KU LEUVEN

FACULTY OF SCIENCES

Educational Assortative Mating and Educational Mobility

Agent-based modeling and simulation to explore their relationship

Seorin KIM

Supervisor: Prof. Jan Van Bavel
KU Leuven

Cosupervisor: Prof. Vincent Ginis
Vrije Universiteit Brussel

Master thesis submitted in fulfillment of the requirements for the degree in Master of Science in Statistics and Data Science

Academic year 2022-2023

© Copyright by KU Leuven

Without written permission of the promoters and the authors it is forbidden to reproduce or adapt in any form or by any means any part of this publication. Requests for obtaining the right to reproduce or utilize parts of this publication should be addressed to KU Leuven, Faculteit Wetenschappen, Geel Huis, Kasteelpark Arenberg 11 bus 2100, 3001 Leuven (Heverlee), Telephone +32 16 32 14 01. A written permission of the promoter is also required to use the methods, products, schematics and programs described in this work for industrial or commercial use, and for submitting this publication in scientific contests.

Preface

I am deeply appreciative of the invaluable guidance provided by my supervisors, Professor Jan Van Bavel and Professor Vincent Ginis, who navigated me through every phase of this thesis project. The insights deriving from their different expertise and fields of research allowed me to explore the thesis subject from various angles and acquire knowledge about various research methodologies. Without their guidance, I would not have been able to produce a thesis of this quality.

I am also grateful for all the support I received from Ph.D. candidate Arne Vanhoyweghen, who is growing his expertise in ergodicity economics. He has hugely inspired and encouraged me to pursue this project. His compelling research projects and unique perspectives on social mobility and inequality profoundly piqued my interest in investigating educational mobility. Furthermore, his unwavering support has helped me go through the challenges I faced during the project.

Lastly, I extend my gratitude to Romy Van Haesendonck and Professor Georg Jäger. The persistent support and encouragement I received from Romy Van Haesendonck are invaluable. Keeping my commitment throughout the year would have been very challenging without these. Professor George Jäger generously shared his codes of agent-based modeling utilizing neural networks with me. His works helped me learn about different aspects of agent-based modeling and truly inspired me. Although I did not employ neural networks to construct my agent-based models in this thesis, I aspire to integrate his methodologies into my future projects.

Summary

When addressing the issue of social mobility in society, an interesting angle to consider is marriage. As it is often believed that like marries like, a mating strategy can ensure a safe transmission of one's relevant resources that can guarantee an equal or even higher level of social position for children. This thesis focuses on educational mobility and how the Educational Assortative Mating (EAM) influences Intergenerational Educational Mobility (IEM) in the long term using a simulation technique called Agent-based Modeling (ABM). Specifically, it delves into whether the prevalence of similarly educated people marrying in society makes it more difficult for individuals to climb the social ladder than the prevalence of marriages between different educational backgrounds.

Agent-based models are created to represent a society where educational opportunities are well distributed among different social strata. The inquiry showed that societies that prefer heterogamous marriages generally demonstrate higher growth in average educational levels at the population level compared to societies that prefer homogamous marriages. Regarding absolute mobility in education, heterogamous societies did not necessarily show the highest mobility level; that distinction went to homogamous ones. However, in the case of homogamous societies, the higher educational mobility appeared to be a result of a few higher social strata dominating higher education, while in heterogamous societies, people from different backgrounds could benefit from the societal growth of the educational level.

The thesis leveraged one of the advantages of Agent-based Modeling: the ability to explore complex realities using a relatively simple model. However, it is important to acknowledge that this simplicity may limit the generalizability or extension of the conclusions to broader contexts. While the findings may be specific to the context and parameters used in the study, they lay the groundwork for future investigations and extensions of the model to address a wide range of research questions in mobility and inequality studies.

Acronyms

ABM Agent-based Modeling.

ANOVA Analysis of Variance.

EAM Educational Assortative Mating.

EBM Equation-based Model.

ESS9 European Social Survey Round 9.

IEM Intergenerational Educational Mobility.

IRL Inverse Reinforcement Learning.

ML Machine Learning.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Throughout history, marriage has been pivotal in perpetuating social class from one generation to another (Nomes and Van Bavel, 2017). Traditionally, loyal families married each other to maintain or extend their power and status (Coontz, 2004). For people of other classes, marriage allowed them to maintain their properties, livestock, money, occupational status, and social connections, *inter alia* (Coontz, 2004). As much as in the past, in today's society, where rigid social class distinctions like the caste system have largely faded and the pursuit of equality prevails, marriage still seems to serve as a tool to reproduce a social class structure and to alter one's social class (Van Leeuwen and Maas, 2005). Marrying a partner with a higher socioeconomic class can lead to upward social mobility for themselves and their offspring, whereas marrying a partner from a lower socioeconomic origin risks downward mobility (Dribe and Lundh, 2010). Consequently, socioeconomic homogamy, which refers to finding a partner from a similar socioeconomic origin, is preferred as it reduces the risks of downward mobility for both individuals within a couple (Dribe and Lundh, 2010; Bourdieu, 2002).

In today's societies, another dimension that plays an important role in shaping socioeconomic structure to influence one's marital decision is education (Nomes and Van Bavel, 2017). In tandem with the increasing female participation in tertiary education, there are more chances to meet a similarly educated partner than before. For instance, in Europe, where women had surpassed men in the enrolment for tertiary education already in the late 1990s (Van Bavel, 2012), marriages where the male partner's education is higher than the female partner's (i.e., hypergamy) have witnessed a decline in prevalence (Erát, 2021). The share of homogamy was predominant in Western European countries across cohort groups between 1941-1980, while the share of hypogamy, characterized by higher female partner's education than male partner's, increased and hypergamy thereof decreased (Grow and Van Bavel, 2015).

Considering the observed educational assortative mating pattern, the thesis questions whether the 'like marries like' strategy stratifies society significantly more than the 'marrying up/down' strategy. In particular, I am interested in investigating intergenerational mobility in education to compare the impact of homogamy and heterogamy, a concept encompassing hypergamy and hypogamy together. In order to examine the long-term impact of a marriage pattern on educational mobility over more than two generations in a controlled environment, Agent-based Modeling (ABM) will be utilized. With synthetic data made with agent-based models, the thesis seeks to gain insights into the intricate relationship between Educational Assortative Mating (EAM) and educational mobility. This way, the thesis hopes to contribute to modeling educational mobility, which sociologists and economists have profoundly studied.

The following sections of the chapter will offer literature reviews on the concept of educational mobility and the Agent-based Modeling. For the literature reviews on the sociological concepts regarding educational mobility, it should be acknowledged that the thesis does not offer detailed nuances of the concepts since the thesis's interest lies in the statistical and data-scientific aspects of the subject matter. After delving into educational mobility, research objectives and scope will be elaborated. Then, the main methodological force, the ABM and the significance of the thesis will be introduced.

Moving into the second chapter, the methodology used for building the model and analyzing the generated data will be discussed in detail. Then, the third chapter will delve into the results of the analysis. The last chapter will conclude the thesis by highlighting the main outcomes, limitations, and suggestions for future research.

1.1 The Role of Education in Mobility

Social inequality is an issue that every society has to grapple with. One key aspect of understanding inequality is the study of social mobility. In a sociological study, social mobility concerns a change in social class (Erikson and Goldthorpe, 2002). The change can be investigated by comparing parents to their children, which is referred to as *intergenerational mobility* or throughout one's lifetime, called *intragenerational mobility* (OECD, nd; Breen, 2004). In both cases, high social mobility leads to a flexible society where people can travel beyond the social class they were born into (Breen, 2004). Contrarily, low social mobility reflects social inequality as people are hindered from climbing up the social ladder (Breen, 2004).

To study mobility, sociologists often utilize the concept of social class, which is often understood as one's socioeconomic status (Mueller and Parcel, 1981). Popular schemata involve occupational status and employment relations to categorize individuals into social classes (Torche, 2015; Erikson and Goldthorpe, 2002). Economists whose interest in social mobility studies grew later than sociologists tend to focus on income (Torche, 2015; Beller and Hout, 2006). Despite the difference in approaches, education plays a vital role in both approaches. One's educational attainment is indirectly implied in one's occupational status and highly correlates with income (Torche, 2015). This suggests that understanding the educational structure in society has far-reaching implications for social mobility studies.

Sociologists and economists involve the educational dimension in social mobility studies differently. Sociologists recognize the mediation effect of educational attainment on the relationship between social origin and social destination (i.e., social mobility) (Goldthorpe, 2014; Reichelt et al., 2019). For economists, the important role of education is identified with human capital, a key factor of social mobility (Yang and Zhou, 2022; Arenas and Hindriks, 2021). With indicator-based human capital measures, which primarily use how much one invested in education (e.g., years spent in education) (Abraham and Mallatt, 2022), some economic studies focus on the mobility in human capital to examine trends in intergenerational mobility (Card et al., 2018). Likewise, the two research fields acknowledge education as an important socioeconomic aspect to consider. In this thesis, the primary focus is on mobility in education, which resembles the economists' approach while still utilizing some sociological concepts in the study, which will be introduced in the next section.

Education is often portrayed as a tool to increase social mobility. Economists, Arenas and Hindriks (2021), demonstrated the role of education in intergenerational mobility as well as human capital using simulation. Unequal educational opportunity increases human capital at the societal level because high-quality schools are predominantly matched to high-income families, whereas it decreases intergenerational mobility (Arenas and Hindriks, 2021). Jerrim and Macmillan (2015) highlighted the vital role of education in the relationship between income inequality and social mobility (viz., Great Gatsby Curve). Namely, unequal access to financial resources (i.e., income inequality) is positively associated with the intergenerational transmission of educational attainment and negatively with the university tuition fee (Jerrim and Macmillan, 2015). Following this logic, one may argue that the impact of education on social mobility must be diminishing in societies where education is increasingly accessible to a larger population (Goldthorpe, 2014; Breen, 2019).

Recently, Ruggera et al. (2023) conducted a study aiming to examine the association between social origin and social destination in Finnish cohorts born between 1951 and 1980. This study was done in Finland, which is considered more egalitarian than other societies (Ruggera et al., 2023). The findings revealed greater fluidity among lower educational levels when looking at the association throughout time per educational level. However, among the tertiary-educated group, social origin appeared to be more influential in determining social destination (Ruggera et al., 2023). The authors highlight that educational opportunity does not increase social mobility in a linear way, as one may suspect from the result of Arenas and Hindriks (2021) mentioned above. The study highlights the persistent influence of parents' social positions on the social destination, even in a society with widespread educational opportunities.

Given that parents' social status is also influenced by their education and cultural assets, which will be introduced as the concept of cultural capital in the next section, it is crucial to delve deeper into how education is transmitted across generations. Furthermore, understanding who genuinely benefits from the overall increase in the mean educational attainment of society can provide valuable insights into the dynamics of social mobility. Consequently, this thesis focuses on educational mobility as a fundamental component of social mobility, aiming to provide insightful perspectives on this critical aspect.

1.2 Objectives and Research Scope

To address the issue of (im)mobility, particularly in the context of education, this thesis adopts a unique approach by examining the phenomenon through the lens of mating strategies. The primary objective of this study is to explore the relationship between educational mating patterns and educational mobility. For this, three **research questions** are formed to be investigated:

1. Does educational mobility increase more in the long term when people with different educational backgrounds marry each other compared to when marriages only happen within the same educational group?

2. Does a higher population-wise educational level (i.e., a higher cross-sectional average educational level of the entire population) reflect a higher degree of intergenerational educational mobility?

3. Does every educational group experience the same improvement or decline in average educational level throughout generations?

Here, it is important to define *educational mobility* as used in this paper. As for social mobility, educational mobility can be examined from two perspectives: (1) absolute mobility and (2) relative mobility. Absolute mobility "captures the total observed change in educational attainment across generations" (Torche, 2019, p.2). Relative mobility, on the other hand, "captures the association between parents' and children's education net of any change in the distribution of schooling across generations" (Torche, 2019, p.2). Absolute mobility does not consider how many people of a generation possess higher education degrees, while relative mobility does. Therefore, under the assumption that both societies experienced a large influx of students in upper tertiary education (e.g., Master's degree), a society where more people are highly educated may manifest relatively lower mobility than a society where most people are lowly educated.

To measure absolute mobility, one can measure the probability of an individual having a higher educational level than their parents when their parents do not possess the highest possible educational level (van der Weide et al., 2021). To measure the relative mobility, the correlation between the parents' education and children's education can be obtained throughout the generations, or a regression coefficient can be obtained (van der Weide et al., 2021). This way, these relative measures take into account the differences in schooling between the generations and the impact of the changes in the educational system over generations.

When measuring absolute mobility, categorical education levels are utilized, which are often referred to as the highest level of education completed or educational attainment. In contrast, continuous education levels, such as years of education completed, are utilized when measuring relative mobility, as the correlation and regression coefficients require continuous variables. As a result, two education variables are incorporated into the thesis, each representing a slightly different aspect of education. The first is the highest level of education completed or **educational attainment**, and the second is **cultural capital**. A continuous variable, years of education completed, is not considered since it is challenging to incorporate it into the simulation in a meaningful way. The main challenge is that a higher number of years of education does not necessarily indicate a greater level of immersion in higher education, making the translation between years of education and educational attainments tricky.

Cultural capital offers a broader view of the distinctions between individuals in terms of education. Cultural capital, a concept deriving from Bourdieu's social reproduction theory, refers to "instruments for the appropriation of symbolic wealth socially designated as worthy of being sought and possessed" (Bourdieu, 1973, p.72). It incorporates three forms with varying tangibility: cultural capital in the embodied state, such as one's knowledge, skills, tastes, and posture; in the institutionalized state, such as educational qualifications; and in the objectified state or cultural goods, such as pictures, books, and instruments (Bourdieu, 1986; Sieben and Lechner, 2019). The embodied cultural capital is less tangible than the other forms and obtained through one's education and individual socialization and experiences (Bourdieu, 1986; Sieben and Lechner, 2019).

Despite many attempts to empirically measure cultural capital (e.g., by the number of books in a household in Sieben and Lechner, 2019; by observing one's taste in cultural goods in Špaček, 2017; and by diet and physical activity in Mudd et al., 2023), it appears challenging to incorporate three forms of cultural capital into one quantity since not every element of cultural capital is observable. The use of simulations in this study offers a distinct advantage by allowing for the representation of cultural capital in a manner that would otherwise be difficult to achieve. Further details regarding the definition of this variable will be elaborated in Chapter 2.

While acknowledging that the cultural capital variable may not perfectly nor correctly capture the entirety of an individual's cultural resources, it still offers valuable insights in addition to the categorical measurement of educational attainments. It is widely acknowledged that even within the category of individuals with tertiary education, certain differentiating factors exist, such as the specific area of study, the prestige of the university attended, the cultural knowledge passed down through the family, and the family's social status. In the simulation, these distinctions can be attributed to variations in the accumulation of cultural capital. Therefore, even if the cultural capital variable in the simulation does not correctly reflect cultural capital, it will allow us to consider other differentiating factors within each educational level, yielding more comprehensive conclusions than relying solely on a simplistic categorical measurement.

1.3 Agent-based Modeling

To build a simulation, Agent-based Modeling (ABM) is used in the thesis. ABM is a computational modeling approach that focuses on simulating a system based on the behavior and interactions of autonomous entities, known as agents (Macal and North, 2009). This modeling can be based on a theory (i.e., theory-based ABM) or on the data researchers have in hand, which then can be used to set the initial stage or to define the behavioral rules (i.e., data-driven ABM) (Turgut and Bozdag, 2023).

A system in a simulation study is a facility or process such as a bank where clients and bank tellers deal with different types of services and a logistic network in a supply chain (Kelton et al., 2015). In the thesis, a system is a specific world where agents interact to marry and reproduce. These agents are autonomous and operate the transitions between states of the system based on the behavioral mechanisms put into them (Conte and Paolucci, 2014). Therefore, in the thesis, the agents autonomously look for a partner based on their preferences and procreate, influencing the system they belong to.

Unlike traditional modeling techniques, Equation-based models (EBMs), that often rely on aggregate variables and focus on system-level information, ABM offers a more granular perspective, capturing the heterogeneity and individual decision-making processes of agents (Parunak et al., 1998). In an EBM, the system is simulated from top to bottom. System-level variables are used extensively to build an equation, which typically forms interrelated differential equations, representing the system and to evaluate the equation afterward (Parunak et al., 1998; Tang et al., 2006). An agent-based model, on the other hand, uses a bottom-to-top approach (Jäger, 2021). The modeling focuses on the agents' behavior and interaction with one another, which outputs system-level information at the end (Macal and North, 2009; Parunak et al., 1998).

The fact that building the agent-based models has its main focus on the agents and their interactions allows a larger degree of flexibility in the model-building procedure, especially when the system of interest is complex (Parunak et al., 1998; Tang et al., 2006). Not only is it more natural to describe a system from the perspective of the agents (Bonabeau, 2002), this approach does not require a thorough understanding of all the dynamics in the complex system Jäger (2021). Instead, knowledge about individual-level behavioral rules or equations is sufficient to build the model (Jäger, 2021). By putting the agents in action, emergent phenomena can be captured from the model, allowing insights into the system of interest (Bonabeau, 2002).

Contrarily, to build a system with an EBM, one must acknowledge the rules of the system dynamics that he or she aims to study and extensively use this system-level information in the model to both design and evaluate the system (Parunak et al., 1998). The reliance on the system-level information makes it difficult to investigate the simulation results at the individual level or at the level of a specific group (Tang et al., 2006).

Despite the strengths, it should also be noted that the flexibility in building an agent-based model comes with requiring many assumptions in the model. The individual-level rules about how the agents behave and interact with other agents do not reflect the entire dynamics and complexity of the world that one aims to represent. Moreover, some of these assumptions are rather ad hoc (Grow and Van Bavel, 2017). To account for this, some authors suggested Machine Learning (ML) driven approach in the agent-based modeling (Turgut and Bozdag, 2023). For instance, Jäger (2019; 2021) uses neural networks to observe types of agents' behaviors and later uses these types in the model to define the behavioral rules for the agents, and Lee et al. (2017) utilize the Inverse Reinforcement Learning (IRL) technique to have a better definition of behavioral rules.

