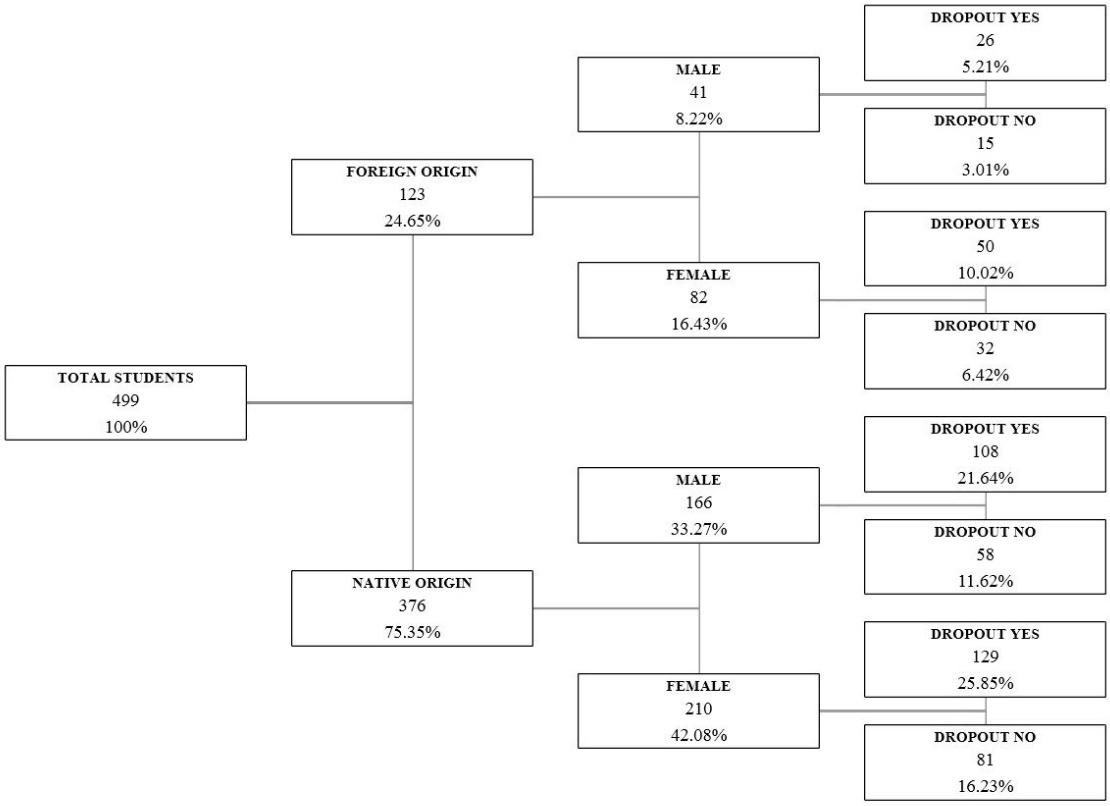

Figure 3.1: Distribution of Students in the Sample

Some interesting points can be derived from the graphic above. The dropout rate for male students of foreign origin within our sample is $\frac{26}{41} * 100 = 63.41\%$, while the dropout rate of male students of native origin is $\frac{108}{166} * 100 = 65.06\%$. This is relatively similar, though students of native origin drop out slightly more. The dropout rate for female students of foreign origin within our sample is $\frac{50}{82} * 100 = 60.98\%$, whereas the dropout rate of female students of native origin is $\frac{129}{210} * 100 = 61.43\%$: Also very similar, but again natively originated female students drop out marginally more. Overall, it seems that in our sample the dropout is somewhat higher among the female students, regardless of their origin. The dropout among gender and origin can be furtherly displayed and clarified using scatterplots as done below in figure 3.2.

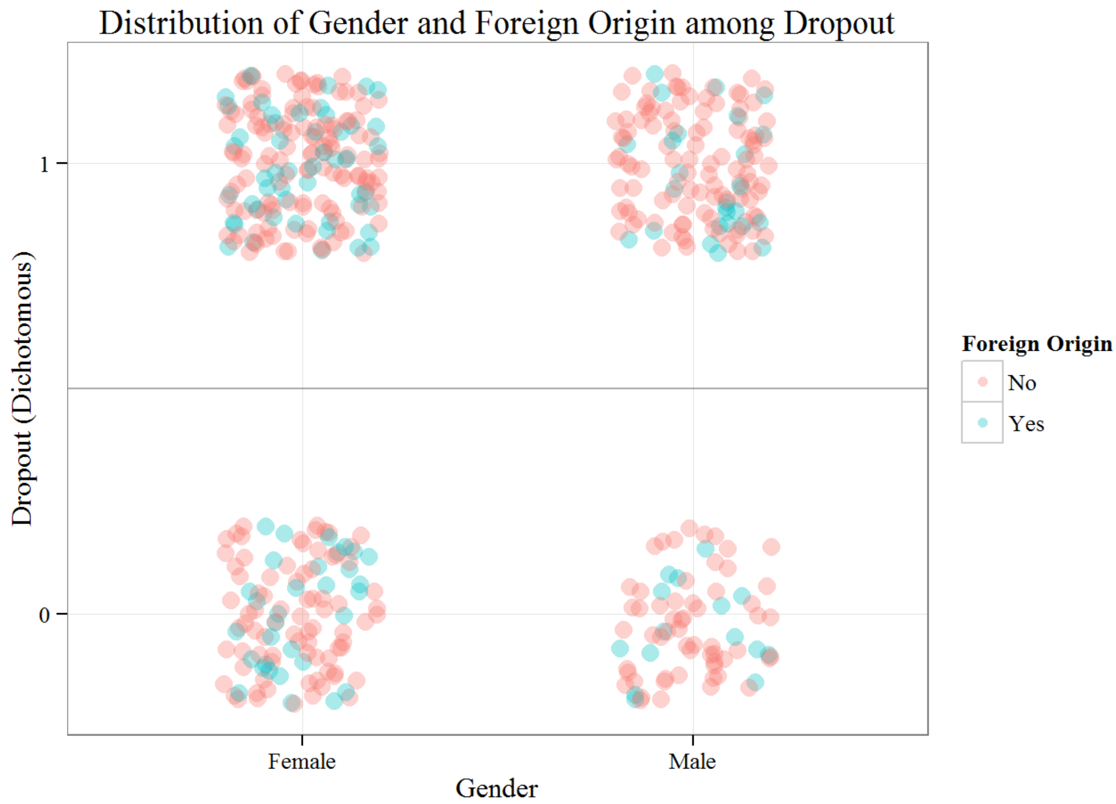

Figure 3.2: Gender, Origin and Associated Dropout among Students

Student Dropout among Gender and Origin

In the figure above one can clearly distinguish the trend in our sample which has already been described using figure 3.1. Large amounts of both female and male students of foreign and native origin drop out of vocational education (denoted by the 1 on the y-axis).