In this thesis, agent-based modeling is the preferred approach, as it offers a valuable means to examine the intricate relationship between educational assortative mating patterns and educational mobility. This relationship is complex, as it involves understanding the educational levels of both partners and individuals themselves compared to their parents. Incorporating the factors that influence an individual's educational level and that of their partner, as well as comprehending how mating patterns impact educational mobility and vice versa, presents a significant challenge when attempting to capture it in a single equation. Therefore, agent-based modeling serves as a convenient and effective tool to explore this complex relationship by simulating various agent behaviors.

The aforementioned approaches, which aim to complement ABM, were not adopted due to their complexity. Although intriguing, the thesis opts not to pursue the black box characteristics of the machine learning (ML)-driven approach, as it aims to conduct a thought experiment to test hypotheses (Turgut and Bozdag, 2023). The use of Inverse Reinforcement Learning (IRL) in ABM allows for studying the motivations behind an agent's behaviors and extracting rich yet concise behavior rules (Lee et al., 2017). However, applying this technique requires access to behavioral data, which proved challenging to obtain for the research objectives of this thesis (Turgut and Bozdag, 2023).

1.4 Significance of the Thesis

The thesis holds a twofold significance. Firstly, it provides a foundational agent-based model that can be adapted for studying various types of mobility. As noted by Turgut and Bozdag (2023), sharing codes in the field of ABM is less common compared to other disciplines. Therefore, this thesis makes a valuable contribution to the field by sharing the agent-based models developed using the `mesa` package. Additionally, it demonstrates a method for transferring simulated data to a graph database management system, specifically `Neo4j`. Although the analysis presented in the later chapters does not primarily rely on this system, the integration of `Neo4j` enables broader analytical opportunities for future research, particularly when dealing with complex systems and diverse outcome measures.

Secondly, the thesis contributes to the broader understanding of educational mobility. While some scholars challenge the significance of the role of education in explaining social mobility, this thesis' experiment offers a more comprehensive view of educational mobility dynamics. The inherent flexibility of the simulation allows for broader reflections on the conclusions, as will be further elaborated in Chapter 4. Beyond exploring the relationship between mating patterns and mobility, the thesis' experiment encourages a deeper examination of the dynamics between diversity and societal growth, inviting questions about their interplay and impact on broader social dynamics.

This chapter has served to introduce the research questions of the thesis, outline the relevant concepts that underpin the research inquiry, and elucidate the chosen methodology. It has also shed light on the significance of the thesis. Moving forward, the subsequent chapter will delve into the specifics of constructing the agent-based models and the used analytical tools.

Chapter 2

Methods

Agent-based models are constructed using `mesa` package in Python (Kazil et al., 2020) to investigate the relationship between educational assortative mating and educational mobility. The main idea behind the built agent-based models is that the agents meet a partner who satisfies their requirements and then procreate the next-generation agent together with the partner. In order to build this system where agents interact in such a way, two main rules must be defined: (1) How do agents marry? and (2) How does parents' educational level transmit to their children? Moreover, before incorporating these rules into the agents, some characteristics of the agents and the spatiotemporal environment of the system should be set up.

As discussed in Section 1.3, setting these rules and characteristics requires assumptions, which will be discussed based on literature in the later sections. Although the thesis aims to experiment with the relationship between mating patterns and mobility in education rather than empirically prove it, the initial state of the system is tailored to the context of Belgian society. However, since the behavioral rules are put into the system to perform a thought experiment, the results will not be limited to Belgian society. Instead, the thesis hopes to offer a simulation model to explore the relationship as well as to serve as a base model to develop an agent-based mobility model.

The agent-based models incorporate global and local level parameters. At the global level, the model establishes the agents' spatiotemporal environment and its global characteristics (Kazil et al., 2020). At the local or agent level, the model outlines how agents behave and interact with each other (Kazil et al., 2020). In other words, while the global level provides typically unchanging rules governing the agents' environment, the local level defines the rules governing the agents' behavior and interaction.

This chapter will begin by describing the spatiotemporal environment of the system. The second section will introduce the local-level parameters that characterize the agents. The third and fourth sections then delve into the rules applied in the models. The former discusses system-wise rules, while the latter focuses on the behavioral rules governing individual agents. Lastly, the chapter concludes by introducing analytical tools employed for analyzing the simulated data.

2.1 Spatiotemporal Environment of the System

Agent-based models are constructed using a *network grid* to represent the **space** where the agents interact with each other. Note that `mesa` package utilizes `networkx` package for the network grid (Kazil et al., 2020). Each agent is assigned to one of the initial *n* nodes, and on average, they are connected to *m* other agents in the network (i.e., *m* edges) (Kazil et al., 2020). For instance, when one sets the model to have 50 nodes and 50 edges, a complete network is formed where agents sit in one of the 50 nodes and have a connection with 50 agents, including themselves on average. To randomly generate edges (or connections) between nodes, an Erdős–Rényi model with *n* nodes is employed, and the probability is determined by the average node degree given in the initial setup divided by the total number of nodes in the network (i.e., m/n) (Hagberg et al., 2008). Each of the initial agents can be connected to another determined by this probability (Hagberg et al., 2008). This probability may change slightly as the simulation evolves further and as some agents die and reproduce.

One may argue that the use of the random model generator, Erdős–Rényi model, is unrealistic because, in reality, people tend to associate with others who share similar characteristics, and some individuals are connected to more people, becoming a hub in the network. Nevertheless, the use of a random network model with equal probability for each edge can be justified. Firstly, concerning the fact that the randomness may not represent the invisible 'bubbles' in societies created by similar status and socio-economic class, it should be highlighted that our main research interest is not who is acquainted with whom but rather who will marry whom. Who meets whom to marry can be determined by the agent's behavioral rules at a later stage instead of already establishing conditions for what types of agents should connect with each other. This way, additional assumptions about edge creations are not necessary.

Secondly, Watts-Strogatz's or Newman-Watts-Strogatz's new world models could be used as they allow hubs in the model (Watts and Strogatz, 1998; Hagberg et al., 2008). However, the considered research question does not require knowing who is most popular in the network as it is not assumed to impact agents' mating behavior. Moreover, the simplicity of the Erdős–Rényi model is preferred over these two models. Still, the thesis offers an example of

Newman-Watts-Strogatz's new world model on a GitHub repository as an additional source (See Appendix on Github).

In the thesis, a set of 500 initial agents will be utilized to simulate the models. While a larger number of agents could potentially enhance the validity of the simulation, a trade-off with computational efficiency had to be considered. The average number of edges employed will also be set at 500, forming a complete network. This decision aims to provide a broader array of connections for agents to explore and select potential partners. By offering a wider net of connections, the simulation aims to isolate the influence of the partner-seeking rules, reducing the impact of other external factors. Further details regarding the behavioral rules will be discussed in Section 2.4.

Regarding the aspect of **time**, a *random activation rule* is employed to schedule agents in each time step of the simulation. Under this rule, every agent is activated once per time step, ensuring that all agents have an opportunity to behave and interact with others (Kazil et al., 2020). The activation order follows a random sequence, meaning that agents are activated in a randomized order in each time step. For instance, agent number 1 may be the first one to initiate a partner search, followed by agent number 100, and so forth, until the final agent is reached. The order is then shuffled in the subsequent time step (Kazil et al., 2020).

The order of activation can have implications when the same set of mating rules is applied to both male and female agents, as introduced in Sections 2.3 and 2.4. Consider a scenario where individuals of both genders seek partners with higher income prospects than their own. If every agent starts looking for a partner simultaneously, numerous clashes between agents may arise, generating a significant number of agents without a partner. However, with a random activation rule, agents seek partners individually, one by one. Consequently, if a first-order agent finds a male partner with a higher income prospect, it implies that the male partner has settled for a partner with lower income prospects than his own.

2.2 Agent Characteristics

At the local level, parameters in Table 2.1 are set to define the agents' characteristics. Note that regarding the construction of the agents' characteristics and behavioral rules, the thesis is inspired by the agent-based model of Grow and Van Bavel (2015), where the marriage market of Western European societies is simulated. The variables can be understood in two categories. There are variables to keep track of information, including State, Spouse, Parents, Cohort, Generation, and Children. The other variables, including Age, Gender, Education, Income, Weight, Cultural, Economic, Social, and Capital, influence the agent's mating strategy.

Table 2.1: Local-level variables used for agent characteristics in the ABMs

Variable	Description
State	It indicates the marital status of an agent: married, married and reproduced, or single. The initial state of agents is single.
Spouse	It contains information about the spouse of married agents. The initial value of this variable is `None`.
Parents	It tracks the parents' IDs. The initial value of this variable is an empty list, implying no parents' information for the initial n agents.
Children	It tracks the children's IDs when a couple procreates. The initial value of this variable is an empty list.
Cohort	It enumerates the cohort number of each agent. The initial agents are assigned one of the first three cohort numbers (1 to 3).
Generation	It enumerates the number of generations in a family tree. For the initial n agents, generation is 1, regardless of their age or cohort.
Age	It indicates the age of the initial agents as an integer, which is assigned following a rule described below.
Gender	It indicates the gender of agents.
Education	It indicates an agent's highest educational level completed.
Income	It indicates the earning prospect of an agent from 0 to 800.
Weights	It expresses, from 1 to 10, how much importance an agent conceives for cultural, economic, and social capital.
Cultural	It expresses an agent's cultural capital based on the agent's educational level and weights for cultural capital.
Economic	It expresses an agent's economic capital based on the agent's income prospect and weights for economic capital.
Social	It expresses an agent's social capital based on the agent's connections and weights for social capital. The initial value at time step 0 is zero.
Capital	It is a sum of Cultural, Economic, and Social capital values.

To initiate the agent-based model, some characteristics need to be specified to n initial agents, which can be arbitrarily given following a theory or based on existing data. In this thesis, some existing data are obtained to set the distribution of Cohort and Education for the initial n agents.

The **cohort** variable indicates the cohort number of each agent, with lower numbers corresponding to older agents. The initial agents are assigned one of the first three cohort numbers (1 to 3). The assignment probabilities for each cohort number are derived from the age groups between 15 and 59 years old in the ESS9 dataset. Specifically, 47% of the initial agents are assigned to Cohort 1, 43% to Cohort 2, and only 10% are allocated to the youngest cohort, 3. Cohort 1 includes individuals aged between 40 and 59 in the ESS9 dataset, Cohort 2 comprises those between 20 and 39, and Cohort 3 consists of individuals aged between 15 and 19 (European Social Survey European Research Infrastructure (ESS ERIC), 2021).

The agent's **age** is determined based on their cohort, represented as an integer. For agents in Cohort 1, their age is randomly assigned between 40 and 59. For Cohort 2, it ranges from 20 to 39, and for Cohort 3, it ranges from 15 to 19. Consequently, the initial agents in the model have an age range between 15 and 59. This choice serves two purposes. Firstly, as in the model proposed by Grow and Van Bavel (2015), where the age range for the marriage market was limited to 16 to 80, it is necessary to start with initial agents aged around 16 to facilitate partner-seeking and procreation already from time step 1. Secondly, this thesis incorporates the reproduction of the agents, leading to the need to consider the fertility of women. According to the (World Health Organization, 2023) definition of fertility indicators, fertile women are typically aged between 15 and 49. This indicates that the maximum age in the model can be smaller than 80 since procreation is unlikely to occur after a woman reaches 49 years old and also typically after a man reaches 59 years old.

It is crucial to note that setting the cohort first and then the age for the initial agents reduces the variation between replications of the same simulation model more compared to the reversed strategy using, e.g., a triangular distribution to assign one's age and then deciding the cohort number based on the age.

The initial agents are randomly assigned a **gender**, with an equal probability of being assigned as either male or female. The equal probability is employed for simplicity. All the initial agents of different cohorts and ages are considered **generation** 1 in the model to indicate that they are the initial agents given to the system, not the ones generated from procreation. These initial agents also do not have parents – hence, no value for the **parent** variable, which tracks the parents' ids. At the initial stage, the agents' **state** is 'single'. When the agents meet a partner, the spouse's information is stored in the **spouse** variable, and their state changes to 'married'. When the married agents reproduce, their state changes to 'reproduced', and their children's ids are stored in the **children** variable.

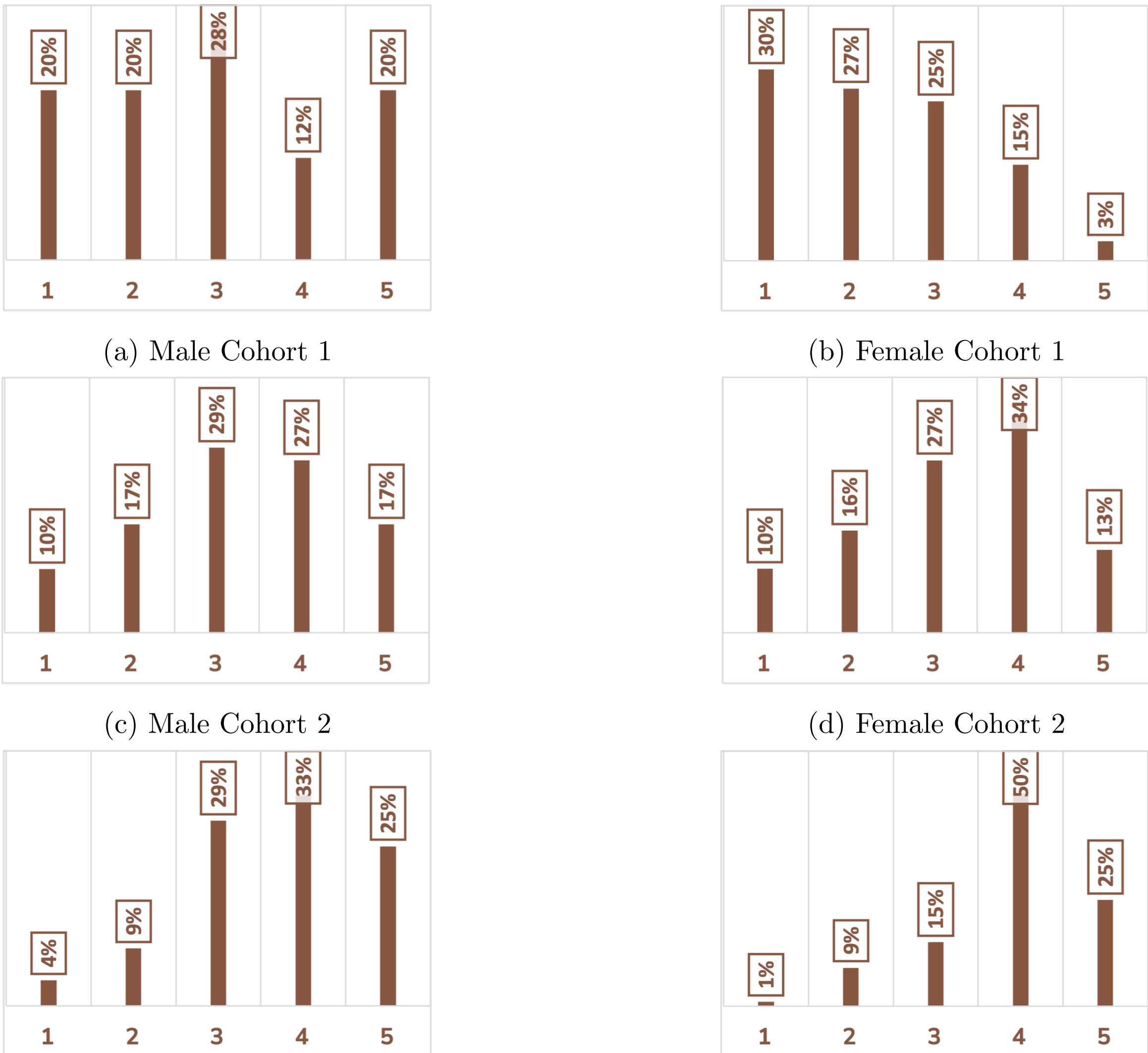
The other variable that utilizes existing data is **education**, which can be considered the most important variable for the research questions. Education expresses an agent's educational attainment with five categories following the ESS9: i) Primary education, ii) Lower secondary education, iii) Upper secondary education, iv) Lower tertiary education (bachelor's degree and technical college degree), v) Upper tertiary education (master's degree or higher). The category representing individuals without any education is not included, as the ESS9 result indicates the absence of individuals without education in the Belgian data set (European Social Survey European Research Infrastructure (ESS ERIC), 2021).

It is important to notice that this variable should be understood separately from the agents' age since it indicates an agent's final educational level, not the agent's educational level at a specific time step. Therefore, one's education value does not change over time steps. For the initial agents, the educational level is distributed based on probabilities resembling the education distributions of Belgian respondents in ESS9. Six different distributions are used for the three initial cohorts of two genders, as shown in Figures 2.1 below. When adopting the distributions from the ESS9 data set, only the age groups equal to or above 30 years old are considered since the respondents younger than 30 years old in the data set still had the potential to pursue further education.

As depicted in Figures 2.1, there is a notable difference in primary education attainment between males and females in Cohort 1, with a larger proportion of females having primary education. Across both genders, the number of individuals with lower tertiary education increases across cohorts, while the number of individuals with only primary education dramatically decreases. In Cohort 3, approximately half of the female population holds a lower tertiary education degree, whereas only 1% have completed primary education. For males, the decline in the population with primary education is relatively less drastic than for females. In Cohort 3, 4% of males possess a primary education degree, while 29% have an upper secondary education. In contrast, only 15% of females in Cohort 3 have completed an upper secondary education degree. Interestingly, both genders have an equal proportion of the population with upper tertiary education in Cohort 3.

Education is conceived influential in shaping one's income. In the model, the **income** variable represents the earning prospect of an agent ranging from 0 to 800, as adapted from (Grow and Van Bavel, 2015). Similar to education, an agent's income remains constant throughout the simulation, reflecting their earning prospects over their entire lifespan. The first n agents' income is defined based on their educational level with triangular distributions. These distributions are informed by the observed correlation between earnings and educational level in Belgium (OECD.Stat, 2019), where higher educational attainment reflected a higher average income.

The income distributions are designed to allow all educational categories the potential to reach the maximum income value of 800. However, in order to provide a generalized representation of the population among the first n agents, distinct minimum values are assigned to each educational group. The agents' income can vary between 0 and 800, with a mode of



(a) Male Cohort 1 (b) Female Cohort 1
(c) Male Cohort 2 (d) Female Cohort 2
(e) Male Cohort 3 (f) Female Cohort 3

Figure 2.1: Distribution of educational levels per cohort and sex for the initial agents.

100 for the primary education group, between 50 and 800 with a mode of 200 for the lower secondary education group, between 100 and 800 with a mode of 300 for the upper secondary education group, between 150 and 800 with a mode of 400 for the lower tertiary education group, and between 200 and 800 with a mode of 600 for the upper tertiary education group. Although this approach is not entirely realistic, this decision is made to ensure that the model can capture general societal trends among the initial agents. It is important to note that these rules are solely applicable to the initial agents, while different rules will be applied to the child agents based on the parent's income.

Based on the education and income values, three types of capital of agents are given. In Section 1.2, cultural capital is briefly introduced, which is part of the capital. In Bourdieu's reproduction theory, capital consists of three types: economic, cultural, and social capital. Affluence in the three capitals often reflects a greater potential to have a higher social standing. In the simplest terms, economic capital represents one's material assets that can be immediately converted into money, and social capital is an aggregate of potential resources deriving from one's relationships and networks, or "membership to a group" (Bourdieu, 1986, p.21).

As for cultural capital, expressing these types of capital in quantity presents a challenge. In this thesis, some observable factors attributing to each of these capitals are utilized: for instance, one's educational levels for cultural capital, one's income for economic capital, and one's number of useful acquaintances for social capital. In order to arbitrarily express the complexity around one's capitals, a variable called weights is introduced to the observable variables.

The **weights** variable expresses, from 1 to 10, how much importance an agent conceives for three types of capital, and agents receive one of the 10 values for each capital randomly. The weights are integers between 1 being 'Not at all important' to 10 being 'Very important', and the three weights are stored in a list in the order of the weights for cultural, economic, and social capital. Instead of using a continuous scale, the ordinal scale is used to reduce the randomness between replications. The weights will often be used not as an integer but as a ratio to the maximum value, 10 (i.e., $weight/10$), when calculating capitals.

Capitals are calculated in the composition of an observable variable and a weight. Firstly, the **cultural** variable aims to depict agents' cultural capital based on their educational level, with the weight for cultural capital noted as `WeightCC`. The cultural capital for an agent i is formulated as follows:

$$Cultural_i = (Education_i + WeightCC_i * 0.1 * Education_i)/10 \tag{2.1}$$

The denominator is 10, such that the upper bound is 1 and the lower bound of this value is 0. Therefore, when one's educational level is high and he/she finds cultural capital very important, his/her cultural capital is high.

Secondly, for economic capital, the **economic** variable utilizes the income variable (i.e., earning prospect) and the weights on economic capital noted as `WeightEC`. Therefore, the economic capital of an agent i is expressed as follows:

$$Economic_i = (Income_i + WeightEC_i * 0.1 * Income_i)/1600 \tag{2.2}$$

Similar to the cultural capital formula, the denominator is 1600, which is twice the maximum bound of income value, 800. Therefore, when both income and weight for economic capital are the highest, the agent's economic capital is 1, while it is 0 when both are the lowest.

Lastly, the agents' social capital is expressed with the **social** variable based on how many useful connections an agent has. In this thesis, the usefulness of a connection is defined by high cultural and economic capital. At the time step 0, where the agents are initialized, all the agents have zero social capital. Once all the initial agents' characteristics are formed, and the connections are made, their social capital is calculated based on how many people with either higher cultural capital or higher economic capital than the self an agent knows and is connected in the network. The extracted number of connections is divided by the total number of existing agents in the network at that time step, which can be expressed as follows:

$$Social_i = \Pr[Cultural_j > Cultural_i] \cup \Pr[Economic_j > Economic_i] \tag{2.3}$$

Where it compares the capitals of agent i with its neighboring agent j to calculate the social capital. Notably, the weight is not explicitly utilized in formulating an agent's social capital since the cultural and economic capitals already incorporate the appropriate weights. Moreover, as no other grouping mechanism is modeled, the approximation of an agent's social capital is achieved by considering the count of useful connections relative to the total number of neighbors. Although this approach does not account for the quality of connections, the exclusion of weight can be justified by the fact that social capital primarily relates to the potential – not actual – resources derived from one's networks (Bourdieu, 1986).

By summing up the three capitals discussed, one can obtain a value for the **capital** variable. As each capital is bounded between 0 and 1, the maximum value of capital is 3, and the minimum value is 0.

2.3 System Rules

At the system level, several rules are put in place to facilitate the simulation and, hopefully, its extension in future research.

Tuition fee is a variable that defines the minimum cost of education required for a child. This variable is set to express the economic constraints in deciding reproduction within a couple. Given that at least three stages of education (i.e., primary, secondary, and tertiary) are often funded by parents, only the couples whose total income prospect is more than three times the set tuition are allowed to reproduce in all the scenarios. Moreover, for this thesis, the default tuition fee is set rather low to 50 with 6.25% of the maximum income prospect to minimize the number of obstacles to procreation. This way, it is possible to better observe the relationship between mating patterns and educational mobility. Therefore, a couple with a household income prospect larger than 150 should be able to have children. The value for this variable can be easily changed in the simulation platform to explore.

Unfairness is a scalar, later determining a quantile of the population in the rank of agents' capital that can experience a higher likelihood for upward mobility. This variable is constructed to add complexity to the transmission rule of educational level from parents to children, and its function will become clearer in Section 2.4. This variable is called unfairness because, based on its degree, more agents can benefit from higher certainty to maintain the expected educational level, which is calculated based on the weights and the parents' educational levels (see the *influence cultural* variable in Section 2.4). More intuitively, it can be understood as how accessible higher education is to the agents in society. As a scalar, the value of unfairness can vary from 0 to 1. Therefore, the lower the value is, the more inclusive the society is in terms of educational opportunities.

At the stage of reproduction, **fertile women's ages** are defined to be between 20 and 45 inclusive. Despite the age range of all the agents between 15 and 59, the fertile women's ages are set to focus on the general patterns in marriage and reproduction. The birth rate per 1000 women aged between 15 and 19 in Belgium is low, for instance, 5.3 in 2021 (World Bank, 2023), and only about 1% of the total births by Belgian mothers in 2021 come from women aged between 15 and 19 (Statbel, 2021). While the mean age of Belgian mothers in 2021 is 31, the number of births by Belgian mothers aged over 46 years takes up only 0.1% of the total births in 2021 (Statbel, 2021). This reassures that limiting the mother's age for births allows the model to focus on the most frequent behavior and type of interactions between agents.

To facilitate the analysis, **the maximum number of children** that agents can have is restricted to two. A model allowing for more children may be more realistic, but this restriction is necessary as too many children make the data too large to handle and hinder a clear view of a family's dynamics. But note that this restriction can result in generating many single agents if the partner-seeking rules are too strict.

Similarly, a divorce is not considered in the model, meaning that the agents do not change their partners in the model. While considering divorces may be more realistic given that the influence of a stepparent may be considerable for some people, the thesis assumes that such influence can be treated together with other subsidiary influences from the environment (i.e., weights for capitals) and focuses on the relationship between the biological parents' educational levels and the children's educational levels. Lastly, to accelerate the dynamics between the agents, the model increases the agents' age by two years at each **time step**. Therefore, one time step can be interpreted as two years in human life.

Table 2.2 summarizes the default settings of the system-level parameters discussed above.

In the models, the total number of initial agents is 500, who have 500 connections on average and age two years per time step. These agents can have at most two children in their life, and only a couple holding at least three times the tuition fee of 50 as income (hence, 150) with a female partner aged between 20 and 45 can procreate. And the default unfairness factor is 0.2.

Table 2.2: Default settings of the system-level parameters

Parameters	**Default Setting**
Total number of initial agents	500
Average number of edges	500
Tuition fee	50
Unfairness	0.2
An increase in time step	2 years
The maximum number of children possible	2
Fertile woman's age range	20-45

2.4 Agent Behavioral Rules

As mentioned at the start of this chapter, two main behaviors should be defined: (1) How do agents marry? and (2) How does parents' educational level transmit to their children? Basic rules will be discussed first and then the extension of them will be discussed at the end of this section.

For the first question, the models focus on heterosexual marriages without divorce to facilitate the tracking of parents-children relationships. In the partner-seeking phase, potential partners of an agent (the seeker) are first gathered from neighbors. Here, it is crucial to realize that if the average number of edges (m) assigned to the system is higher and close to making a complete network, more neighbors are yielded at this stage. Among the neighbors, agents of a different gender who are not the spouse, child, or parent of the seeker, and the age gap with the seeker is equal to or less than 20 (i.e., $\left|\text{age}_i - \text{age}_j\right| \leq 20$ for the seeker i and a neighbor j) are considered potential partners. Once the potential partners are gathered, a score is calculated for each potential partner, and the potential partner who has the highest score becomes the final partner of the seeker. The score for a potential partner j is calculated as follows:

$$\begin{aligned} Score_j = & (1 - |cultural_i - cultural_j|)^{weightCC_i*0.1} \\ & * (1 - |economic_i - economic_j|)^{weightEC_i*0.1} \\ & * (1 - |social_i - social_j|)^{weightSC_i*0.1} \end{aligned} \tag{2.4}$$

Where the values for the weights for cultural, economic, and social capital, weightCC, weightEC, and weightSC, are $1, \ldots, 10$ and the values for cultural, economic, and social are $0, \ldots 1$. This formula is inspired by the mating value formula in Grow and Van Bavel (2015). The formula calculates a score based on the differences in cultural, economic, and social capital between the seeker and each of his/her potential partners as well as the seeker's weights. Therefore, while generally a smaller difference between the seeker's and the potential partner's capitals is preferred, if the seeker weighs more on a certain capital, the larger the difference in capital is, the more punishment it applies. For example, for the same difference between capital, e.g., 0.8, the score differs when the weight is 1 ($(1-|0.1-0.9|)^{0.1} = 0.617$) and when the weight is 9 ($(1-|0.1-0.9|)^{0.9} = 0.235$). In other words, because the seeking agent finds the capital important, even a smaller difference matters to dislike a potential partner.

On top of these general conditions, additional conditions are applied to encompass the age preferences of each gender that Grow and Van Bavel (2015) highlighted: women prefer slightly older men than them, and men prefer women in the mid-20s. Accordingly, if the female potential partners are between 23 and 27 for male seekers and if male potential partners are between 0 to 10 years older than the female seeker, an additional 0.3 score is added for those potential partners. Figure 2.2 summarizes the partner-seeking process.

For the second question regarding the transmission of educational attainments, one should focus on the phase where a child is born and how the characteristics of the child are defined. At the reproduction phase, the child's **generation** is defined as the maximum parents' generations plus one (i.e., $max(generation_i, generation_j) + 1$ for a parent agent i and his/her spouse j). Similarly, the **cohort** of the child agent is one plus the maximum value of parents' cohorts (i.e., $max(cohort_i, cohort_j) + 1$). When the child is born, his/her **age** is 0, and as the initial agents, his/her **state** is set to be single. As the initial agents, the child agents' **weights** for cultural, economic, and social capital are defined randomly as an integer between 1 and 10.

While the social capital follows the same formula as for the initial agents (see Equation 2.3), the education and income variables that influence cultural and economic capital follow unique transmission rules. To determine a child's **educational attainment**, a variable called **influence cultural** is created based on the cultural capital of the agent i and his/her spouse

Figure 2.2: General flow of the partner-seeking process from the perspective of a potential partner. The score is calculated by Equation 2.4.

j as follows:

$$Influence\_Cultural = round(max(cultural_i, cultural_j) * 5) \tag{2.5}$$

If the sum of the parents' capital (i.e., $capital_i + capital_j = cultural_i + economic_i + social_i + cultural_j + economic_j + social_ij$) is larger than $6 * unfairness$, unfairness being a floating number between 0 and 1, then the child's educational level is chosen randomly between min(influence cultural, 5) and 5. Here, if the influence cultural variable is already 5, the child's educational level is defined as 5. If the sum of the couple's capital is smaller than

6*unfairness, the child's education is chosen randomly between 1 and min(influence cultural, 5) or is defined as 1 if the influence cultural is already 1.

Here, the unfairness is a scalar representing a quantile of the population as introduced in Section 2.3 and 6 is the maximum capital value a couple can hold. For instance, with the unfairness of 0.2, the model sends the top 80% of the population in terms of household capital to have their children's education to be *at least* the calculated influence cultural value while the bottom 20% will have *at most* the calculated influence cultural value as the child's education. As the scalar value increases, the less privileged children experience more difficulty having a higher educational level than their parents. Finally, the child's **cultural** capital is calculated the same as for the initial: $(education + (weightCC/10 * education))/10$.

As for the initial agents, a correlation between income prospects and education is considered for the child agents. The **income** of the child is defined by a variable called **influence income** based on the parents' income and their weights for economic capital as follows:

$$Influence\_Income = max(income_i, income_j) * max(weightEC_i, weightEC_j) * 0.1 \quad (2.6)$$

for an agent i and his/her spouse j who are the parents of the child agent being created. It is noticeable that this will usually result in a slightly smaller value than the highest income value of the parents unless the maximum value of the parents' weights is 10, which returns the maximum income value of the parents.

Once the influence income value is estimated, a condition is set to divide the child agents' behaviors into two based on their educational level. If the child's educational level is the same as or higher than the highest educational level of the parents, or the sum of the parents' weights for economic capital is equal to or larger than 12 (i.e., either upward mobility reached or parents have a high economic capital), the child agent's income is randomly chosen between the estimated influence income value and the maximum income value, 800. If the said conditions are not met, the child agent's income is determined from the range between the minimum income value, 0, and the influence income value. Lastly, like the economic capital for the initial agents, the child's economic capital is calculated as $(Income + WeightEC * 0.1 * Income)/1600$.

It is crucial to recognize that transmitting educational levels from one generation to the next based on capital is less involved than specifying rules for a child's educational level based on the parents' educational backgrounds. In the latter approach, a specific set of transmission rules must be entirely or partially known. For instance, one can give a rule such that children of upper tertiary-educated parents have a 90% of chance to also have upper tertiary education. In contrast, the former aims to introduce additional complexities in the process to mitigate the impact of these transmission rules, which are unknown. Specifically, three key factors come into play during the transmission process, contributing to the intricate relationship between parental educational levels and their child's educational level: i) capital, ii) weights, and iii) unfairness.

Firstly, the *capital* variable is a composite of three types of capital: cultural, economic, and social capital. Therefore, the transmission rule is dependent on these three factors, reflecting how the educational background of parents, their financial capacity to support their child's education, and the connections they possess to facilitate their child's success (e.g., tutoring) play a significant role altogether in defining the child's education.

Considering the added complexity introduced by assigning *weights* to each capital variable, the transmission rule may lead to situations where even parents with lower educational levels can have highly educated children, depending on their other capitals and the respective weights assigned to them.

Lastly, the model incorporates the *unfairness* factor, which further complicates the transmission rule. It determines who can gain easier access to upward mobility and, as a result, reinforces or challenges existing disparities in educational opportunities. This way, the child's educational level is not solely dependent on the parents' educational levels.

Taking into account all the variables and rules described above, the model operates as follows. Initially, agents are created in a network, and each agent's characteristics are defined. Once the network is established and populated with agents possessing specific characteristics, the social capital of the initial agents is determined, a process that requires the completion of the network. Subsequently, the model advances the agents' age by two years at each time step. During each time step, certain conditions are evaluated. If an agent is older than 65 or belongs to the oldest cohort group (i.e., the lowest cohort number), there is a 40% chance of that agent being removed from the network. If not removed, single agents have the opportunity to meet a partner, and subsequently, married agents engage in reproduction. For reproduction to occur, female agents must fall within the inclusive age range of 20 to 45 years. After reproduction takes place, the resulting child agents undergo a similar cycle of seeking partners and reproducing with their chosen partners. The general flow of the model operation is summarized in Figure 2.3.