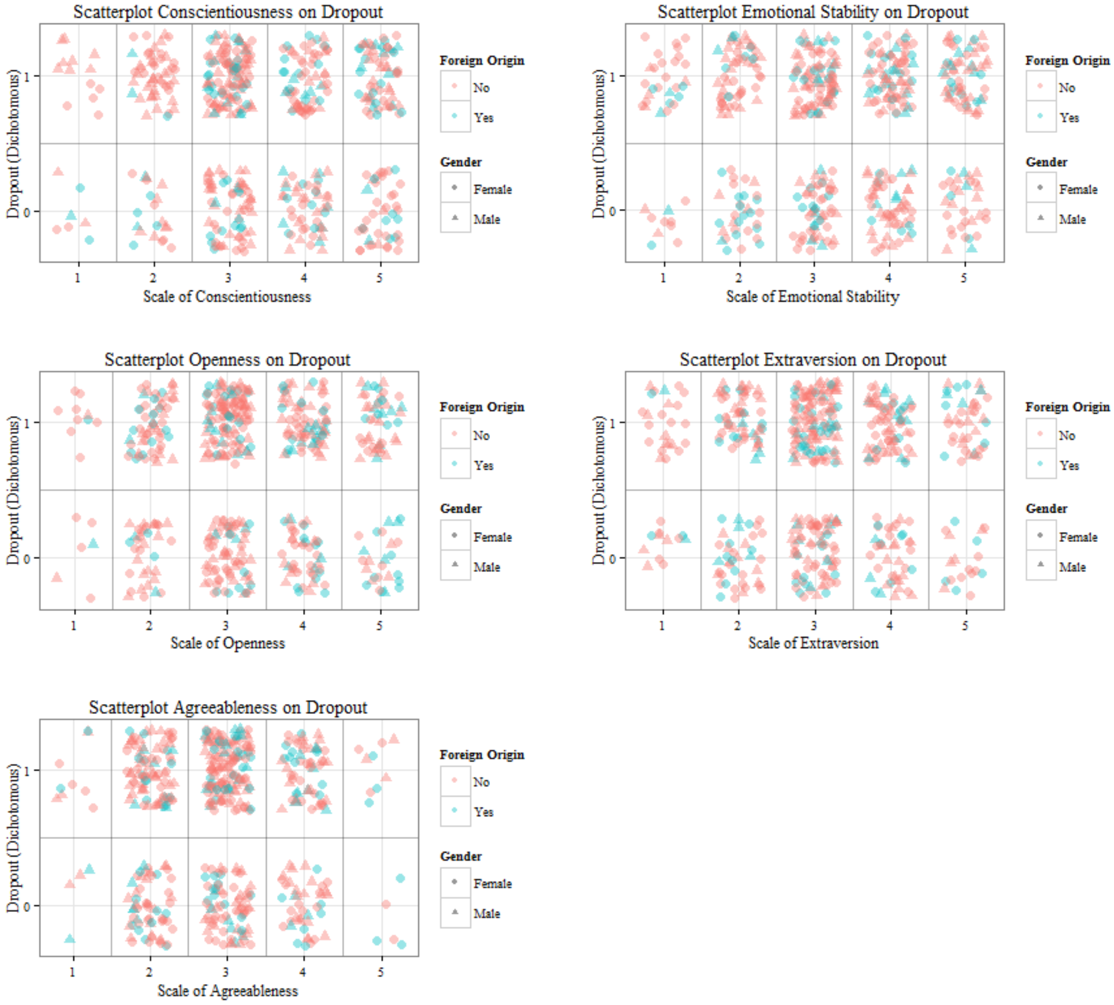

Figure 3.3: Dropout and the Big Five by Gender and Foreign Origin

Dropout and the Big Five Personality Traits

In figure 3.3 scatterplots are shown displaying the relation between each Big Five variable and dropout. As the trend among the plots is roughly similar, we will treat and interpret the various plots as though it is one. Although the observations seem dispersed, they all refer to one and the same intersection in the graph: This is the result of jittering, so that we can observe the density of the observations in each cluster more clearly.

At a first glance, there appears to be no significant visual relationship between dropout and any of the Big Five personality traits. Considering the plot for conscientiousness, it can be observed that dropout does not seem to decrease with increasing scores of conscientiousness. Intuitively, the same is true for the opposite. Whereas we hypothesised low scores of conscientiousness to correlate with high dropout rates, this is in no sense supported by the plot. The same seems to be true for the other Big Five. A second general trend that can be observed in the plots is a surprisingly large absolute, as well as relative, amount of dropouts. To illustrate this, refer to the emotional stability and extraversion plots. In these plots, one can observe that vastly more observations occur for dropout being equal to one, than for dropout being equal to zero. Numerically illustrated: from the 499 observations a grand total of 313 students dropped out. In percentage terms this corresponds to $\frac{313}{499} * 100 = 62.73\%$. In a section further on in this chapter, we will explore the aforementioned causality more in-depth using regression analysis.

3.1.2 Exploring the Big Five and Gender

This section will explore the difference in scores among the Big Five personality traits among males and females in our sample. It will do so using boxplots and numerical summaries. These are displayed in figure 3.4 and table 3.1. However, first and foremost we note that our sample has 207 male students and 292 female students. It may be possible that the larger proportion of female students elevates or decreases some Big Five scores, when taking the average of both males and females.

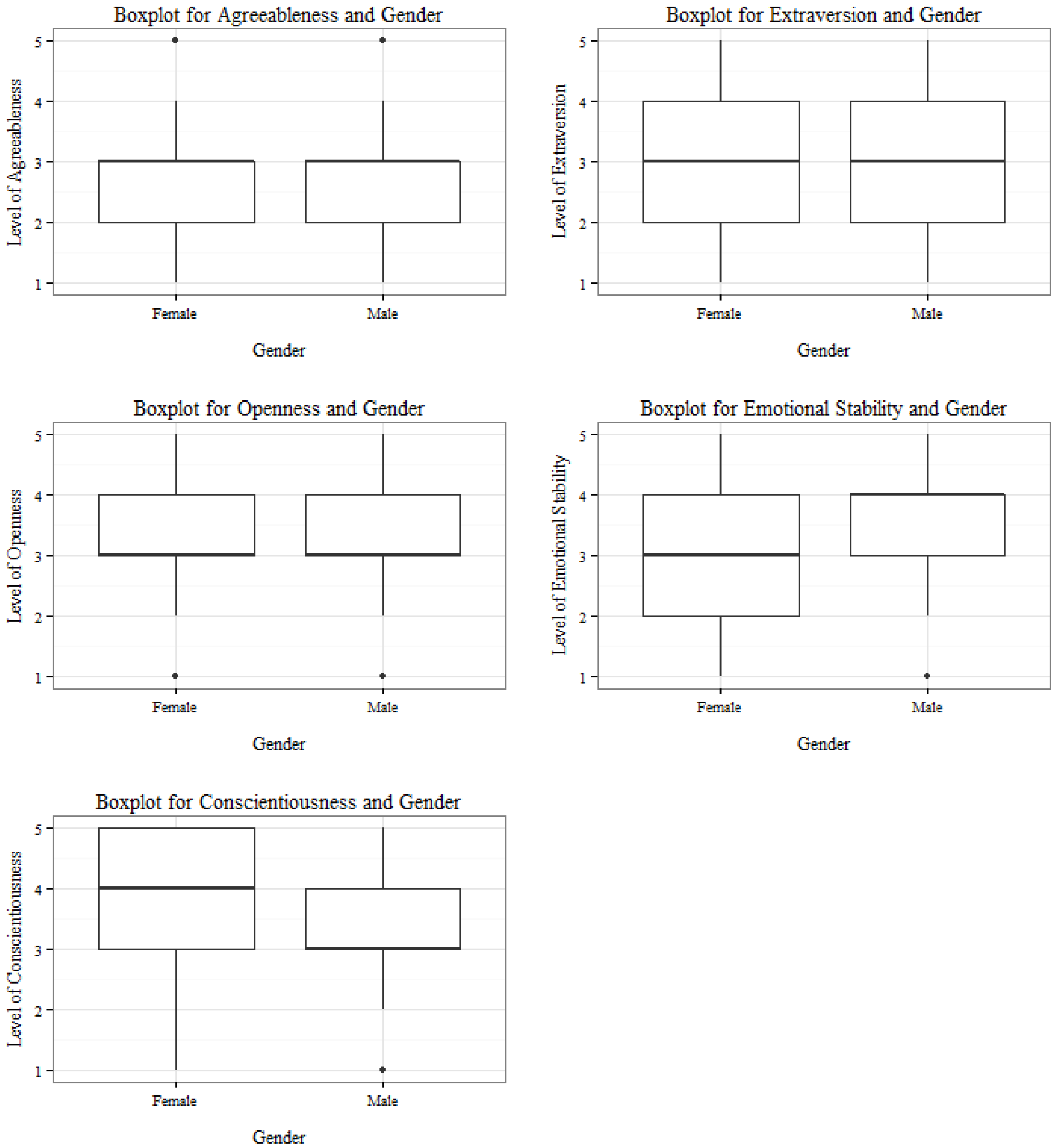

Figure 3.4: Boxplots for the Big Five and Gender

Numerical Summaries

Conscientiousness	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Female	1.000	3.000	4.000	3.606	5.000	5.000
Male	1.000	3.000	4.000	3.227	4.000	5.000
Emotional Stability	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Female	1.000	2.000	3.000	3.147	4.000	5.000
Male	1.000	3.000	4.000	3.551	4.000	5.000
Openness	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Female	1.000	3.000	3.000	3.209	4.000	5.000
Male	1.000	3.000	3.000	3.382	4.000	5.000
Extraversion	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Female	1.000	2.000	3.000	3.127	4.000	5.000
Male	1.000	2.000	3.000	3.111	4.000	5.000
Agreeableness	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Female	1.000	2.000	3.000	2.938	3.000	5.000
Male	1.000	2.000	3.000	2.879	3.000	5.000

Table 3.1: Numerical Summaries for the Big Five and Gender

Conscientiousness

We can observe that in our sample females have higher mean and median levels of conscientiousness. This is on par with the theories we discussed earlier. It can also be observed that females have a larger interquartile range in conscientiousness compared to males. The spread of scores for females is divided across the entire interquartile range plus the whiskers of possible scores, whereas for males a score of one is considered an outlier. Therefore, we can conclude that in our sample females score higher on conscientiousness and males score lower but more consistent.

Emotional Stability

Next, it is shown that females score on average lower than males on emotional stability. As for the median score, males also score one point higher than females. Females also show a larger overall interquartile range than males. Just like in the previous plot, a score of one is not uncommon for females. For males, again, this is considered an outlier. We can conclude that male students in our sample score higher and more consistent on emotional stability than female students.

Openness

On openness males and females score more or less equal. When referring to the numerical summary however, it can be observed that males score on average slightly higher on the scale of openness, as shown by the means. The score of one on openness is rare and can be considered an outlier for both sexes, as shown by the boxplot.

Extraversion

The scores on the extraversion scale for males and females are even more similar than for openness. The boxplots appear identical, however the numerical summary suggests that on average females score slightly higher on the extraversion scale in our sample. Another meaningful thing to point out is that for extraversion there are no outlier scores in our sample.

Agreeableness

The last big five characteristic, agreeableness, once again seems rather similar between the two sexes when looking merely at the boxplot. However, the numerical summary gives the insight that females are slightly superior to males on the scale of agreeableness. Noteworthy is that the highest score, five, is rare and considered an outlier for both genders.

3.1.3 Exploring the Big Five and Foreign Origin

Here, the difference in scores among the Big Five personality traits between foreign and native students will be reviewed. Again it must be stressed that the number of native and foreign students in the sample vastly differs: There are 376 native students but only 123 students of foreign origin. One must keep in mind that this discrepancy may distort the results. Refer to figure 3.5 for the boxplots and to table 3.2 for the numerics.

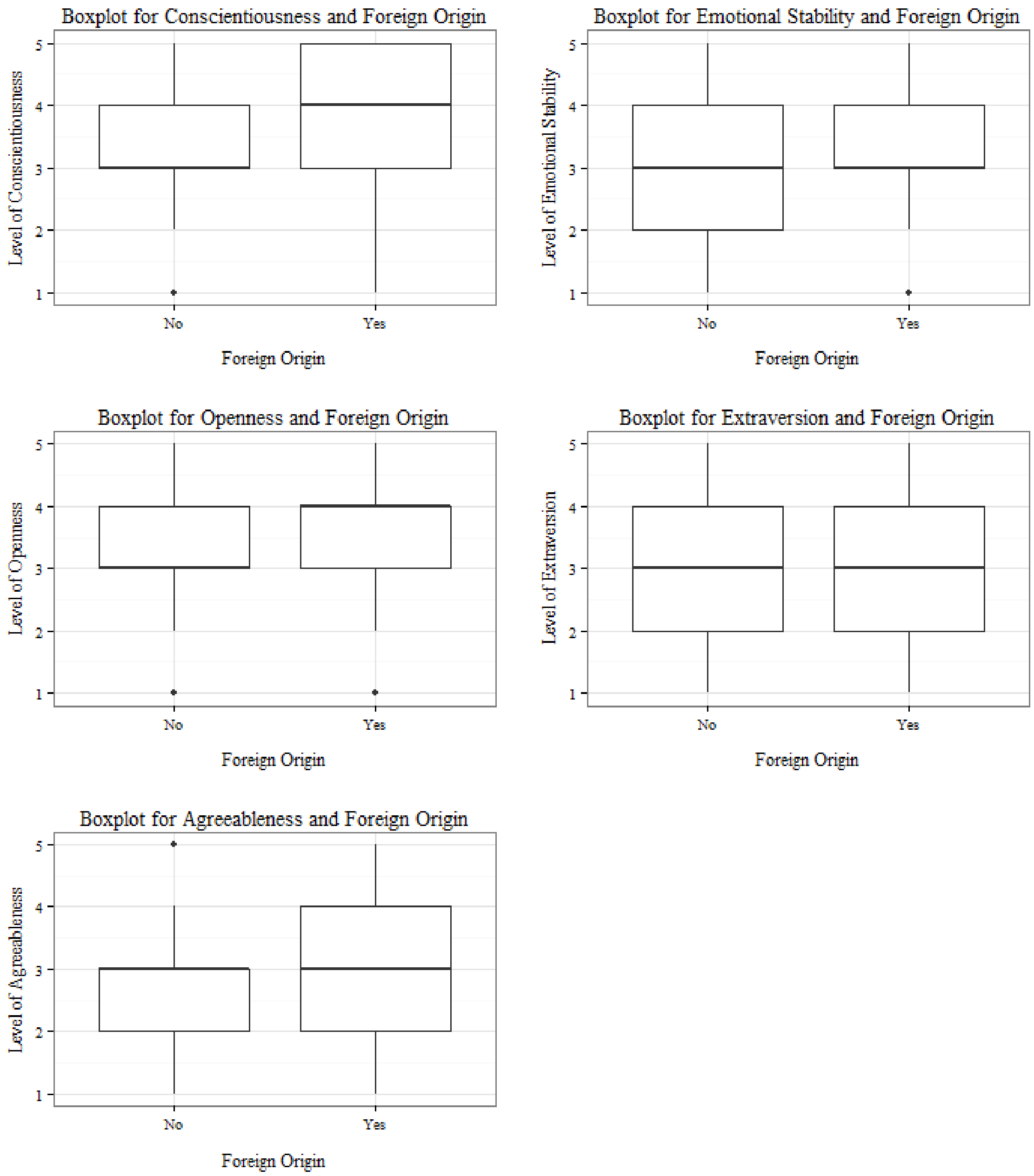

Figure 3.5: Boxplots for the Big Five and Foreign Origin

Numerical Summaries

Conscientiousness	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Native Origin	1.000	3.000	3.000	3.367	4.000	5.000
Foreign Origin	1.000	3.000	4.000	3.699	5.000	5.000
Emotional Stability	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Native Origin	1.000	2.000	3.000	3.303	4.000	5.000
Foreign Origin	1.000	3.000	3.000	3.350	4.000	5.000
Openness	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Native Origin	1.000	3.000	3.000	3.186	4.000	5.000
Foreign Origin	1.000	3.000	4.000	3.569	4.000	5.000
Extraversion	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Native Origin	1.000	2.000	3.000	3.104	4.000	5.000
Foreign Origin	1.000	2.000	3.000	3.171	4.000	5.000
Agreeableness	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Native Origin	1.000	2.000	3.000	2.891	3.000	5.000
Foreign Origin	1.000	2.000	3.000	2.984	4.000	5.000

Table 3.2: Numerical Summaries for the Big Five and Foreign Origin

Conscientiousness

Referring to the boxplot and numerical summary for conscientiousness, it is immediately clear that the students of foreign origin in the sample score higher on both mean and median scores. The spread of their scores is also broader as can be seen by the boxplot's whiskers and interquartile range. The students of native origin score more consistently, but also score lower.

Emotional Stability

The mean scores of emotional stability tend to be slightly higher for students of foreign origin. Moreover, the spread of scores is lower and a score of one is considered an outlier. Students of native origin display scores occurring across the entire score spectrum.

Openness

In terms of openness the students of foreign origin in our sample score higher on mean and median scores. The spread of scores among the quartiles is equal between both groups. Furthermore, a score of one on openness is considered an outlier in both groups.

Extraversion

The boxplot for extraversion shows a lot of similarity between both groups. The numerical summary suggests a small difference in mean extraversion score in favour of the students of foreign origin.

Agreeableness

Lastly, the score on agreeableness shows a higher mean and a broader spread among the quantiles and whiskers for students of foreign origin. Striking is that the maximum score of agreeableness is considered an outlier for students of native origin.

3.1.4 Remarks

From the exploration of the performance of students among the Big Five personality traits it can be concluded that the students of foreign origin in our sample performed more or less better. The observant reader will note that this goes directly against our hypothesis as well as against the findings of academic literature mentioned earlier. We therefore expect that the students of foreign origin in our sample are relatively well performing students.

3.2 Logistic Regression Analysis

This section will describe the procedure by which we have built, estimated, calibrated and diagnosed our logistic regression models.