Scenario Expansions

From the base model (hereafter, Scenario 1) with the basic rules employed as described above, the model is slightly modified to explore the relationship between educational assortative mating and educational mobility. Firstly, Scenario 2 is modified from Scenario 1 by giving

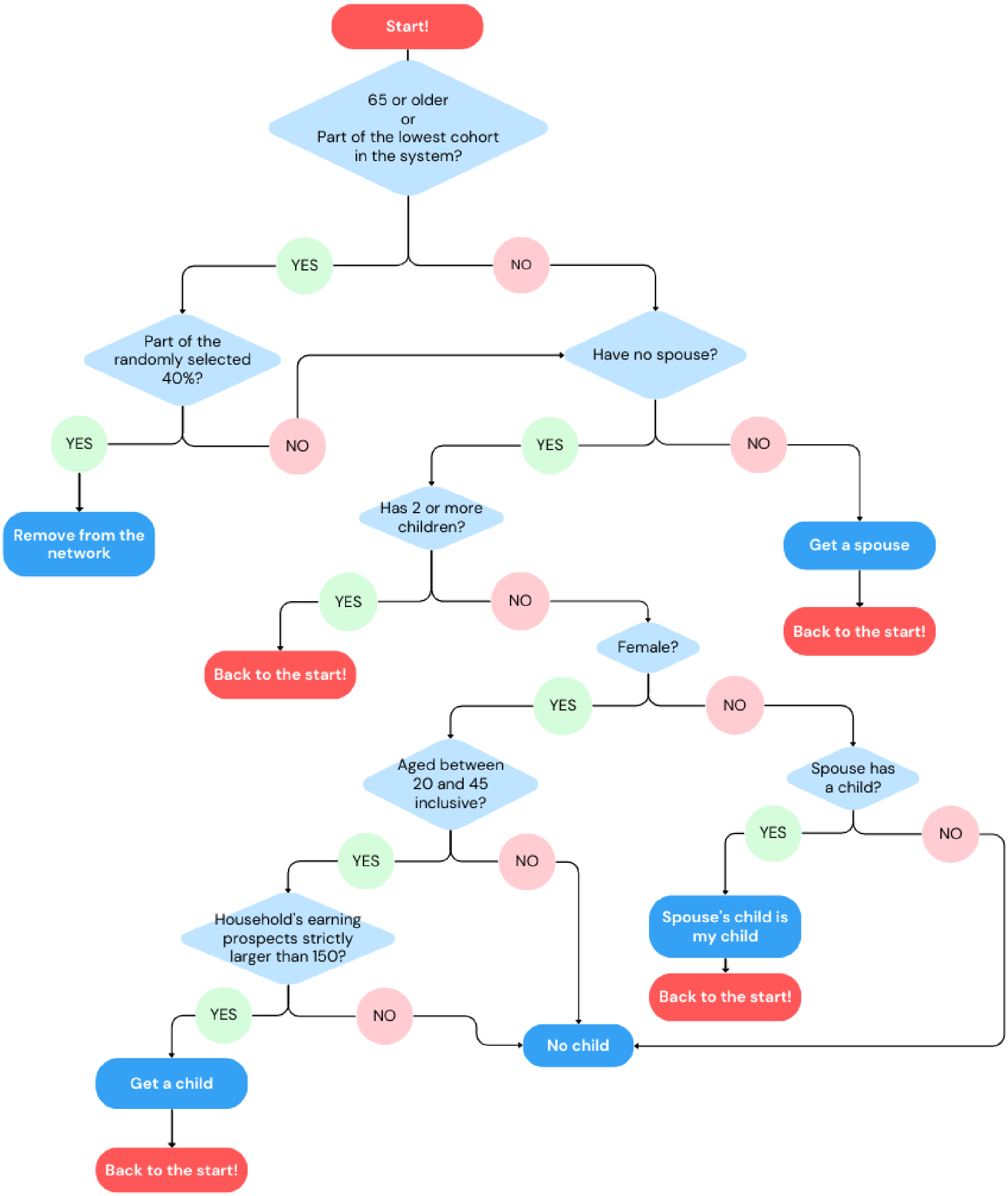


Figure 2.3: General flow of the model operation, from the perspective of an agent.

more incentives to equally educated potential partners. When the potential partners possess the same educational level as the seeker, their score is increased by 0.5, while when they satisfy the age preferences by gender, the incentive is still 0.3. A slightly higher bonus is given to the equally educated potential partners than those satisfying the age preferences because the main factor of educational mobility the thesis aims to discern is mating patterns.

Scenario 3, on the other hand, gives more incentives to differently educated potential partners. Similarly to Scenario 2, when the potential partners' educational level differs from the seeker, their score is increased by 0.5, while when they satisfy the age preferences by gender, the incentive is still 0.3.

In order to verify the pure impact of homogamy and heterogamy on educational mobility, two extreme cases are additionally made. Scenario 4 simulates true homogamy, where everybody marries an equally educated partner, and Scenario 5 simulates true heterogamy, where everybody marries a differently educated partner.

Here, one may notice that homogamy is not necessarily characterized by 'equally' educated couples but 'similarly' educated couples. Nevertheless, Scenarios 2 and 4 restrict homogamic marriages to equally educated couples because the model only allows 5 categories of educational attainments. Therefore, it is assumed that the small differences between partners can be captured by their cultural capital rather than their educational attainments.

2.5 Analysis

Four measures exploring absolute and relative mobility in van der Weide et al. (2021) are adopted to investigate educational mobility from the simulated data in this thesis. With the main goal of the thesis to analyze educational mobility, as mentioned in Section 1.2, two types of mobility are investigated: absolute mobility in terms of educational attainments and second is relative mobility in terms of cultural capital.

As will be clarified with the relative mobility measures formula, using a categorical variable, educational attainments, to calculate relative mobility measures has a limitation. As the relative mobility measures utilize correlation coefficients or ranking, a larger spectrum of educational values is required. For the correlation coefficients, an ordinary logistic regression may be used, which allows a categorical variable as a dependent variable. However, since there are only 5 educational levels, creating 8 dummy variables to express the mother's and father's educational level to explain the child's educational level results in overfitting. Here, utilizing a maximum educational level of the parents does not facilitate the process because as the cohort advances, the maximum value of parents' educational level becomes predominantly 5 for all the agents. No distinction in the explanatory variable leads to an error in the model. A similar limitation prevails when ranking the agents based on educational attainments. Since the range of the educational levels is small, an adequate distinction between the agents becomes difficult.

For these reasons, the cultural variable will be used when employing relative mobility measures, while categorical educational attainment is still apt for absolute mobility measures.

Here, the maximum educational value of the parents will be utilized as in van der Weide et al. (2021). This decision was supported by the following reasons: (1) "It provides a more accurate measure of the parental human capital that can be transferred to the parents' children" (p.5) ; (2) it arguably better represents the household resources that contribute to the human capital growth of the children ; (3) it may be a more stable measure given its robustness to the expanding female educational attainment (cf., the reversal of the gender gap in education) and the increasing educational assortative mating (van der Weide et al., 2021, 5).

Absolute Mobility Measures on Educational Attainments

The following two measures will be calculated in the analysis to assess absolute intergenerational mobility in education: (1) the probability of a child's educational level being strictly higher than both parents' educational levels, given that the parents' educational levels are strictly lower than the upper tertiary education (i.e., level 5). (2) the probability of children with strictly higher educational levels than the parents who do not have an upper tertiary education or with the upper tertiary education whose either parent has an upper tertiary education. Unlike World Bank (van der Weide et al., 2021), the upper tertiary is considered a reference category instead of tertiary education. This is because in van der Weide et al. (2021), the lower and upper tertiary levels are not distinguished. Moreover, since more than half of 25-34 year-olds have a tertiary education in Belgium (OECD, 2022a), upper tertiary education appears to distinguish the population more than lower tertiary education.

To formalize, the first measure is as follows:

$$\Pr(Education_{child} > Education_{parents} | Education_{parents} < 5) \tag{2.7}$$

Where $Education_{parents}$ implies that the condition must be satisfied for both mother and father.

The second measure comprises the part that is the same as the first measure, and the second part takes into account the cases where the children of an upper tertiary-educated parent keep the parent's educational level. As such, the second measure considers all the individuals who have progressed to the upper tertiary level as an indication of upper mobility.

$$\begin{gathered}\Pr(Education_{child} > Education_{parents} | Education_{parents} < 5) \\ \cup \\ \Pr(Education_{child} = 5 | Education_{mother} = 5 \text{ or } Education_{father} = 5)\end{gathered} \tag{2.8}$$

Relative Mobility Measures on Cultural Capital

The following two measures are employed for the relative mobility measures: (1) the regression coefficient from children's educational levels regressed on the parents' educational levels ; (2) the expected educational rank of children born to parents in the bottom half.

For the first measure, a linear model is used to obtain a regression coefficient, β, as follows:

$$Cultural_{child} = \beta_0 + \beta_1 \cdot Cultural_{parents} + \epsilon \tag{2.9}$$

In this equation, β represents the extent to which parents' cultural capital predicts the child's cultural capital. A higher β value indicates a more constrained social mobility. To prevent overfitting, the equation uses the maximum cultural value of parents denoted as $Cultural_{parents}$. The term ϵ in the equation represents the residuals, which are the unexplained variations in the child's cultural capital that cannot be accounted for by the parents' cultural capital in the linear model.

The second measure ranks children's cultural capital, considering only those whose parents have a maximum cultural capital ranking in the bottom half.

$$E(rank_{child} | rank_{parents} < 50) \tag{2.10}$$

In addition to these four measures, absolute mobility in education will be further investigated per group defined by the parents' educational levels (e.g., a couple between an upper tertiary education and a lower secondary education). Moreover, the thesis will also look at the evolution in the population-wise average educational level over generations. By inquiring into different aspects of educational mobility in varying settings for mating strategies, the thesis aims to discover the relationship between educational assortative mating and educational mobility.

Scenario Comparisons

After analyzing educational mobility, the scenarios' distinctiveness will be assessed within each cohort and in general. By examining their uniqueness across cohorts, one can gain valuable insights into the far-reaching impact of a mating pattern on mobility. Furthermore, inquiring into the general effect of each scenario will shed light on how different mating patterns influence the overall evolution of educational mobility.

Firstly, Analysis of Variance (ANOVA) are used to evaluate the cohort-wise scenario differences. Unlike t-tests which allow a comparison between two groups, one-way ANOVA permits comparisons between two or more groups. For the comparisons, a linear model is fitted at each cohort as follows:

$$y_i = \beta_{0i} + \sum_{j=1}^{J} \beta_{ji} Scenario_j + \epsilon_i \tag{2.11}$$

Where Cohort i is {3, ..., k} and Scenario is {1, ...,J}.

To investigate the general effect of each scenario, a linear mixed model is utilized, which allows for dependence or hierarchical structures in the observations (Verbeke and Molenberghs, 2013). The `lmer4` R package will be used for conducting this model. It is important to note that multiple replications will be conducted for each scenario to increase the analysis's validity when analyzing the data. The analysis requires more than one replication of each scenario since it is crucial to account for the possibility that the findings and conclusions depend on the randomness of the simulation. Here, it should not be confused that the same numbered replications are not assumed to be correlated across scenarios since they come from different scenarios, although each scenario has the same values (say, 1, ..., 10), indicating the number of the replication.

To facilitate the statistical analysis, the linear mixed model is used on aggregated data where each scenario has an average value of replications for each cohort. Each absolute mobility data set contains repeated measures, as each data set has values for each cohort, and the observations across cohorts are assumed to be correlated due to the transmission rules for education between generations. Given a data set involving all the scenarios' aggregated information on the average educational mobility, a simple linear mixed model will be composed as follows:

$$y_{ij} = \beta_0 + \beta_1 Cohort_{ij} + u_{0j} + \epsilon_{ij} \tag{2.12}$$

Where the outcome variable for a $Cohort_i$ and $Scenario_j$ is the probability for the children to have a better educational level than their parents (i.e., absolute mobility), using the aforementioned two measures. The term β_0 is the overall average across the cohorts, and u_{0j} is the scenario-specific random intercept. Lastly, ϵ_{ij} represents the residuals that are not explained by the overall and scenario-specific averages. With this equation, one can discern whether the scenarios differ in the absolute measures by comparing it to the null model without the random intercept, u_{0j}. The comparison will be done utilizing the log-likelihood ratio test.

The log-likelihood ratio test for a nested model in R is performed as follows:

$$LRT = -2log_e(\frac{\mathcal{L}_s(\hat{\theta})}{\mathcal{L}_g(\hat{\theta})}) \tag{2.13}$$

Which is the ratio of the likelihood function of the simpler, null model (s) and the general model (g) (MacKenzie et al., 2018). Asymptotically, this test statistic is chi-squared distributed with 1 degree of freedom as only one parameter is added in the general model from the null model (i.e., u_{0j}) (MacKenzie et al., 2018; Lewis et al., 2011). In the model selection process, if the test statistic is significant with a p-value lower than 0.05, the more complex model, g is chosen.

This chapter delved into the methodologies used to build agent-based models and analyze educational mobility with simulated data. The next chapter will discuss the results of the analysis for each scenario.

Chapter 3

Results

This section will discuss the results of the simulation by agent-based models. The three scenarios discussed in Section 2.4, together with the two extreme cases, are analyzed. Scenario 1 is the base scenario where some parameters for the initial agents are tuned with Belgian data. Scenario 2 rewards the homogamous mating patterns between agents, whereas Scenario 3 rewards the heterogamous marriages. The two extreme cases are true homogamy (Scenario 4), and true heterogamy (Scenario 5), where everybody must mate homogamically and heterogamically, respectively.

Given that the conclusion can depend on the randomness in each scenario, 10 replications have been generated for each scenario as explained in Section 2.5. The results of the 10 replications are aggregated to report and compare with other scenarios. As introduced in Chapter 2, some restrictions are applied to the system in the agent-based models. Specifically, agents are restricted to having a maximum of 2 children in order to have a clear view of the association between the parents and the children agents. As a consequence, the population does not sustain for long and eventually dies around the age of 65 in all the scenarios, which may affect the conclusions regarding educational mobility. Moreover, the unfairness factor is arbitrarily set to 0.2, assuming that it will represent a society such as Belgium with a relatively accessible higher education. A different unfairness factor can easily change the shape of the evolution of educational mobility over cohorts. Therefore, verifying and understanding the impact of these restrictions on the conclusion through sensitivity analyses is crucial to draw a valid conclusion.

This chapter will begin with the findings in terms of general trends and educational mobility, and then mobility in cultural capital will be discussed. In these sections, the two extreme cases are discussed to demonstrate the impact of mating patterns on mobility by eliminating other possible factors influencing mobility. The three scenarios (Scenario 1-3) will exhibit more nuanced results. Once the main findings are elaborated, the results of sensitivity analyses will be presented to verify the validity of the main findings.

3.1 Average vs. Mobility

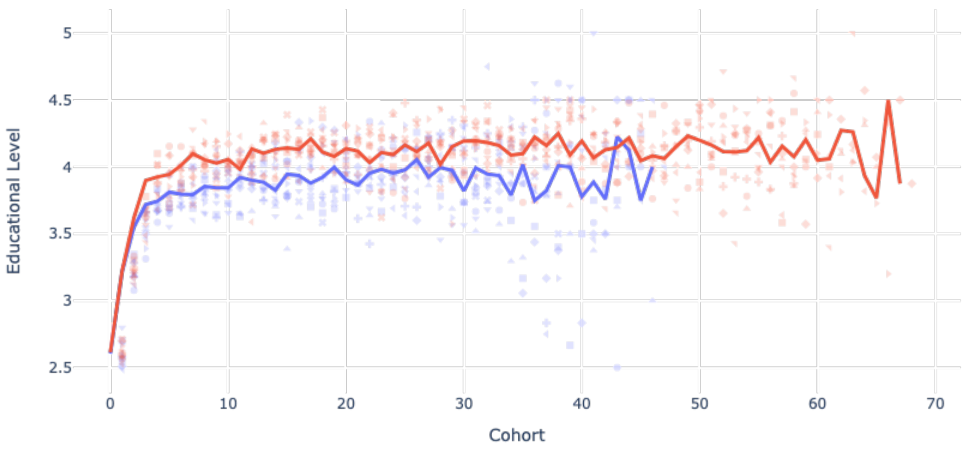
The first section delves into the results of the agent-based models in terms of population-wise averages and absolute mobility as well as the relationship between the population-wise average and mobility. Moreover, a detailed investigation of how people of different background experience the improvement or decline in the population-wise averages will be offered. This chapter will be subdivided into two parts, the first discussing the results of the extreme cases (Scenario 4-5) and the second for the three scenarios (Scenario 1-3).

As already outlined in Section 2.5, two absolute mobility measures are applied. Measure 1 examines cases where both parents have educational levels strictly lower than upper tertiary education (i.e., < 5), and their children have advanced their educational levels beyond those of their parents (as defined in Equation 2.7). Measure 2 additionally includes the upper tertiary-educated agents whose parents are also upper tertiary-educated.

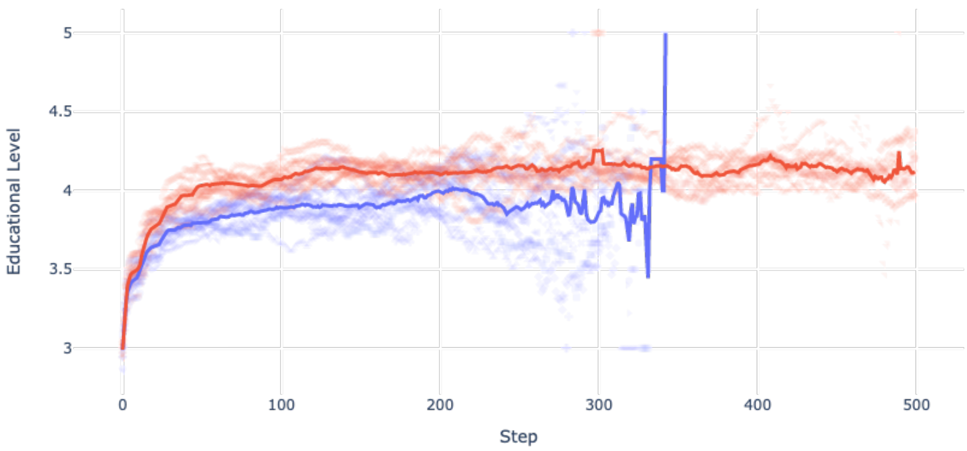
For both measures, the probabilities are aggregated to obtain mean probabilities that summarize the results from 10 replications per scenario. The error bands in the graphs represent the 95% confidence intervals with the standard errors between replications. To generate these graphs, the initial agents with a generation value of 1 are excluded from the data as they serve as parents in subsequent generations. Consequently, the minimum cohort number presented in the graphs is 3. Given the smaller population size and greater variance in population size in the later cohorts, particularly cohorts 50 and above, the focus will be on cohorts prior to Cohort 55.

Extreme Cases

The two extreme scenarios that explore the effects of strict heterogamy or homogamy will be considered first in this section, which are referred to as the true heterogamy (Scenario 4) and true homogamy (Scenario 5) scenarios. Investigating these extreme mating patterns allows one to gain insights into their impact on educational mobility. Figure 3.1a illustrates that averaging over the 10 replications, the true homogamy scenario exhibits a lower mean educational level across cohorts, except for a few cohorts towards the end. Similarly, Figure 3.1b reveals that the true homogamy scenario, averaging over the 10 replications, has a lower mean educational level across time steps, except for the last few time steps. Again, it should be reminded that 1 time step represents 2 years in the agent's lifetime. Notably, the true homogamy scenario concludes earlier than the true heterogamy scenario in terms of both cohorts and time steps. This observation aligns with the notion that exclusively seeking partners with similar educational backgrounds is more restrictive than the alternative scenario.



(a) Mean educational levels over cohorts



(b) Mean educational levels over time-steps

Figure 3.1: Population-wise mean educational levels are depicted, with the true homogamy scenario shown in blue and the true heterogamy scenario in red. The dots illustrate the mean educational level of each of the 10 replications, while the bold lines connect the scenario-wise means at each point. These scenario-wise means are calculated as the averages of the 10 dotted values. The true heterogamy scenario generally shows higher mean educational levels across cohorts and time steps and a longer-lived population than the true homogamy scenario.

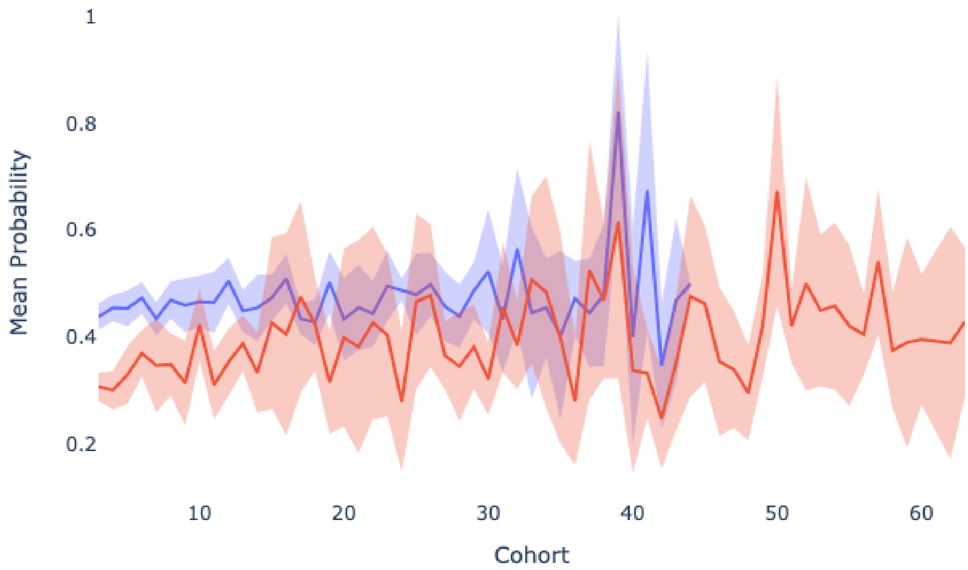
Regarding absolute mobility in education shown in Figures 3.2a and 3.2b, for both measures, the true heterogamy scenario lies mostly below the true homogamy scenario across cohorts with considerably larger 95% confidence intervals. Moreover, except for the last few cohorts containing fewer agents, the mean probabilities for the true homogamy scenario appear relatively stable, around 0.5, looking at the two absolute measures. In contrast, the mean probabilities by the two measures for the true heterogamy scenario exhibit gradually increasing trends – albeit not linearly – starting near 0.3 and later soaring close to 0.7.

In order to investigate who benefits from mobility in both extreme scenarios and whether this differs between the scenarios, agents are grouped into 15 groups based on their parents' educational levels. Although these groups do not distinguish who in the married couple has a higher educational level than the other, 15 groups instead of 25 are created to facilitate the analysis. Figure 3.3 presents the average educational levels of each group across cohorts, revealing a stark contrast between the true homogamy and heterogamy scenarios. In the true homogamy scenario, high-ranking groups with at least one tertiary-educated parent exhibit higher mean educational levels compared to groups with parents at lower educational levels. In contrast, the true heterogamy scenario displays groups that are relatively more closely aligned compared to the true homogamy scenario. Furthermore, the ordering of the groups appears to vary, although the high-ranking groups predominantly occupy the top tier.

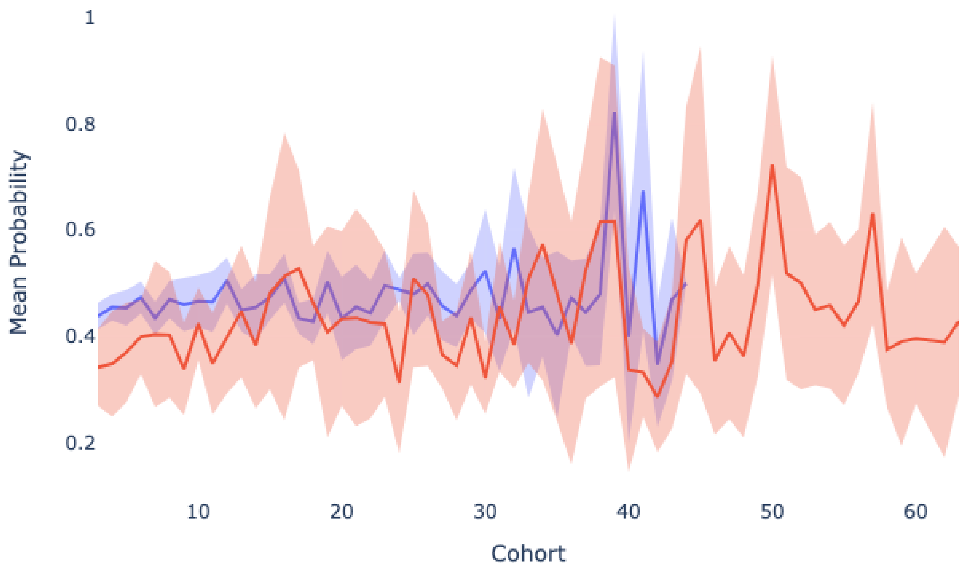
The examination of extreme cases highlights that heterogamy is associated with higher average educational levels across cohorts and time steps compared to homogamy. In terms of educational mobility, homogamy ensures a more stable pattern of intergenerational mobility, with relatively smaller variances observed between replications and generally higher mobility levels than heterogamy. Conversely, the mean probability for children's upward mobility varies much more across the replications in the true heterogamy scenario. These variances are particularly pronounced when additionally considering cases where children of highly educated parents maintain their upper tertiary degrees (i.e., Measure 2 of absolute mobility).

Furthermore, when analyzing the population-wise average educational levels within groups defined by the parents' educational background, the true homogamy scenario exhibits more noticeable differences between groups in each cohort (i.e., greater between-group variance). This finding aligns with the previous observations of generally lower population-wise average educational levels and stable mobility patterns over cohorts in the true homogamy scenario. In other words, children of highly educated parents tend to attain higher education, while children of less educated parents tend to achieve lower levels of education. This distinction results in a lower average educational level within cohorts but a consistent pattern of mobility compared to the true heterogamy scenario.

Lastly, when comparing the two scenarios at each cohort and overall, their distinctiveness is observed. Especially in the earlier cohorts before Cohort 20 with Measure 1, most of the absolute mobility values appear significantly different at a significance level of 0.05. With Measure 2, some significant differences appear in the early cohorts but less frequently than



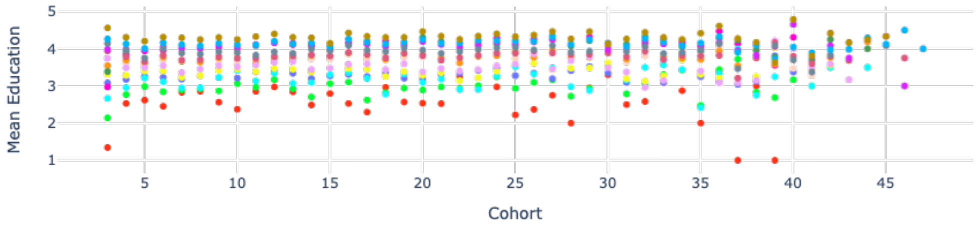
(a) Measure 1: The probability of less privileged children obtaining a higher educational level than their parents.



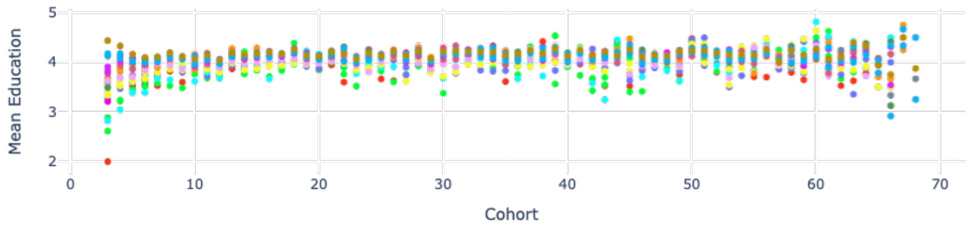
(b) Measure 2: It adds the probability of children with upper tertiary-educated parents maintaining the status to Measure 1.

Figure 3.2: The absolute mobility in education over cohorts averaged over 10 replications of each extreme scenario with the true homogamy scenario in blue and the true heterogamy scenario in red. 95% error bands express the variability across the replications of each scenario. The true heterogamy scenario appears to have generally lower absolute mobility with both measures but a larger error band.

with Measure 1. Moreover, the effect of scenarios seems to resonate throughout the cohorts, albeit sparsely. Some later cohorts still show significant differences between the scenarios (e.g.,



(a) True homogamy scenario



(b) True heterogamy scenario

Figure 3.3: The mean educational levels of groups characterized by parents' educational levels averaged over 10 replications of the two extreme scenarios are depicted. The groups are colored as in the legend above. E.g., [1,2] means one of the parents has an educational level of 1, and the other has 2. Various groups share a higher mean value in the true heterogamy scenario, while each group seems to maintain its position throughout cohorts in the true homogamy scenario.

Cohort 30 and 41 with measures 1 and 2).

The overall effect of scenarios is investigated with a linear mixed model as described in Section 2.5. As shown in Table 3.1, introducing the scenarios to distinguish the observations improves the model using both measures with a significance level of 0.5.

Three Scenarios

Analyzing the trends of mean educational levels across the replications of the three scenarios (Scenario 1-3) at each cohort group in Figure 3.4a, it is evident that Scenario 3 generally remains above the other scenarios from Cohort 2, creating a substantial gap at Cohort 3. This trend continues until Cohort 40, with some exceptions where Scenario 1 slightly surpasses Scenario 3 at Cohorts 23, 31, 36, and 38 and where Scenario 2 briefly exhibits a higher mean educational level at Cohorts 36 and 37. However, beyond Cohort 40, the number of agents in each replication decreases, sometimes falling below 100 and even as low as 30, while Cohort 1 still comprises over 200 agents. As previously mentioned, this disparity in replication sizes can impact the variation between replications within a scenario, consequently affecting the overall mean of the scenario and complicating the comparison among the three scenarios.

Table 3.1: Log-likelihood ratio tests for linear mixed models

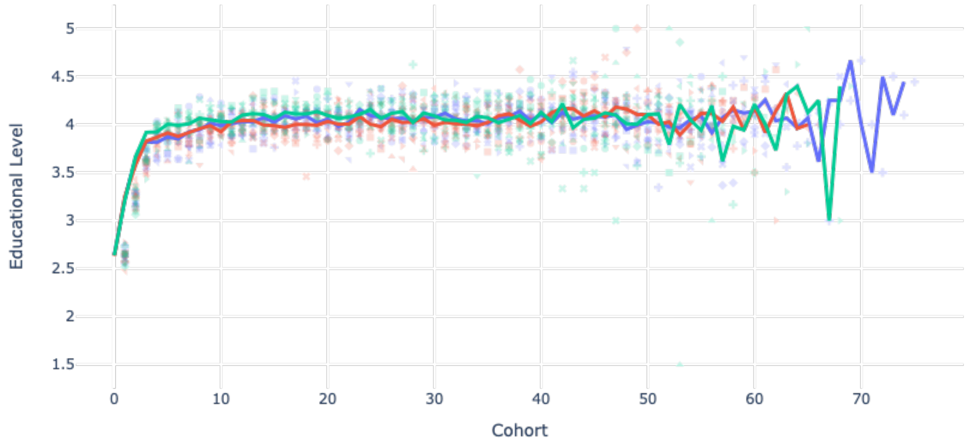
	Nr.Parameters	logLik	AIC	LRT	Df	$Pr(> \chi^2)$
Extreme Cases						
Measure 1						
Null	4	82.433	-156.87			
(1 | Scenario)	3	73.997	-142	16.871	1	4.00E-05 ***
Measure 2						
Null	4	79.072	-150.14			
(1 | Scenario)	3	76.836	-147.67	4.4709	1	0.03448 **
3 Scenarios						
Measure 1 (Singularity issue)						
Null	4	251.76	-495.53			
(1 | Scenario)	3	251.76	-497.53	0	1	1
Measure 2						
Null	4	236.53	-465.06			
(1 | Scenario)	3	226.13	-446.26	20.794	1	5.11E-06 ***

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

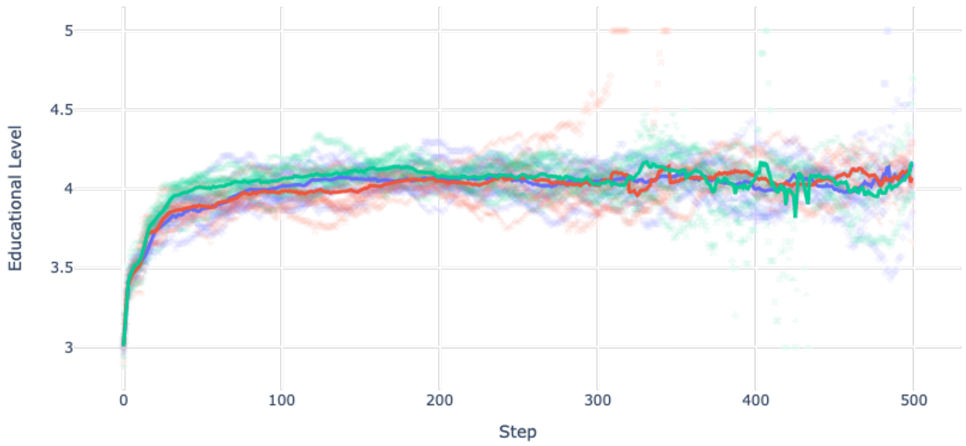
Examining the mean educational levels over time steps in Figure 3.4b, a similar trend emerges where Scenario 3 mostly maintains a higher mean educational level compared to the other scenarios until around time step 300. Scenario 2 initially follows a similar pattern as Scenario 1 until time step 100, but between time step 100 and 200, a considerable gap between Scenario 2 and the other scenarios becomes apparent. Beyond time step 300, the average number of agents per scenario drops below 300, and in extreme cases, even below 100.

Based on the investigation of mean educational levels across cohorts and time steps, Scenario 3 consistently yields higher mean educational levels on average compared to the other scenarios. Scenario 2 exhibits a similar pattern to Scenario 1 but consistently maintains the lowest average educational level. These results align with the findings in the extreme cases above, where true heterogamy demonstrated higher mean educational levels than true homogamy.

Figure 3.5a illustrates the mean probability for the children of relatively less educated parents (< 5) to achieve upward educational mobility across cohorts for each scenario (mean Measure 1). Initially, Scenario 2 consistently exhibits the highest mean probability, occasionally surpassed by Scenario 1. Scenario 3 consistently manifests the lowest mean probability



(a) Mean educational levels over cohorts

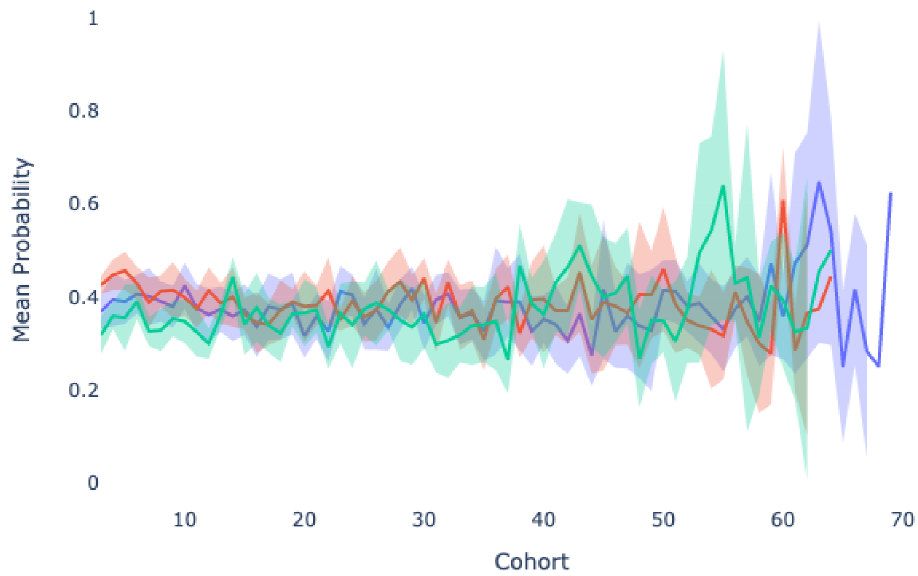


(b) Mean educational levels over time-steps

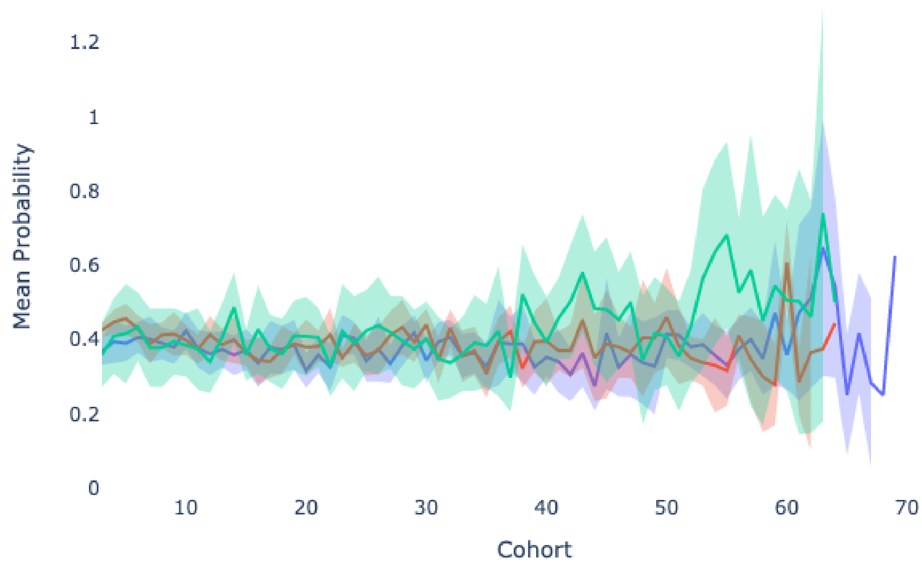
Figure 3.4: Population-wise mean educational levels are depicted, with Scenario 1 shown in blue, Scenario 2 in red, and Scenario 3 in green. The dots illustrate the mean educational level of each of the 10 replications, while the bold lines connect the scenario-wise means at each point. These scenario-wise means are calculated as the averages of the 10 dotted values. Scenario 3 generally shows the highest mean educational levels across cohorts and time steps, while Scenario 2 stays rather low in both graphs.

among the three scenarios before Cohort 40, with a few spikes observed at Cohort 14, 16, 25, and 35. Gradually, between Cohort 35 and 50, the mean probability in Scenario 3 shows a further increase beyond the other scenarios. However, this notable increase may be influenced by the disparities in the number of agents across cohorts in the scenario.