3.2.1 Model Structure

In order to determine the causal relation between dropout and the Big Five, the first step we have undertaken is to produce a base logit model where dropout depends on the Big Five, controlled for age. See logit model 3.1 below, where vrv_1 is the dichotomous dependent variable for dropout. This model shall function as the base model upon which we build further.

$$\begin{aligned} ln(\frac{\pi_i}{1-\pi_i}) = \widehat{vrv\_1}_i = \widehat{\beta}_0 + \widehat{\beta}_1 \cdot cons_i + \widehat{\beta}_2 \cdot altr_i + \widehat{\beta}_3 \cdot open_i \\ + \widehat{\beta}_4 \cdot stab_i + \widehat{\beta}_5 \cdot extr_i + \widehat{\beta}_6 \cdot age_i + \varepsilon_i \end{aligned} \tag{3.1}$$

The results of the estimation of this model are shown in table 3.3, column one. However, such a base model suffers from scarcity in explanatory variables. In technical terms, omitted variable bias (OVB) is likely to occur in the first model because important control variables, as suggested by existing literature, are left out. What ensues is a trade-off between eliminating, or reducing omitted variable bias, while not letting multicollinearity come into play. In column two, a more specific model is presented, where not only age is controlled for, but also foreign origin and overall learning intelligence. Age, foreign origin and learning intelligence are necessary control variables, because learning intelligence is used as a proxy for the effects of human capital theory, whereas foreign origin should cover the consequences of social capital. Additionally, interaction effects between conscientiousness, openness and emotional stability with foreign origin will be tested. These interaction effects are of serious importance, for both theory and our EDA identify different Big Five magnitudes depending whether a student is of foreign origin or not. To see the functional form of this model, refer to equation 3.2 below.

$$\begin{aligned} ln(\frac{\pi_i}{1-\pi_i}) = \widehat{vrv\_1}_i = \widehat{\beta}_0 + \widehat{\beta}_1 \cdot cons_i + \widehat{\beta}_2 \cdot altr_i + \widehat{\beta}_3 \cdot open_i + \widehat{\beta}_4 \cdot stab_i \\ + \widehat{\beta}_5 \cdot extr_i + \widehat{\beta}_6 \cdot age_i + \widehat{\beta}_7 \cdot foreign_i \\ + \widehat{\beta}_8 \cdot intell_i + \widehat{\beta}_9 \cdot cons_i \cdot foreign_i \\ + \widehat{\beta_{10}} \cdot open_i \cdot foreign_i + \widehat{\beta_{11}} \cdot stabiel_i \cdot foreign_i + \varepsilon_i \end{aligned} \tag{3.2}$$

3.2.2 Model Interpretations

This subsection will interpret our estimated logistic models in two ways: Using odds ratios and marginal effects. Researchers of the social sciences are usually more familiar with odds ratios, while economists and econometricians tend to prefer marginal effects. With us belonging to the latter group, but our research treading on both social as well as economic ground, we have decided to include both metrics. Note that the section on odds ratios includes full interpretations and explanations, whereas the section on marginal effects is considered supplementary and therefore only contains interpretations.

Odds Ratios

Results for the logistic regressions we have run for model 3.1 and 3.2 can be found in table 3.3. In this table variable names are listed in Dutch to maintain consistency throughout the paper. In a previous chapter, we have discussed and translated most of these variables in detail. Nonetheless, a fully English translation of this results table is available in the appendix in table B.1.

Consider the base model in the first column of the table with significant effects for *zorgvuldig* (conscientiousness) and *age* on dropout. The odds ratios reported in the table convey the likelihood of an event. Thus, when a student is rated one point higher on the scale of conscientiousness, he or she is 0.813 times *less* likely to drop out of vocational education than his or her fellow students, ceteris paribus on the other exogenous variables. The base model also shows a significant result for the age of the specific student: A student is 1.057 times *more* likely to drop out of vocational education when he or she is one year older than his or her peers, ceteris paribus on the other covariates.

Now consider model two in the second column of table 3.3. Please note that every interpreted effect discussed here is ceteris paribus on the other variables and relative compared to the other observations (i.e. other students). First of all, the variable *zorgvuldig* (conscientiousness): A student rated one point higher on the scale of conscientiousness is 0.699 times *less* likely to drop out of vocational education. This effect is much stronger than in the first model. The age of a student is also a slightly stronger factor in the second model: When a student is one year older than his or her peers, he or she is 1.059 times *more* likely to drop out of vocational education. A new significant effect in the second model is *leerintelligentie* (educational intelligence). For every one point higher scored on the scale of educational intelligence, a student is 1.227 times *more* likely to drop out of vocational education. Initially, this seems an obscure result. However, existing research and literature suggest that students with high scores on educational intelligence might be likely to leave vocational education for higher types of education, such as a University of Applied Sciences.

Furthermore, our analysis pinpoints strong interaction effects between students of foreign origin and *zorgvuldig* (conscientiousness), *open* (openness) and *stabiel* (emotional stability). The principal effect of foreign origin on vocational dropout rates is negative but (only slightly) statistically insignificant for $p < 0.1$. Therefore, if we interpret it (although we treat it cautiously), it suggests that students of foreign origin are 0.209 times *less* likely to drop out. This corresponds with what we have shown in our EDA earlier on. As for the significant interaction effects: A student of foreign origin is 2.093 times *more* likely to drop out for each point increase in conscientiousness, 0.532 times *less* likely to drop out for every point increase in openness and 1.497 times *more* likely to drop out for each additional point in emotional stability. Given the magnitude of these findings, as well as the fact that these phenomena are not mentioned in the existing literature, we suggest not to focus too much on their value and we recommend further investigation in different samples. As we argued earlier in the section on EDA, our data sample is likely biased. Students of foreign origin, while scarce in the sample, score relatively high on all scales compared to students of native origin.

Altogether, when looking at all the results of the Big Five in model two, it is notable that emotional stability and conscientiousness show *negative* effects on dropout (i.e. reduce the odds of it occurring, ceteris paribus on the other covariates). Extraversion, openness and agreeableness, on the other hand, seem to display a *positive* effect on dropout (i.e. increase the odds of it occurring, ceteris paribus on the other covariates). Except for agreeableness, which is theoretically supposed to reduce dropout, these results are in line with our theoretical framework. Further examination tells us the odds ratio found for agreeableness is 1.056, with a standard error of 0.127. Insignificance is the key here: The estimator is found to be almost equal to one, which is the threshold value for the effect's sign. An estimator larger than one indicates a positive relation, is the coefficient less than one, it appears to be a negative relation. In other words, we are very close to proving the expected relation, but in this case, even a little error can already flip the sign of the effect. Openness might seems out of place at first sight. However, low levels of openness are, as literature suggests, associated with students who seek to outperform others and might therefore not soon abandon there studies. Conversely, students showing high levels of openness could therefore be more prone to drop out. Lastly, age has been found to display a small but *positive* effect on dropout. The magnitude corresponds with the findings of [Grift et al., 2009], as mentioned earlier in the theoretical framework. The direction of the relation, however, is exactly the opposite. A possible explanation can be found in the fact that, although we made use of the same dataset, this research has only been able to employ 499 observations after filtering the data for student observations containing data on the Big Five.

Table 3.3: Regression Results (Odds Ratios)

	Dependent variable:	
	Dropout (Dichotomous)	
	(1)	(2)
stabiel	0.964	0.855
	(0.087)	(0.103)
open	0.990	1.104
	(0.103)	(0.123)
zorgvuldig	0.813*	0.699**
	(0.096)	(0.112)
extravert	1.083	1.060
	(0.105)	(0.108)
altrusme	1.012	1.056
	(0.122)	(0.127)
allochtoonBinYes		0.209
		(1.044)
age_opleiding	1.057***	1.059***
	(0.016)	(0.017)
leerintelligentie		1.227*
		(0.113)
zorgvuldig:allochtoonBinYes		2.093**
		(0.250)
open:allochtoonBinYes		0.532**
		(0.242)
stabiel:allochtoonBinYes		1.497*
		(0.214)
Constant	1.020	0.797
	(0.531)	(0.682)
Observations	499	499
Log Likelihood	-319.901	-306.682
Akaike Inf. Crit.	653.802	637.364
Nagelkerke R^2	0.052	0.119
Note:	*p<0.1; **p<0.05; ***p<0.01	

Marginal Effects

The interpretation of the statistically significant marginal effects of the logit models will be described here. All interpreted effects are ceteris paribus on the other covariates. A comprehensive overview of all marginal effects is available in table 3.4 of which, again, a translation is available in appendix table C.1. As mentioned earlier, refer to the odds ratio interpretation section above for theoretical explanations of the interpreted effects.

The model in the first column of table 3.4 is the base model and only shows significant effects for conscientiousness and the student's age. A one point increase in conscientiousness decreases dropout likelihood by 4.8%. As for age: A one year increase in the age of a student makes him or her 1.3% more likely to drop out of vocational education.

With respect to the more elaborate model in column two, we observe that a one point increase in a student's conscientiousness decreases the probability of him/her dropping out by 8.2%. Equivalently, a student that is one year older than the other students is 1.3% more likely to drop out. A one point increase in the educational intelligence of a student increases dropout likelihood by 4.7%. Although insignificant, the main effect of a student's origin on dropout is negative. Being of foreign origin decreases the dropout probability by a vast 37%, which corresponds to the findings of the earlier EDA section. As for the interaction effects: A student of foreign origin is 16.9% more likely to drop out for each point increase in conscientiousness, 14.5% less likely to drop out for each point increase in openness and 9.3% more likely to drop out for each point increase in emotional stability.