Figure 3.5b presents the mean probability of children, including the cases where the children of upper tertiary-educated parents achieve the upper tertiary level (mean Measure 2). The probabilities among the three scenarios appear even more similar compared to Figure 3.5a. However, the 95% error band of Scenario 3 are significantly larger than those of Scenario



(a) Measure 1: The probability of less privileged children obtaining a higher educational level than their parents.

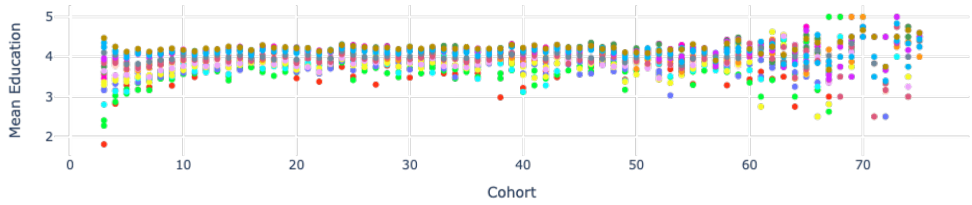


(b) Measure 2: It adds the probability of children with upper tertiary-educated parents maintaining the status to Measure 1.

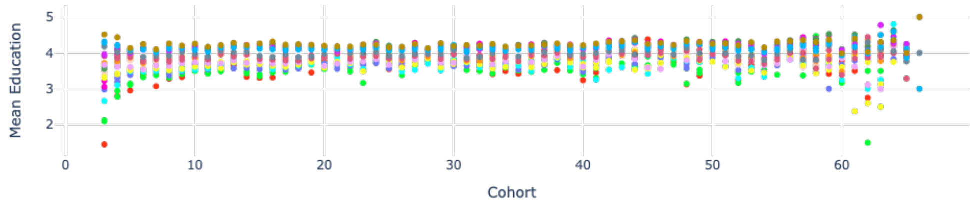
Figure 3.5: The absolute mobility in education over cohorts averaged over 10 replications of each scenario with Scenario 1 in blue, Scenario 2 in red, and Scenario 3 in green. 95% error bands express the variability across the replications of each scenario. Scenario 3 starts with the lowest mobility and Scenario 2 with the highest, and Scenario 3 with Measure 2 shows a wider error band than other scenarios.

2. Aligned with the results in extreme cases, this difference confirms that in societies where heterogamy is favored, a greater variance in mobility prevails compared to societies where

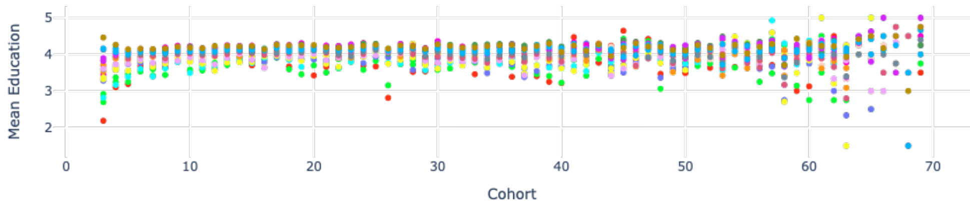
homogamy is preferred.



(a) Scenario 1



(b) Scenario 2



(c) Scenario 3

Figure 3.6: The mean educational levels of groups characterized by parents' educational levels averaged over 10 replications of the three scenarios are depicted. The groups are colored as in the legend above. E.g., [1,2] means one of the parents has an educational level of 1, and the other has 2. Scenario 3 appears to have smaller gaps between groups throughout the cohorts than other scenarios, denoting more diversity in groups contributing to higher population-wise educational levels exists.

To investigate who actually benefits from high educational mobility, mean educational levels are calculated for groups defined by parents' educational levels and presented in Figure 3.6. Figure 3.6 reveals that the group of agents with both parents holding upper tertiary education ([5, 5]) consistently attains the highest mean educational level across all three scenarios. However, in Scenario 3, the gaps between the groups are narrower compared to the other scenarios, and more groups overlap with the highest mean educational level. Scenarios 1 and 2, on the other hand, exhibit larger distances between the groups, with Scenario 2 showing a particularly wider variance among the groups. Comparing Scenarios 2 and 3, the results resonate with the results in extreme cases where the true heterogamy scenario showed a narrower gap between the groups.

The differences in variance between the scenarios are calculated for cohorts below 50, taking into account the large disparities in the sample sizes across the replications, resulting in a total of 47 cohorts. With few exceptions, Scenario 3 consistently demonstrates smaller variances compared to Scenario 2. Even in cases where Scenario 3 shows larger variances, the difference remains relatively small (ranging from a minimum of 4.34E-04 to a maximum of 5.39E-02). Similarly, when compared to Scenario 1, Scenario 3 generally exhibits smaller variances with only 19 cohorts showing exceptions, while Scenario 2 tends to have larger variances except for 14 cohorts.

The smaller variances observed in Scenario 3 resonate with the greater variances observed in the absolute mobility measures discussed earlier. Scenario 2 characterized by a preference for homogamy, displays a smaller variance in educational mobility. This can be attributed to a more structurally stable distribution of higher education attainments in the population, and this is reflected in a larger variance in the mean educational levels of the different educational backgrounds.

Lastly, more cohorts appear significantly distinct from one another when comparing Scenario 2 and 3, compared to when comparing all three scenarios. Specifically, Cohort 3 showed the most significant differences with Measure 1 between the two scenarios but not with Measure 2. From Cohort 8 onwards, the differences became increasingly less significant, although some significance persisted in later cohorts (e.g., Cohort 37 and 38 in measures 1 and 2). Compared to the extreme cases discussed above, these scenarios do not differ as prominently, aligning with the fact that the extreme cases do not allow a mix of homogamy and heterogamy in the data, while the three cases relax such regulations.

With linear mixed models discerning the significance of the hierarchical effect of scenarios on the two absolute mobility measures, the scenario appeared to be significant with Measure 2 but not with Measure 1. As described in Table 3.1, the model for Measure 1 suffered from a singularity issue, which often indicates overfitting in the model. Here, an alternative model is not constructed to ensure a consistent testing methodology across all analytical data sets. Due to the singularity, the log-likelihoods of the null model and the model for Measure 1 are equal, resulting in a p-value of 1. Nevertheless, based on Measure 2, one can discern that introducing the scenario variable in the model still improves the model. Thus, the scenarios matter in understanding educational mobility in terms of Measure 2.

3.2 Mobility in Cultural Capital

The utilization of cultural capital, as discussed in Section 2.5, allows for the application of methods for assessing relative mobility in education. The inclusion of the cultural capital variable offers the advantage of capturing the transmission of one's educational status in a more comprehensive manner. In this section, the results of mobility in cultural capital will be presented through regression coefficients and rankings. Unlike the previous section, each subsection will focus on regression coefficients and ranking.

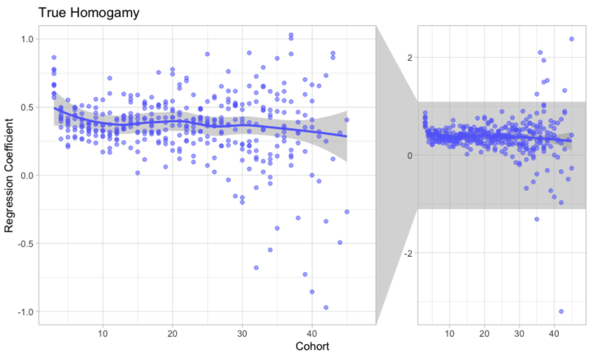
Although this chapter investigates mobility in cultural capital, the term educational mobility will still be used since this thesis utilizes cultural capital to express a broader and more complex aspect of education. Moreover, it should be recalled that the maximum value between parents' cultural capital is used for both measures, as explained in Section 2.5.

Regression Coefficients

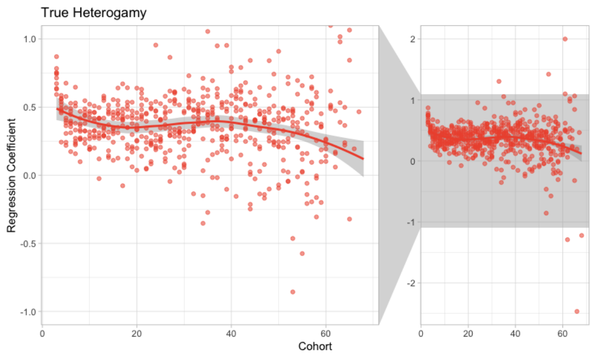
Regression coefficients expressing the effect size of the maximum cultural capital of parents on the agent are calculated for each cohort of the replications. These coefficients provide insights into the relationship between parental cultural capital and educational mobility. The regression coefficients are then plotted across the cohorts, and a loess line is added to summarize the trend observed across the 10 replications of each scenario. This trend in the regression coefficients serves as an indicator of whether relative mobility in cultural capital is improving or not. A high regression coefficient denotes a lower relative mobility since it means that the parents' cultural capital is highly correlated with the child's cultural capital. Accordingly, an upward trend illustrates increasingly lower mobility across the cohorts, while a downward trend stands for increasingly higher mobility.

When looking at Figure 3.7, it is notable that the loess line remains relatively stable in the true homogamy scenario, indicating a consistent level of mobility. Conversely, in the true heterogamy scenario, the regression coefficient exhibits a deeper decline in the earlier cohorts and a gradual decline over the cohorts. This downward trend suggests increasing mobility, where the cultural capital of parents becomes less influential over time, although the relationship may not be strictly linear. It is important to note that a direct comparison between these two scenarios is challenging, as the population in the true homogamy scenario diminishes more rapidly than in the true heterogamy scenario. The longer lifespan of the population in the heterogamy scenario allows for larger sample sizes in the younger cohorts across the replications, which is reflected in the smaller 95% error band in the true heterogamy scenario.

Focusing on the individual standard errors for the regression coefficients of each replication in Figure 3.8, a notable observation is the relatively larger standard errors exhibited by the true heterogamy scenario's replications in the earlier cohorts compared to the true homogamy scenario. This indicates that within each cohort, there is greater variation between observations in the true heterogamy scenario, particularly in the earlier cohorts. In contrast, the true homogamy scenario shows larger standard errors only in later stages when the population size has significantly decreased. In other words, homogamy ensures a certain level of correlation between generations, maintaining a more stable mobility pattern. On the other hand, heterogamy introduces flexibility that allows for modification of the mobility trends, ranging from subtle to more pronounced variations.

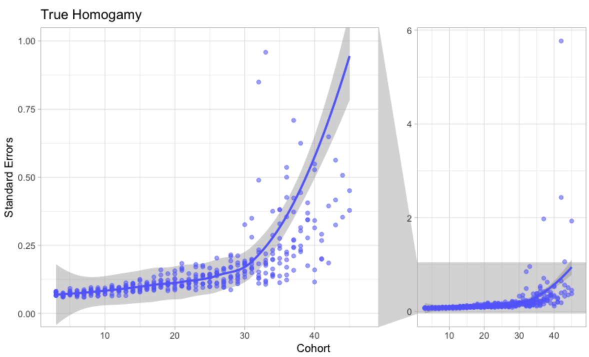


(a) True Homogamy

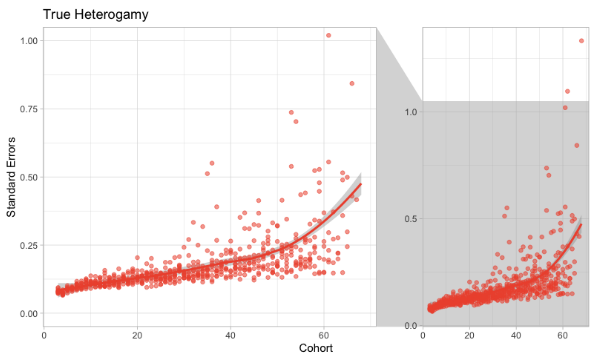


(b) True Heterogamy

Figure 3.7: The effect of the maximum parent's cultural capital on the child's cultural capital (i.e., regression coefficient) at each cohort is depicted as a dot for 10 replications of the two extreme scenarios. The effects are obtained using linear regression. A bold line is a loess line with 95% confidence interval. The true homogamy scenario shows a more stable loess line than the true heterogamy scenario.



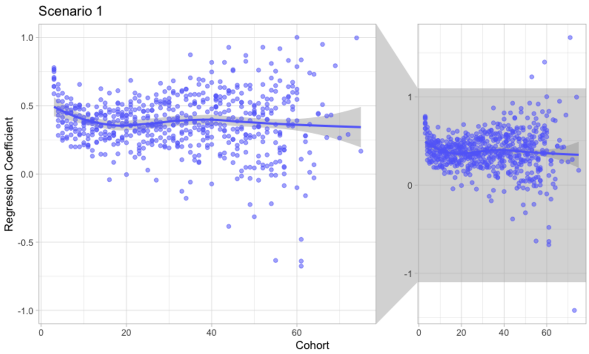
(a) True Homogamy



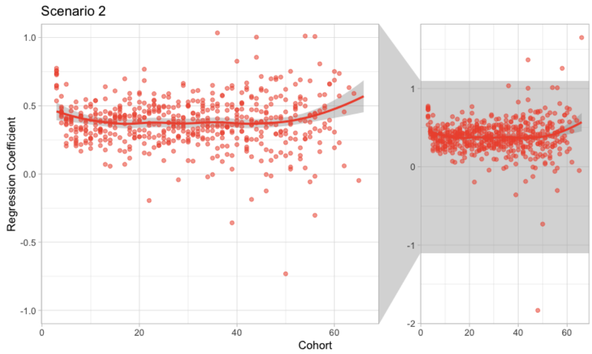
(b) True Heterogamy

Figure 3.8: The standard error (SE) of the regression coefficient explaining the child's cultural capital with the maximum parent's cultural capital at each cohort is depicted for 10 replications of the two extreme scenarios. The true heterogamy scenario generally has larger SEs except for Cohort 35, where the true homogamy scenario's SEs increase dramatically due to the small population size.

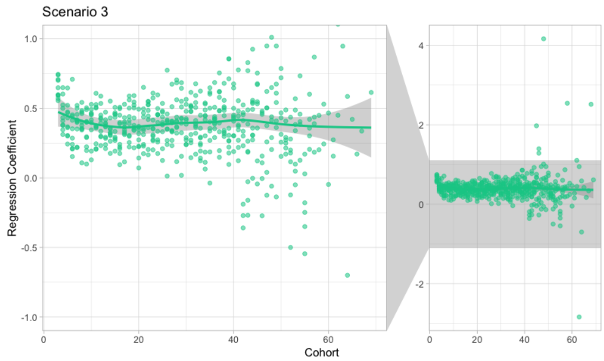
The patterns observed in the regression coefficients across the three scenarios, which aimed to simulate more realistic settings, align with the findings from the two extreme cases. Scenario 1, depicted in Figure 3.9a, exhibits a combination of the patterns observed in the true homogamy and heterogamy scenarios. Initially, the mean regression coefficients show a slight decrease around Cohort 20, followed by a gradual reduction with a peak around Cohort 40. While the true heterogamy scenario typically demonstrated a significant downward trend, in Scenario 1, the line appears to hover around 0.3, albeit slowly decreasing.



(a) Scenario 1: The base setting



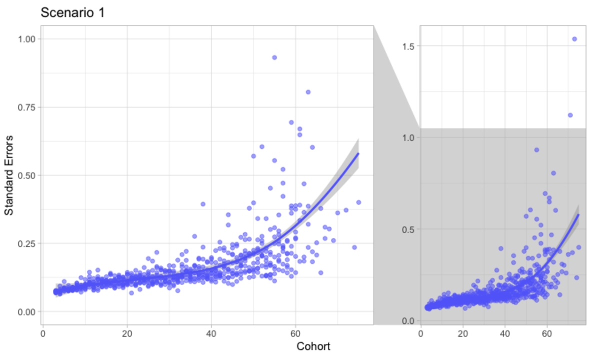
(b) Scenario 2: Incentives to homogamy



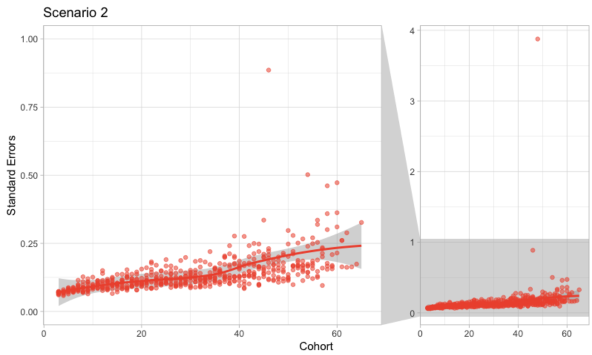
(c) Scenario 3: Incentives to heterogamy

Figure 3.9: The effect of the maximum parent's cultural capital on the child's cultural capital (i.e., regression coefficient) at each cohort is depicted as a dot for 10 replications of the three scenarios. The effects are obtained using linear regression. A bold line is a loess line with 95% confidence interval. Scenarios 1 and 3 show a deeper U-shape trend in the earlier cohorts, while Scenario 2 maintains a relatively stable trend across cohorts.

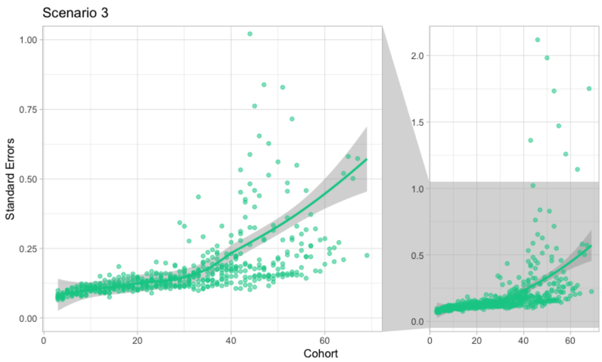
In Figure 3.9b, Scenario 2 explores the encouragement of homogamous marriages and displays an unexpected pattern: an upward trend toward the end of the cohorts, which contain the younger population. Similar to the true homogamy case, the regression coefficients exhibit consistency across the cohorts initially, but as of Cohort 40, the loess line gradually rises. This upward trend suggests that the maximum cultural capital of parents is becoming increasingly important in determining the child's cultural capital across the cohorts. Again, it is worth noting that the observed trend could be influenced by the decreasing number of agents as the



(a) Scenario 1: The base setting



(b) Scenario 2: Incentives to homogamy



(c) Scenario 3: Incentives to heterogamy

Figure 3.10: The standard error (SE) of the regression coefficient explaining the child's cultural capital with the maximum parent's cultural capital at each cohort is depicted for 10 replications of the three scenarios. Scenario 3 generally has larger SEs and increases over cohorts more rapidly than other scenarios.

cohort number increases. Scenario 3, illustrated in Figure 3.9c, follows a pattern similar to Scenario 1, but with more frequent occurrence of outlying cases. Already at Cohort 43, one of the replications has a regression coefficient as high as 4.167, whereas none of the replications in Scenario 1 exceeds a coefficient of 2.

Regarding the standard errors of the coefficients shown in Figure 3.10, Scenario 3 generally exhibits a larger standard error across the cohorts, having the largest standard error at Cohort 40. Comparing Scenarios 2 and 3, standard errors in Scenario 2 do not increase across cohorts as much as in Scenario 3. Similar to the extreme cases, the higher prevalence of homogamy seems to ensure the correlation between the parents' and child's cultural capital more than heterogamy thereof.

Ranking

The ranking of agents' cultural capital whose maximum parent's cultural capital is strictly lower than the median is obtained to inquire into mobility in cultural capital. Comparing the two extreme cases, as shown in Figure 3.11, both cases exhibit an upward trend as the cohort increases. However, the true homogamy scenario demonstrates a steep increase in the percentile of mean rankings for the children in the lower half, ranging from around 50% to around 75%. Considering that 100% implies the largest cultural capital one can possess, a percentile of 75% corresponds to being among the top 25% in terms of cultural capital possession within the given population. On the other hand, the true heterogamy scenario shows a slower and smaller increase, with the percentile rising from around 50% to around 64%.

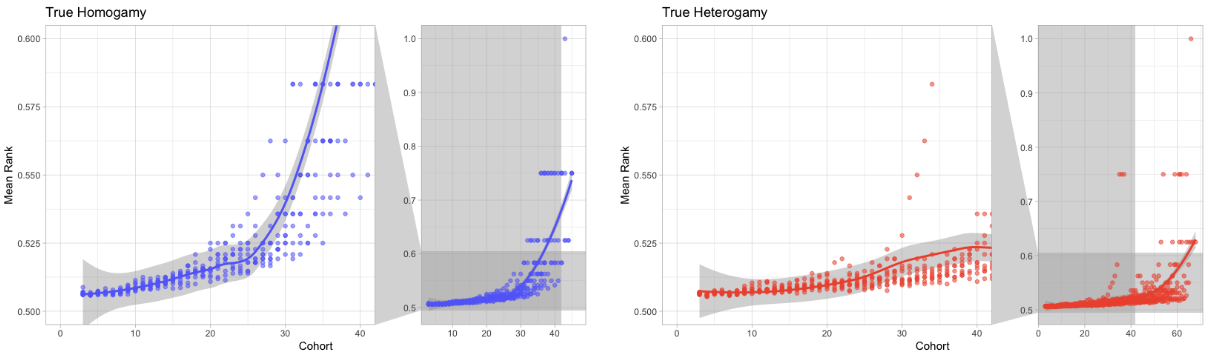
It is worth noting that although the percentiles are calculated while considering the population size of each cohort, the smaller population size in larger cohorts may contribute to increased variance between replications. Nevertheless, some discrepancies between replications are already evident in the earlier cohorts (around Cohort 20 and 30) in the true heterogamy scenario, where the population size did not decrease dramatically.

Similar to the extreme cases, all three scenarios exhibit an upward trend in percentiles (see Figure 3.12). However, unlike the extreme cases, Scenario 2, where homogamy is encouraged, shows the lowest peak point at around 60% among the three scenarios. Scenario 1 reaches approximately 65% and Scenario 3 hovers around 63%. Notably, Scenario 3 demonstrates more outlying replications than the other scenarios, potentially contributing to the upward trend observed in the loess line. Moreover, unlike the extreme cases, Scenario 3, which encourages heterogamous marriages, demonstrates a steeper increase compared to Scenario 2.

3.3 Sensitivity Analyses

While the analysis presented above has yielded interesting findings, it is important to assess the robustness of these findings by conducting sensitivity analyses. Several factors may significantly influence the results, and understanding their impact is crucial for drawing reliable conclusions. In this section, two key sensitivity analyses are performed.

Firstly, the influence of the unfairness variable is examined by investigating how changes in this variable affect the results. By altering the unfairness variable, one can observe the variations in outcomes and assess the sensitivity of the model to this parameter. This analysis provides insights into the relationship between the unfairness variable on the population's mean



(a) True Homogamy (b) True Heterogamy

Figure 3.11: Percentiles of the mean ranking of the bottom half's children are depicted across cohorts for 10 replications of two extreme scenarios. The graphs are zoomed to Cohorts between 3 and 40 and mean percentiles between 0.5 and 0.6. The loess lines in bold summarize the percentiles at each cohort. Both show increasing patterns, indicating that the position of the next generation in society is increasingly improving.

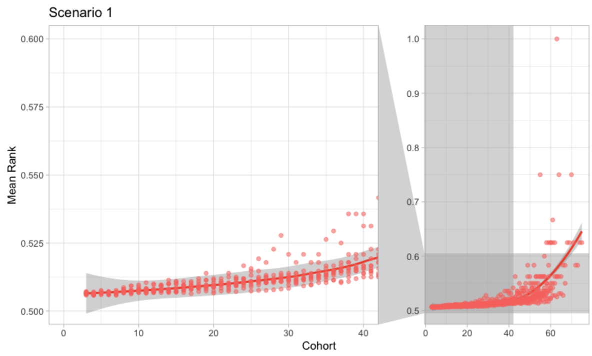
educational level and the impact of the unfairness on educational mobility.

Secondly, the impact of the maximum number of children allowed in the system is explored. By relaxing the constraint on the maximum number of children by one degree, the scenarios can accommodate a larger population and extend the lifespan of the agents. This allows us to uncover hidden trends in mobility that may not have been apparent in the previous analysis due to the limited population size. This sensitivity analysis provides a broader perspective on educational mobility and offers valuable insights into the dynamics of the system.

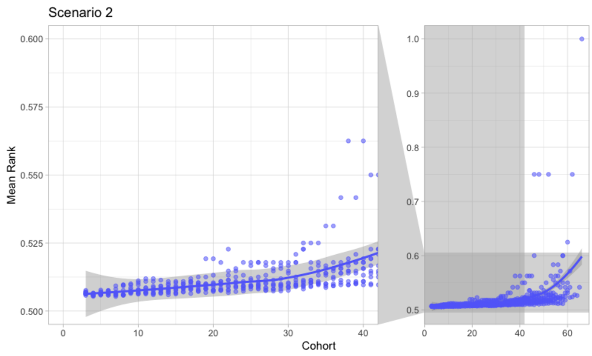
Changing the Unfairness Factor

In the analysis, an unfairness factor of 0.2 is utilized to represent a system where education is relatively accessible. By running the base scenario at least 10 times with different unfairness factors (0, 0.3, 0.4, 0.5, 0.6, and 1.0) in the built application, insights can be gained into the influence of this factor on the average education across time steps and its trends. As the unfairness factor increases, the probability of the average education exhibiting a downward trend also increases. Notably, already at an unfairness factor of 0.5, downward trends are more prevalent in the replications compared to an unfairness factor of 0.2. Additionally, the factor of 0.2 showed less variation between replications compared to the factor of 0.3, which demonstrated either an upward pattern or a slow U-shape, a pattern that decreases in the first half and then slowly increases later (See Appendix on Github for the logbook).

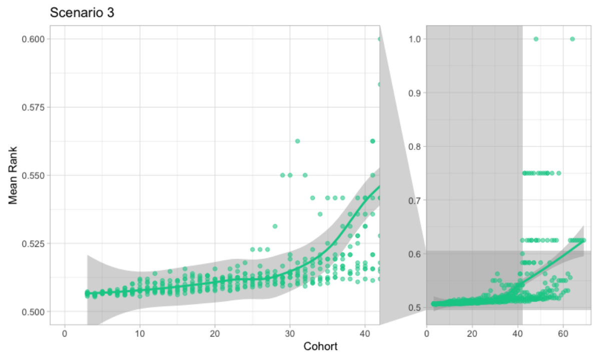
The influence of the unfairness factor on the average educational level is prevalent. When comparing an unfairness factor of 0.5 to the previously analyzed 0.2, it is observed that



(a) Scenario 1: The base setting



(b) Scenario 2: Incentives to homogamy

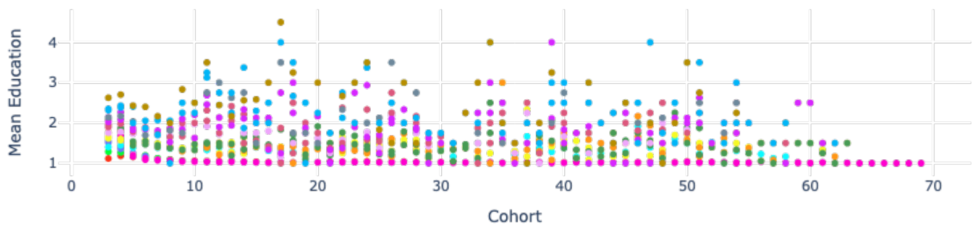


(c) Scenario 3: Incentives to heterogamy

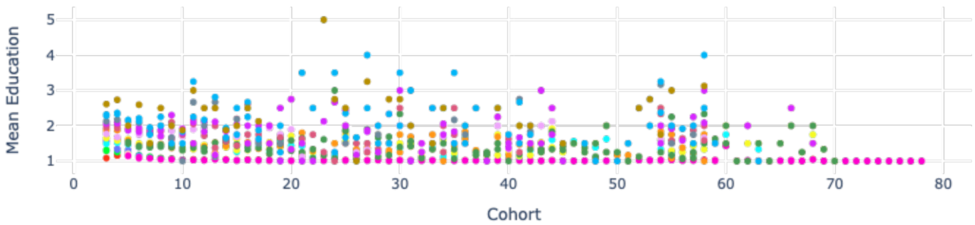
Figure 3.12: Percentiles of the mean ranking of the bottom half's children are depicted across cohorts for 10 replications of the three scenarios. The graphs are zoomed to Cohorts between 3 and 40 and mean percentiles between 0.5 and 0.6. The loess lines in bold summarize the percentiles at each cohort. All of them show increasing patterns, indicating that the position of the next generation in society is increasingly improving but Scenario 3 shows more variability between replications than others.

the increased probability of lowering the average educational level obscures the effects of homogamy and heterogamy when examining general trends over cohorts and time steps, as well as absolute mobility measures. All the scenarios have downward trends in the mean educational levels, reaching around 1 already at Cohort 10 and around time step 70. When it comes to educational mobility, all the scenarios show nearly 0 for both absolute measures. Still, as in the previous analysis, Scenario 3 still predominantly lies on the top in terms of mean educational levels both across cohorts and time steps, and it starts with the lowest probability for upward mobility.

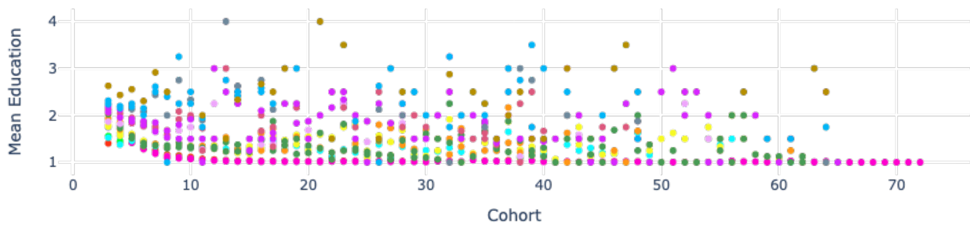
The obscured distinctiveness among the scenarios with the unfairness factor of 0.5 is evident. When assessing the differences between the scenarios, the number of cohorts where the three scenarios significantly differ is considerably lower than when the unfairness factor is 0.2. Especially in the earlier cohorts before Cohort 20, significant differences rarely are captured



(a) Scenario 1: The base setting



(b) Scenario 2: Incentives to homogamy



(c) Scenario 3: Incentives to heterogamy

Figure 3.13: The mean educational levels of groups characterized by parents' educational levels averaged over 10 replications of the three scenarios are depicted for a sensitivity analysis. The unfairness factor is set to 0.5 rather than 0.2. The groups are colored as in the legend above. E.g., [1,2] means one of the parents has an educational level of 1, and the other has 2. The scenarios are hard to distinguish from one another, except that before Cohort 10, Scenario 3 shows a smaller variance than others.

among the three scenarios nor between Scenario 2 and 3 at a significance level of 0.05. For instance, for two-scenario comparisons, only Cohort 8 appeared significant with Measure 1, and Cohort 8 and 10 with Measure 2, while three-scenario comparisons exhibit Cohort 8 and 20 with Measure 1 and Cohort 8, 15 and 20 with Measure 2. Furthermore, as shown in Table 3.2, introducing scenarios as a structuring variable did also not improve the model with Measure 1, while with Measure 2, the linear mixed model had a singularity issue. This confirms that increasing the unfairness factor obscures the effect of mating patterns on educational mobility.

When assessing mean educational levels per group defined by the parent's educational level, Figure 3.13 shows that the three scenarios are less distinguishing, unlike with the unfairness factor of 0.2. Albeit very small, Scenario 3 exhibits more variation between groups across cohorts compared to Scenario 2. This contradicts the findings observed in the previous analysis, where Scenarios 3 and 5, preferring heterogamy, showed smaller variances. However, consistent with the previous analysis, all scenarios generally place groups with parents holding the lowest educational level at the bottom. Additionally, greater diversity is observed among the higher educational levels. Above an educational level of 2.5, Scenario 2 includes groups [2, 5], [3, 5], [4, 4], [4, 5], and [5, 5], while Scenario 3 additionally includes groups [3, 3] and [3, 4].

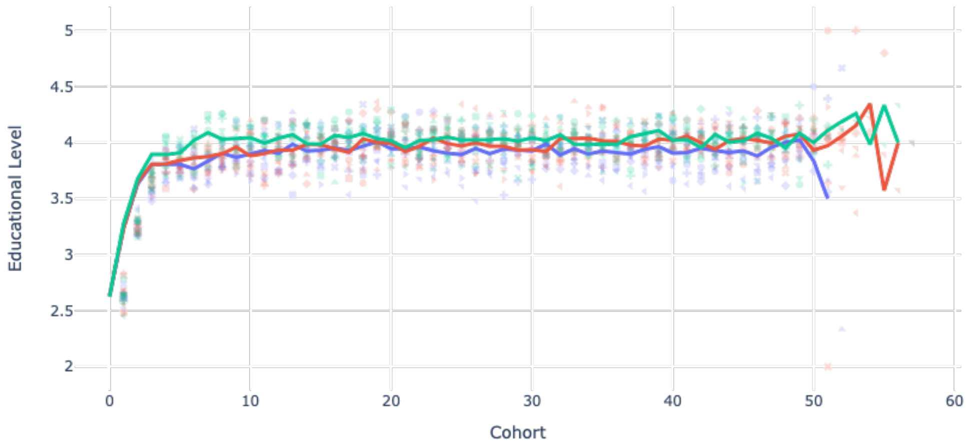
In conclusion, the unfairness factor has a notable effect on educational mobility as a higher unfairness factor increases the chance for downward trends and obscures the effect of mating patterns on educational mobility. Therefore, identifying a factor that accurately represents the system of interest is crucial, considering that a higher unfairness factor negatively affects the population's average educational levels and mobility. However, determining the precise factor for a specific context can be challenging when the relevant data are unavailable. In the thesis, a choice of 0.2 seems to generate a good mix of scenarios with both downward and upward trends, leaning towards more upward trends in population-wise educational level. The general upward trends also seemingly resonate with the increased tertiary education population on average in OECD countries (See OECD, 2022b and more specifically to Belgium, OECD, 2022c).