Table 3.4: Regression Results (Marginal Effects)

	Dependent variable:	
	Dropout (Dichotomous)	
	(1)	(2)
stabiel	-0.008	-0.036
	(0.020)	(0.024)
open	-0.002	0.023
	(0.024)	(0.028)
zorgvuldig	-0.048^{*}	-0.082^{**}
	(0.022)	(0.026)
extravert	0.018	0.013
	(0.024)	(0.025)
altrusme	0.003	0.013
	(0.028)	(0.029)
allochtoonBinYes		-0.369
		(0.231)
age_opleiding	0.013^{***}	0.013^{***}
	(0.004)	(0.004)
leerintelligentie		0.047^{*}
		(0.026)
zorgvuldig:allochtoonBinYes		0.169^{**}
		(0.057)
open:allochtoonBinYes		-0.145^{**}
		(0.055)
stabiel:allochtoonBinYes		0.093^{*}
		(0.049)
Observations	499	499
Log Likelihood	-319.901	-306.682
Akaike Inf. Crit.	653.802	637.364
Nagelkerke R^2	0.052	0.119
Note:	*p<0.1; **p<0.05; ***p<0.01	

3.2.3 Model Diagnostics

In order to verify the validity and fit of the models used in our research, we have applied some common techniques to diagnose the models. These diagnostics are displayed comprehensively in Appendix C, but will be briefly elucidated here.

The plots of the model's residuals versus fitted values, shown in figure D.2 and D.4, show that model two is generally a better fit for our data. This is underlined by the Nagelkerke R^2 values shown in table 3.3. An important thing to note, is that residual versus fitted plots are not as influential in making model fit decisions for logistic models as they are for linear (OLS) models. Therefore, we have also plotted Cook's distance for both regression models. These plots can be observed in figure D.1 and D.3. To properly interpret these plots, one has to take note of the Cook's Distance threshold calculated in equation 3.3. In the plots in the appendix the threshold levels are denoted by the dashed red lines. It can be observed that for the second model Cook's distance is relatively smaller compared to the first model. There are also less influencing points for the second model (i.e. points that exceed the red threshold line).

$$D = \frac{4}{N-k-1} \quad \longrightarrow \quad \begin{cases} D_{model_1} & \frac{4}{499-6-1} & \approx 0.0081 \\ D_{model_2} & \frac{4}{499-11-1} & \approx 0.0082 \end{cases} \tag{3.3}$$

Finally, to check our model for multicollinearity, we examine the variance inflation factors (VIF) of the model. Tables containing the VIF values are shown in table D.1 and D.2. We test for multicollinearity by checking if the conditions as stated in equation 3.4 are true. The last column of the table shows the results of the test. Model one appears to be completely without multicollinearity. In model two, only the variables interacting with the dummy variable *foreign origin* display signs of multicollinearity. However, because of the dichotomous nature of this variable, this can be safely ignored.

$$\begin{aligned} H_0 &: \quad No\ multicollinearity \\ H_A &: \quad Multicollinearity \end{aligned}$$

$$\sqrt{VIF} > 2 \quad \longrightarrow \quad Reject\ H_0 \tag{3.4}$$

We can conclude that the second model is a better fit for the data at hand: It is a significant improvement to our base model and it is without an influencing presence of multicollinearity. However, the second model still has a respectable amount of influencing observations, as shown by Cook's distance. It may be possible to improve the model further and make it more accurate, yet with the severely limited documentation on the dataset at our disposal, as well as the pressing timeframe of this research, this venture currently lies beyond our options.

Chapter 4

Discussion Results

"It's the economy, stupid!"

— James Carville

This chapter will discuss the effects of dropout in Dutch vocational education theoretically in a context of economic implications. We also attempt to extend this analysis to the economic consequences as a result of dropout in higher education. To do so, we employ human and social capital theory as presented in the theoretical framework.

4.1 Dropout in Vocational Education

Although the analysis of many researchers (cf. [Catterall, 1987]; [Hankivsky, 2008]; [King, 1999] and [De Witte, 2014]) has produced estimates of costs of dropout for both the society as a whole and the individual in particular, few of them approached this problem purely theoretically. In what ways is a regional or national (macro-)economy actually influenced by such a micro phenomenon as dropout behaviour? Analysing this problem comprises two kinds of costs: *tangible* and *intangible* costs [Hankivsky, 2008]. The former represents direct costs related to dropout, such as tuition fees and the institution's cost of providing education. The most vital and largest aspect of tangible costs however, is foregone earnings. Consequently, it is there that the focus will lie in this section. Foregone earnings is income which could have been realised if education were retained. Intangible costs, on the other hand, are harder to attach values to. The costs follow from emotional distress and reduced quality of life. These seem rather hard to define and therefore require attention in a separate research.

One way in which education relates to growth is through labor productivity [Hanushek and Woessmann, 2007]. In general, vocational education and training prepares students for relatively low-skilled occupations such as security guard, baker or construction worker. Following equation 1.2, dropping out nullifies the level of investment in future human capital. With that in mind, leaving vocational education reduces the amount of human capital (h) that would otherwise have increased the value of labour (L), as shown in equation 1.1 from chapter 1.3. As a result, the economy suffers from foregone earnings caused by the dropout.

As Mincer [1981] proposes, the larger the amount of physical capital, the larger the benefits of additional human capital. This also works vice versa: a high level of human capital facilitates growth of physical capital, as it can be more skillfully employed [Mincer, 1981]. Following his reasoning, higher dropout rates as a macro phenomenon lower the degree of human capital in the production function in 1.1. Consequently, this not only reduces the marginal benefit of physical capital, it also dilutes the need for more physical capital. Hence a low human capital ratio (h) is a two-fold deteriorating factor to growth [Mincer, 1981].

4.2 Dropout in Higher Education

There are multiple mechanisms that constitute the effect of higher education on economic development: While vocational schooling is mostly associated with labour productivity, attaining higher education is beneficiary in terms of (a) more innovative ability and (b) improved capacity to deal with new information. Dropping out of higher education therefore has a larger economic magnitude.

Effect (a) refers to the larger toolbox higher-educated individuals have built up. With it, one can create, shape and develop ideas more effectively, as a result of the enjoyed education. Hence it implies possible creation of new capital. Effect (b) works through better knowledge accumulation. It states that education makes an economy able to keep up with the rapid technological developments of the present time and thereby empowers economic growth [Hanushek and Woessmann, 2007]. In terms of the production function 1.1, dropping out of higher education reduces h more than does dropping out of vocational education and thereby is likely to have a *larger* negative effect on economic prosperity.

Overall, empirical evidence is present in abundance for this causal relation between human capital and economic growth. Whether education is measured as adult literacy rates, school enrollment ratios or average years of schooling, appears irrelevant: The empirical literature seems to agree that the quantity of education matters. For example, refer to [Benhabib and Spiegel, 1994], [Barro, 2001] and [Mankiw et al., 1992].

However, dropout alone cannot be held responsible for the entirety of decrease in human capital. This incorrectly assumes that human capital can solely be acquired by attending an educational institution, whereas it can be obtained in a social context (e.g. through parents and the direct environment) as well. This has been thoroughly explained through the social capital concept in section 1.5 and has been illustrated by empirical proof Coller has found in earlier research. However, assuming that attending school is the *primary* source of human capital is plausible, see also Hanushek and Woessmann [2007]. Social capital is a stronger determinant in affecting human capital for the higher education segment than it is for the vocational education segment.

Chapter 5

Conclusions and Remarks

"It's more fun to arrive at a conclusion than to justify it."

— Malcolm Forbes

The research in this paper has been fruitful, however several limitations have been encountered. The findings that our research has yielded will be presented, summarised and evaluated here. Furthermore, we will discuss the aspects that held our research back and, therefore, can be improved upon in further research.

5.1 Conclusions

Hitherto, this paper has shown what the psychological and personal causes of dropout behaviour in Dutch vocational education are, and illustrated what the economic implications of dropout are for a regional or national economy. The existing literature already shed some light on the relationship between dropout and students' personalities and it suggested that in particular conscientiousness and agreeableness are important determinants of study retention. Although openness and extraversion did influence academic *motivation*, significant direct effects on dropout were not found. This largely corresponds with our results, which lead us to believe that also in our model conscientiousness is the main Big Five determinant of dropout behaviour. On top of that, age and general learning intelligence appear to be significant major contributing factors.

Through the use of exploratory data analysis, we have visualised key aspects of our dataset. Apart from further familiarising the reader with the data and its structure, this served to inspect the data for relationships as dictated by theories from the theoretical framework. Through EDA, several interaction effects for students of foreign origin have been found. These were subsequently proven by the logit models. Conscientiousness, emotional stability and openness seem to very strongly influence the behaviour of students of foreign origin. However, these results are strongly at risk to be biased, as our sample contained relatively few but extremely well performing students of foreign origin. This belief is strengthened by the fact that no existing research hints at similar effects. Hence, we think we do best interpreting these effects as *strong* and *positive* c.q. *negative* effects that require further attention, thereby not focusing too much on their exact numerical value.

Lastly, the relationship between dropout and the economy has been discussed, where a distinction between vocational - and higher education has been made. We argued that students who prematurely leave vocational education decrease the level of human capital in an economy, ensuing in reduced economic growth via foregone earnings. Dropping out of higher education translates into, besides the effects of reduced labour productivity, diminished intellectual agility and overall potential of the student in question. Therefore, we argue that the effect of dropping out of higher education has a larger magnitude than its vocational counterpart does. As a result of those implications, institutions on both the national and European level are concerned with tackling dropout rates; the phenomenon that has taken on a central role in this research.

5.2 External Validity

Some of the results we have obtained partially contradict the existing theory or are statistically insignificant, as can be seen in both the exploratory data analysis as well as the logistic regression results. Examples of these contradictions are the results of students of foreign origin displaying less dropout behaviour than students of native Dutch origin and that, within our sample, there is a relatively higher dropout rate among female students. When one compares these findings to the aforementioned literature, it seems these phenomena might be a consequence of the particular sample we have used. Therefore, extending all our conclusions to apply to the entire population of students attending Dutch vocational education might be unwise. Besides these doubtful findings, however, the research has also yielded new and significant results which are completely in line with, and contribute to, existing theories.

Conclusions that have been drawn by this research could be utilised by Dutch institutions, such as schools and centres providing vocational education, as well as the government in order to improve school retention rates. When considering our results in the context of a country with a vocational education system significantly different from the Netherlands, please note that every case is different. The intrinsic characteristics of students (Big Five), as well as the external factors at play, may be far from similar to the Netherlands. Hence, repeating the approach of this research in such a country could yield contrasting results.

5.3 Limitations

Several encountered limitations have already been mentioned throughout the text, such as the staggeringly high average dropout rate in our dataset which is most likely a signal of sample bias. Compared to the Dutch national average the rate is extreme. Even though our dropout variable technically describes *stopout*, the percentage remains high. Nonetheless, it would be wise to test our findings in different contexts and samples.

Secondly, the field of econometrics provides numerous tools to approach complex analyses like these. Considering the background and statistical experience of the authors, more advanced and elaborate methods would potentially yield different results.

A further complication of the research is found in the number of observations in the sample: With a dataset containing over 46,000 observations, this research could analyse a mere 1% of all students in the dataset, as a result of missing data in key variables. Another drawback concerning the data has been the absence of a codebook accompanying the dataset. As a result, fully tangible explanations for e.g. the high dropout bias in the sample are barely possible.

Lastly, due to time constraints, as well as the lack of appropriate data for the matter, quantitative analysis was not an option for the economic and monetary consequences of dropout in vocational education. Empirically testing the fit of Mincer's earnings model would have provided excellent additional insight. However, this part remained completely qualitative (i.e. literature based) in our research. In any case, applying statistical analysis to data that describes these phenomena suggests an interesting subject and merits further research.

Bibliography

[Acemoglu and Autor, 2014] Acemoglu, D. and Autor, D. (2014). Lectures in Labor Economics.

[Arrazola et al., 2005] Arrazola, M., de Hevia, J., Risueño, M., and Sanz, J. F. (2005). A proposal to estimate human capital depreciation: Some evidence for Spain. *Revista de Economía Pública*, 172:9–22.

[Barro, 2001] Barro, R. J. (2001). Human capital and growth. *American Economic Review*, pages 12–17.

[Belloc et al., 2011] Belloc, F., Maruotti, A., and Petrella, L. (2011). How individual characteristics affect university students drop-out: a semiparametric mixed-effects model for an Italian case study. *Journal of Applied Statistics*, 38(10):2225–2239.

[Benhabib and Spiegel, 1994] Benhabib, J. and Spiegel, M. M. (1994). The role of human capital in economic development evidence from aggregate cross-country data. *Journal of Monetary economics*, 34(2):143–173.

[Cairns et al., 1989] Cairns, R. B., Cairns, B. D., and Neckerman, H. J. (1989). Early school dropout: Configurations and determinants. *Child development*, pages 1437–1452.

[Catterall, 1987] Catterall, J. S. (1987). On the social costs of dropping out of school. *The High School Journal*, pages 19–30.

[Clark and Schroth, 2010] Clark, M. and Schroth, C. A. (2010). Examining relationships between academic motivation and personality among college students. *Learning and Individual Differences*, 20(1):19 – 24.

[Coleman, 1988] Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, 94:S95–S120.

[Coleman and Hoffer, 1987] Coleman, J. S. and Hoffer, T. B. (1987). *Public and Private Schools: The Impact of Communities*. New York: Basic.

[de Grip and van Loo, 2002] de Grip, A. and van Loo, J. (2002). The Economics Of Skills Obsolescence. *Elsevier Science*, 21:1–26.

[De Witte, 2014] De Witte, K. (2014). Beleid rond vroegtijdig schoolverlaten is de investering waard. *Indicator*, 16(18):20.

[Digman, 1990] Digman, J. M. (1990). Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology*, 41(1):417–440.

[Dutch Government, 2014] Dutch Government (2014). Facts and figures school dropout.

[European Commission, 2011] European Commission (2011). Keeping kids in school.

[Grift et al., 2009] Grift, Y., Van Der Geest, L., Dietvorst, A., and Sheikkariem, G. (2009). Het resultaat telt: Risicoprofiel van voortijdig schoolverlaters. *Onderzoek ROC Eindhoven*.

[Hankivsky, 2008] Hankivsky, O. (2008). *Cost estimates of dropping out of high school in Canada*. Canadian Council on Learning Ottawa.

[Hanushek and Woessmann, 2007] Hanushek, E. A. and Woessmann, L. (2007). The Role of Education Quality for Economic Growth. *Policy research working papers*.

[Hazrati-Viari et al., 2012] Hazrati-Viari, A., Rad, A. T., and Torabi, S. S. (2012). The effect of personality traits on academic performance: The mediating role of academic motivation. *Procedia - Social and Behavioral Sciences*, 32(0):367 – 371. The 4th International Conference of Cognitive Science.

[Hlavac, 2014] Hlavac, M. (2014). Stargazer: LaTeX/HTML code and ASCII text for well-formatted regression and summary statistics tables.

[King, 1999] King, A. (1999). The cost to Australia of early school leaving. *Dusselorp Skills Forum, Sydney*.

[Komarraju and Karau, 2005] Komarraju, M. and Karau, S. J. (2005). The relationship between the big five personality traits and academic motivation. *Personality and individual differences*, 39(3):557–567.

[Komarraju et al., 2009] Komarraju, M., Karau, S. J., and Schmeck, R. R. (2009). Role of the big five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences*, 19(1):47 – 52.

[Mankiw et al., 1992] Mankiw, N. G., Romer, D., and Weil, D. N. (1992). A contribution to the empirics of economic growth. Technical report, National Bureau of Economic Research.

[McDougall, 1932] McDougall, W. (1932). Of the words character and personality. *Journal of Personality*, 1(1):3–16.

[Mincer, 1981] Mincer, J. (1981). Human Capital and Economic Growth. Working Paper 803, National Bureau of Economic Research.

[Nolen-Hoeksema et al., 2009] Nolen-Hoeksema, S., Frederickson, B. L., Loftus, G. R., and Wagenaar, W. A. (2009). *Atkinson & Hilgard's Introduction to Psychology*. Wadsworth Pub Co.

[R Development Core Team, 2008] R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[Schultz, 1961] Schultz, T. W. (1961). Investment in Human Capital. *The American Economic Review*, 51:1–17.

[Statistics Netherlands, 2015] Statistics Netherlands (2015). Definitions: Basic qualification.

[Stratton et al., 2005] Stratton, L. S., O'Toole, D. M., and Wetzel, J. N. (2005). A Multinomial Logit Model of College Stopout and Dropout Behavior. IZA Discussion Papers 1634, Institute for the Study of Labor (IZA).

[University Of Alabama, 2014] University Of Alabama (2014). The Central Limit Theorem.

[Van Bragt et al., 2011] Van Bragt, C., Bakx, A., Bergen, T., and Croon, M. (2011). Looking for students personal characteristics predicting study outcome. *Higher Education*, 61(1):59–75.

[Weil, 2013] Weil, D. N. (2013). *Economic Growth*. Prentice Hall.

[Wickham, 2009] Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York.

[Yao et al., 2013] Yao, J., Yi, H. M., Zhang, L. X., Shi, Y. J., Chu, J., Loyalka, P., and Rozelle, S. (2013). Exploring Dropout Rates and Causes of Dropout in Upper-Secondary Vocational Schools in China. *Rural Education Action Program*.

[Zeigler-Hill, 2013] Zeigler-Hill, V. (2013). Advanced Statistics for Psychological and Behavioral Research.

Appendices

Appendix A

Translated Data Description

Dataset Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
dropout (binary)	499	0.627	0.484	0	1
gender (male=1)	499	0.415	0.493	0	1
foreign origin (foreign=1)	499	0.246	0.431	0	1
age_education	499	20.152	7.817	15	62
overall capacity	499	3.168	0.885	1	5
educational intelligence	499	3.269	0.992	1	5
numerical predisposition	499	3.126	1.048	1	5
verbal predisposition	499	3.184	1.164	1	5
logical reasoning	499	3.253	0.974	1	5
reasoning ability	499	3.060	1.105	1	5
evaluative ability	499	3.429	1.103	1	5
spatial predisposition	499	3.122	1.073	1	5
personality/motivation	499	3.172	0.639	2	5
effort	499	3.156	0.771	1	5
conscientiousness/tidiness	499	3.395	0.889	1	5
want for coaching	499	2.735	0.735	1	5
hope for success	499	3.255	1.065	1	5
anxiety to fail	499	2.611	1.120	1	5
want for individual stimulation	499	2.908	1.070	1	5
enthusiasm	499	2.687	1.021	1	5
decency	499	3.455	0.965	1	5
emotional stability	499	3.315	1.184	1	5
extroversion	499	3.120	1.082	1	5
agreeableness	499	2.914	0.845	1	5
conscientiousness	499	3.449	1.097	1	5
openness to new experience	499	3.281	1.030	1	5

Table A.1: Translated Summary Statistics Dataset

Appendix B

Translated Logit Odds Ratios

Table B.1: Translated Regression Results (Odds Ratios)

	Dependent variable:	
	Dropout (Dichotomous)	
	(1)	(2)
Emotional Stability	0.964	0.855
	(0.087)	(0.103)
Openness	0.990	1.104
	(0.103)	(0.123)
Conscientiousness	0.813*	0.699**
	(0.096)	(0.112)
Extraversion	1.083	1.060
	(0.105)	(0.108)
Agreeableness	1.012	1.056
	(0.122)	(0.127)
Foreign Origin: Yes		0.209
		(1.044)
Age	1.057***	1.059***
	(0.016)	(0.017)
Educational Intelligence		1.227*
		(0.113)
Conscientiousness*Foreign Origin: Yes		2.093**
		(0.250)
Openness*Foreign Origin: Yes		0.532**
		(0.242)
Emotional Stability*Foreign Origin: Yes		1.497*
		(0.214)
Constant	1.020	0.797
	(0.531)	(0.682)
Observations	499	499
Log Likelihood	-319.901	-306.682
Akaike Inf. Crit.	653.802	637.364
Nagelkerke R^2	0.052	0.119
Note:	*p<0.1; **p<0.05; ***p<0.01	

Appendix C

Translated Logit Marginal Effects

Table C.1: Translated Regression Results (Marginal Effects)

	Dependent variable:	
	Dropout (Dichotomous)	
	(1)	(2)
Emotional Stability	-0.008	-0.036
	(0.020)	(0.024)
Openness	-0.002	0.023
	(0.024)	(0.028)
Conscientiousness	-0.048^{*}	-0.082^{**}
	(0.022)	(0.026)
Extraversion	0.018	0.013
	(0.024)	(0.025)
Agreeableness	0.003	0.013
	(0.028)	(0.029)
Foreign Origin: Yes		-0.369
		(0.231)
Age	0.013^{***}	0.013^{***}
	(0.004)	(0.004)
Educational Intelligence		0.047^{*}
		(0.026)
Conscientiousness*Foreign Origin: Yes		0.169^{**}
		(0.057)
Openness*Foreign Origin: Yes		-0.145^{**}
		(0.055)
Emotional Stability*Foreign Origin: Yes		0.093^{*}
		(0.049)
Observations	499	499
Log Likelihood	-319.901	-306.682
Akaike Inf. Crit.	653.802	637.364
Nagelkerke R^2	0.052	0.119
Note:	*p<0.1; **p<0.05; ***p<0.01	

Appendix D

Logit Model Diagnostics

This appendix entry displays regression diagnostics of the models used. Among these diagnostics will be: The variance inflation factor, Cook's distance plots and residual versus fitted values plots. Please note that residuals versus fitted value plots are not as decisive in judging the regression estimation model fit in logistic regression as they are in linear regression (OLS).

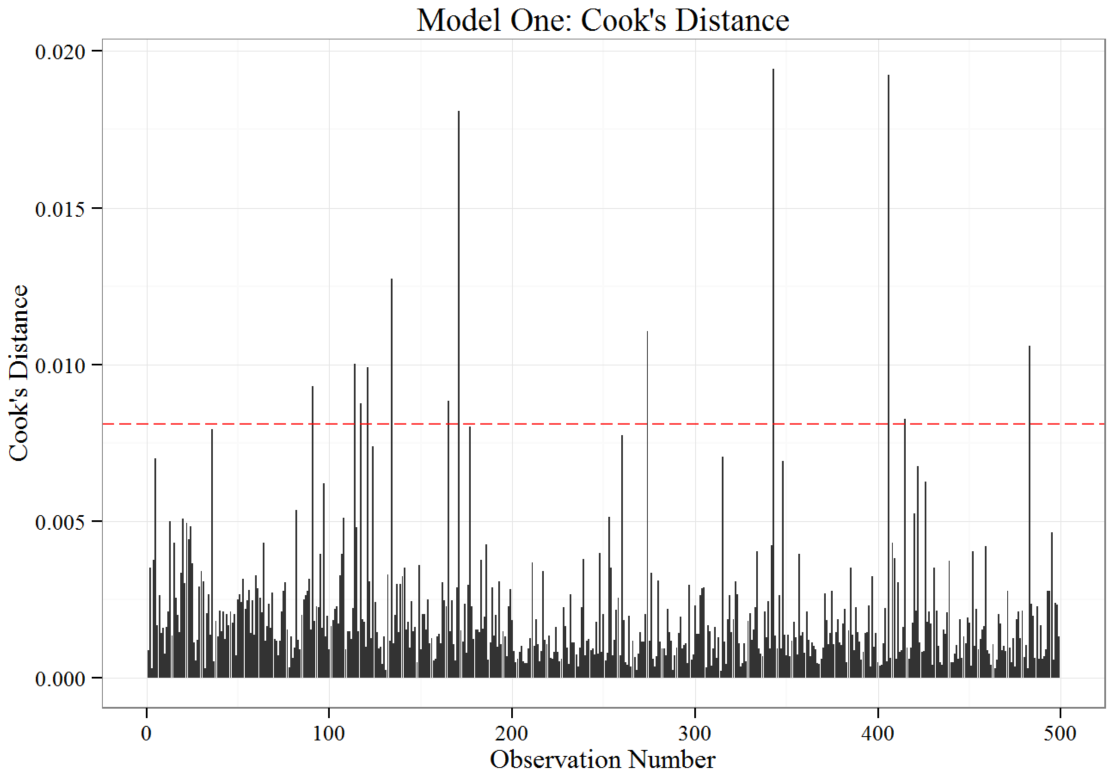

Figure D.1: Model One: Cook's Distance

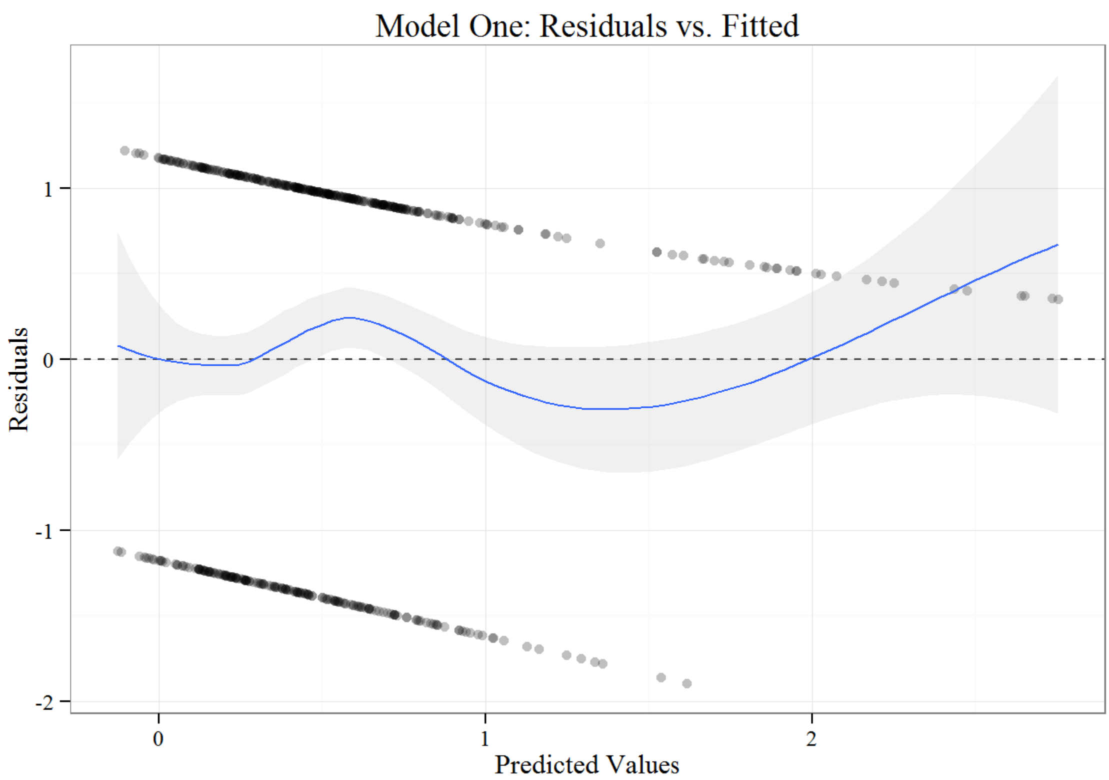

Figure D.2: Model One: Residuals versus Fitted Values

Variance Inflation Factor

Covariate	VIF	$\sqrt{VIF} > 2?$
stabiel	1.183144	No
open	1.257312	No
zorgvuldig	1.223593	No
extravert	1.428105	No
altrusme	1.182988	No
age_opleiding	1.036169	No

Table D.1: Model One: Variance Inflation Factor and Multicollinearity Test

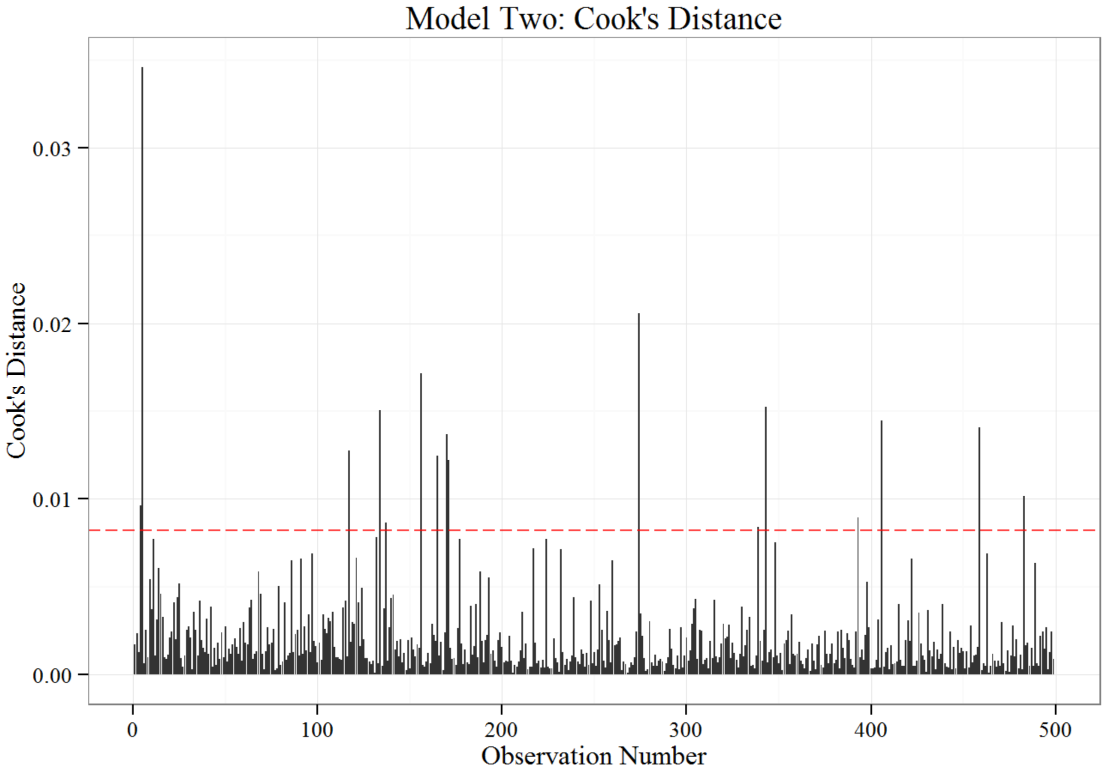

Figure D.3: Model Two: Cook's Distance

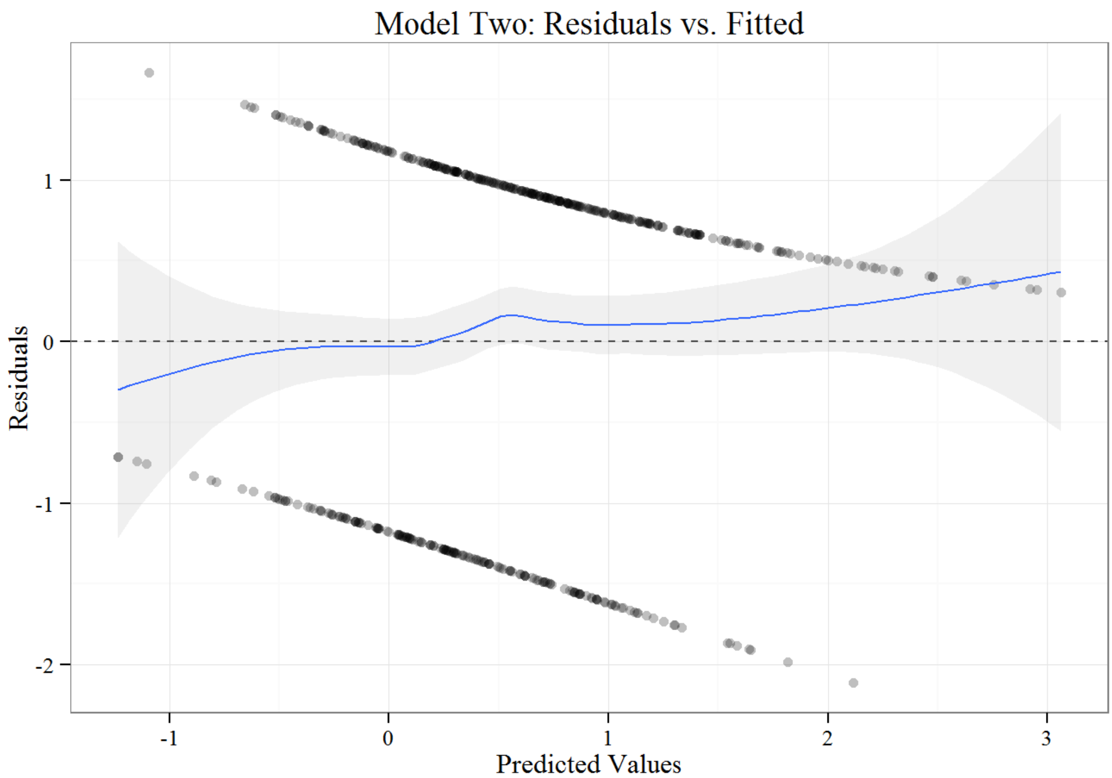

Figure D.4: Model Two: Residuals versus Fitted Values

Variance Inflation Factor

Covariate	VIF	$\sqrt{VIF} > 2$?
zorgvuldig	1.517166	No
open	1.703091	No
stabiel	1.518127	No
extravert	1.454943	No
altrusme	1.214333	No
allochtoonBinYes	21.299938	Yes
age_opleiding	1.055990	No
leerintelligentie	1.260560	No
zorgvuldig:allochtoonBinYes	18.119427	Yes
open:allochtoonBinYes	16.813727	Yes
stabiel:allochtoonBinYes	11.301315	Yes

Table D.2: Model Two: Variance Inflation Factor and Multicollinearity Test