Table 3.2: Log-likelihood ratio tests for linear mixed models in sensitivity analyses

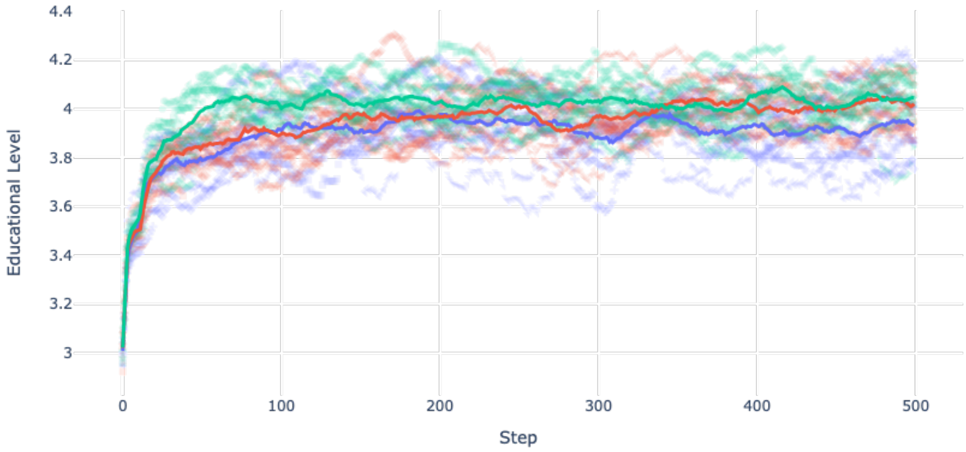
	Nr.Parameters	logLik	AIC	LRT	Df	$Pr(> \chi^2)$
Unfairness of 0.5						
Measure 1						
Null	4	638.16	-1308.3			
(1 | Scenario)	3	658.12	-1310.3	0.034538	1	0.8526
Measure 2 (Singularity issue)						
Null	4	655.5	-1303			
(1 | Scenario)	3	655.5	-1305	-2.2737e-13	1	1
3 Maximum number of children						
Measure 1						
Null	4	253.79	-499.58			
(1 | Scenario)	3	244.14	-482.28	19.295	1	1.12e-05 ***
Measure 2 (Singularity issue)						
Null	4	204.16	-400.32			
(1 | Scenario)	3	204.16	-402.32	-1.7053e-13	1	1

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Changing the Maximum Number of Children



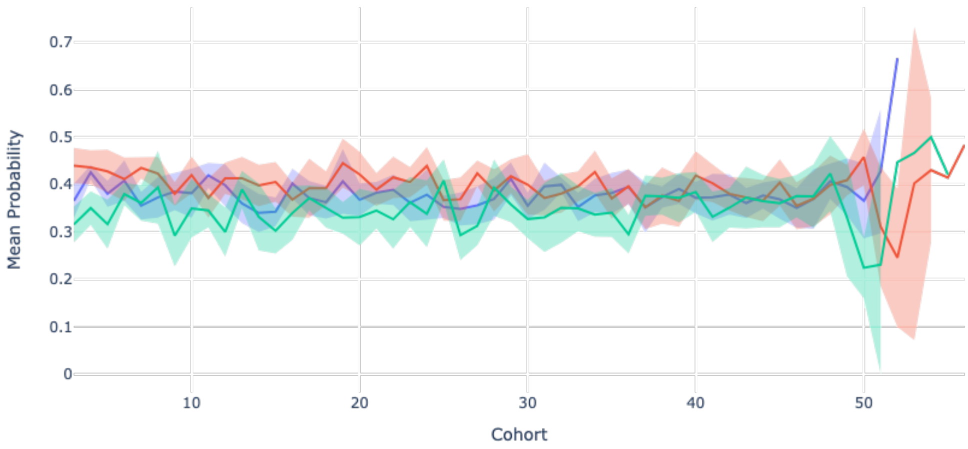
(a) Mean educational levels over cohorts



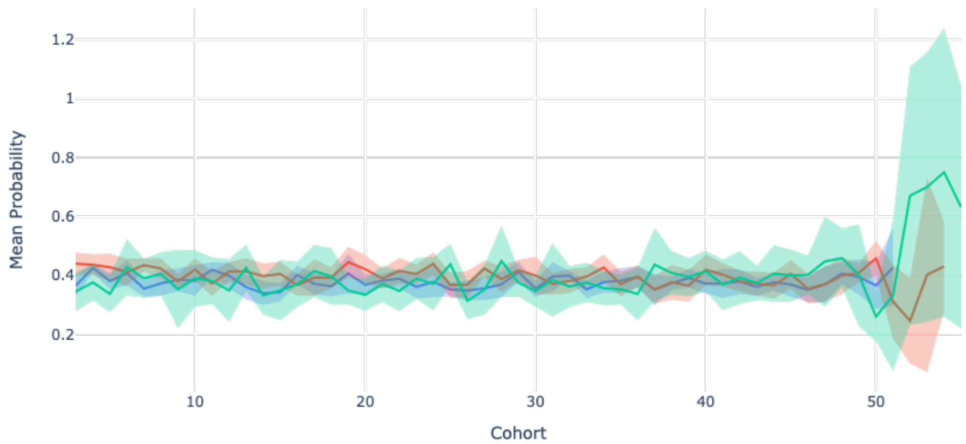
(b) Mean educational levels over time-steps

Figure 3.14: Population-wise mean educational levels are depicted for a sensitivity analysis. Scenario 1 in blue, Scenario 2 in red and Scenario 3 in green have 3 maximum number of children. The dots illustrate the mean educational level of each of the 10 replications, while the bold lines connect the scenario-wise means at each point. These scenario-wise means are calculated as the averages of the 10 dotted values. The trend lines appear similar to the previous analysis with Scenario 3 generally having the highest mean educational levels.

In this sensitivity analysis, the maximum number of children a couple can have is extended from 2 to 3. The restriction on the maximum number of children is still required to facilitate data management and also to provide realistic scenarios, where unlimited reproduction is rare. Therefore, by extending the maximum number, this section aims to provide insights into the impact of this change on the conclusions. It is important to note that the unfairness factor of 0.2 is used for all three scenarios (Scenario 1-3) in this analysis.



(a) Measure 1: The probability of less privileged children obtaining a higher educational level than their parents.



(b) Measure 2: It adds the probability of children with upper tertiary-educated parents maintaining the status to Measure 1.

Figure 3.15: The absolute mobility in education over cohorts averaged over 10 replications of each scenario for a sensitivity analysis. Scenario 1 in blue, Scenario 2 in red and Scenario 3 in green have 3 maximum number of children. 95% error bands express the variability across the replications of each scenario. Aligned with the previous analysis, Scenario 3 starts with the lowest mobility and Scenario 2 with the highest, albeit similar. With Measure 2, Scenario 3 shows a wider error band than other scenarios.

Consistent with the previous analysis, Scenario 3 maintains the highest average educational level across the 10 replications until approximately Cohort 30 and time step 350 (see Figures 3.14a and 3.14b). Regarding the absolute measurements presented in Figures 3.15a and 3.5b, similar patterns to the previous analysis emerge. Specifically, Scenario 3 consistently exhibits the lowest mean probability for child agents to experience upward educational mobility, especially compared to Scenario 2. Additionally, when examining mobility specifically with Measure 2 (Figure 3.15b), including child agents with upper tertiary educated parents, Scenario

3 demonstrates the largest variance among the three scenarios.

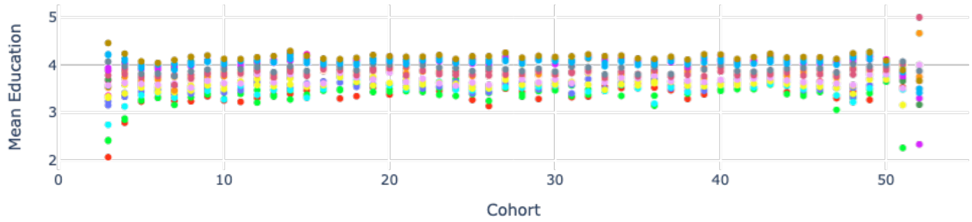
Investigating the mean educational levels per group defined by the parents' educational levels yields similar results as before, as shown in Figures 3.16. The variance in Scenario 2 is larger compared to Scenario 3. This highlights that the higher mean educational level in Scenario 2 is predominantly due to the highly positioned groups' educational levels, while the groups share similar mean educational levels in Scenario 3.

With the maximum number of children set at 3, more cohorts exhibit significant differences in educational mobility with both measures between the scenarios compared to when the restriction was set at 2. Specifically, when comparing Scenarios 2 and 3, as well as all three scenarios with Measure 1, more significant differences are evident, especially in the earlier cohorts before Cohort 20, than in the previous analysis.

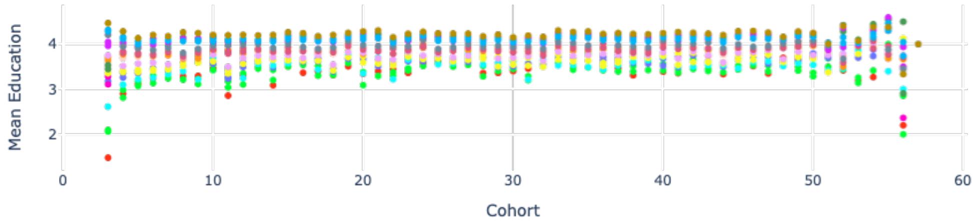
Regarding Measure 2, although more cohorts show significant differences between Scenarios 2 and 3 than in the previous analysis, these differences appear generally less frequently than with Measure 1. The lack of distinctiveness between the scenarios in terms of Measure 2 is reflected in the singularity issue encountered when constructing a linear mixed model with the Measure 2 data set, as shown in Table 3.2. Despite this, with Measure 1, it becomes evident that understanding educational mobility with scenarios is more effective than without it, with a p-value in the log-likelihood ratio test smaller than 0.001. This highlights the continuous impact of mating patterns on educational mobility even with the extended maximum number of children.

When it comes to mobility in cultural capital, some differences emerge compared to the previous analysis with a maximum of 2 children. As in the previous analysis, the regression coefficients over the cohorts remain above 0 on average. However, the patterns in the last cohorts show slight variations. While Scenarios 1 and 3 exhibit relatively stable trends, Scenario 2 demonstrates a slight downward trend in the last cohort. Furthermore, extending the maximum number of children resulted in reduced variation between the replications compared to the previous analysis. When excluding the last cohorts where the error band is wide, while all the scenarios stay rather stable at the end of the cohorts, Scenario 3 shows a steeper decline in the first 20 cohorts compared to Scenario 2. This aligns with the previous analysis, where Scenario 3 demonstrated a slightly greater decline in the first 20 cohorts compared to Scenario 2.

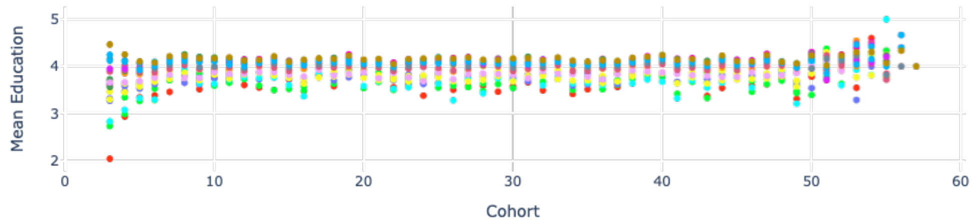
Regarding the rankings in Figure 3.18, the scenarios do not appear much different from one another, except for the last cohorts, where each replication contains fewer agents compared to the initial population. Notably, Scenario 2 exhibits more outlying values and a larger variance in the later cohorts compared to Scenario 3. This finding contrasts with the previous analysis,



(a) Scenario 1: The base setting



(b) Scenario 2: Incentives to homogamy

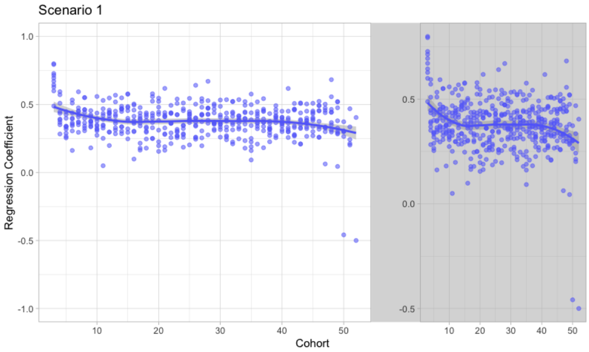


(c) Scenario 3: Incentives to heterogamy

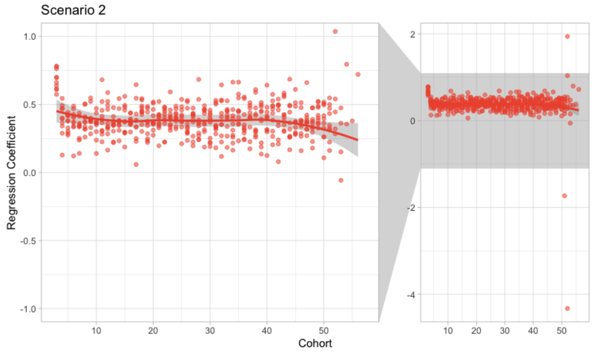
Figure 3.16: The mean educational levels of groups characterized by parents' educational levels averaged over 10 replications of the three scenarios are depicted for a sensitivity analysis. The maximum number of children is set to 3 rather than 2. The groups are colored as in the legend above. E.g., [1,2] means one of the parents has an educational level of 1, and the other has 2. Similar to the previous analysis, Scenario 3 exhibits a smaller variance across cohorts than others.

where Scenario 3 demonstrated a larger variance and more outlying values than Scenario 2. This difference may be attributed to the demographic impact on the mobility measure, as the increase in the maximum number of children and the less restrictive nature of heterogamy in Scenario 3 allows for a larger population size. Consequently, Scenario 3 yields more consistent results with a narrower 95% confidence band compared to Scenario 2.

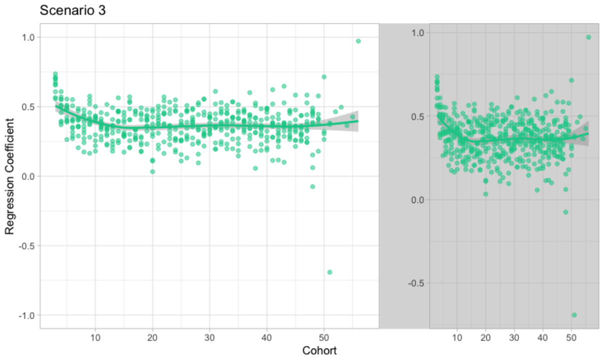
In conclusion, the change in the maximum number of children does not seem to change the general conclusions from the analysis. However, the increase in the maximum number of children increases the population size in each cohort, which affects the size of 95% confidence intervals. It should be noticed that a couple's educational levels also have an impact on deciding how many children they want, together with other factors such as reasons regarding



(a) Scenario 1: The base setting



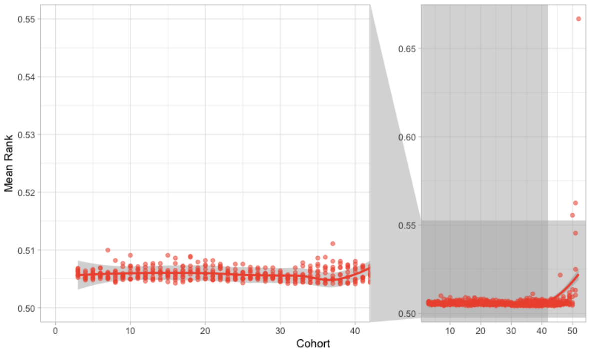
(b) Scenario 2: Incentives to homogamy



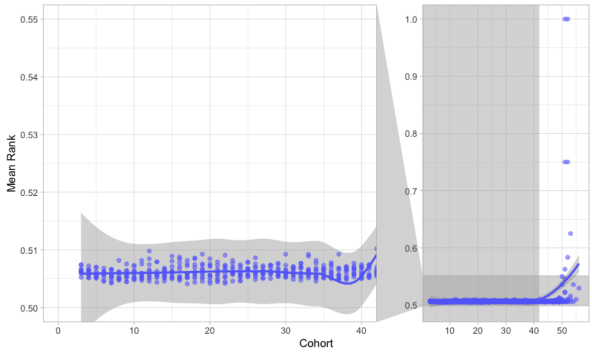
(c) Scenario 3: Incentives to heterogamy

Figure 3.17: The effect of the maximum parent's cultural capital on the child's cultural capital (i.e., regression coefficient) at each cohort is depicted as a dot for 10 replications of the three scenarios for a sensitivity analysis. The maximum number of children is 3. The effects are obtained using linear regression. A bold line is a loess line with 95% confidence interval. Similar to the previous analysis, Scenario 3 shows a deeper decline in the effects in the earlier cohorts than others.

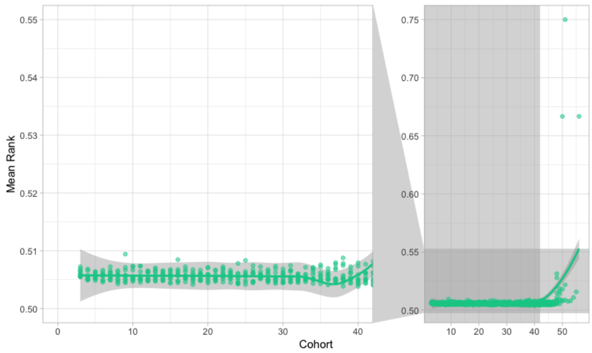
the country's economy and/or global environmental issues. Therefore, when more children are allowed to be born, more complexity may be needed in the reproduction phase of the model.



(a) Scenario 1: The base setting



(b) Scenario 2: Incentives to homogamy



(c) Scenario 3: Incentives to heterogamy

Figure 3.18: Percentiles of the mean ranking of the bottom half's children are depicted across cohorts for 10 replications of the three scenarios. The graphs are zoomed to Cohorts between 3 and 40 and mean percentiles between 0.5 and 0.55. The loess lines in bold summarize the percentiles at each cohort. All the scenarios seem to keep the top 49% position throughout the cohorts.

Chapter 4

Conclusions

With the aim of exploring the relationship between educational assortative mating and educational mobility, agent-based models were constructed and analyzed. This chapter concludes the paper by first offering a summary of the findings with regard to the research questions, the limitations of the thesis, and some suggestions for future research.

4.1 Summary of Findings

Five agent-based models were built to answer the following research questions:

1. Does educational mobility increase more in the long term when people with different educational backgrounds marry each other compared to when marriages only happen within the same educational group?

2. Does a higher population-wise educational level (i.e., a higher cross-sectional average educational level of the entire population) reflect a higher degree of intergenerational educational mobility?

3. Does every educational group experience the same improvement or decline in average educational level throughout generations?

Two scenarios explored extreme homogamy and heterogamy world, where all the formed marriages are homogamous or heterogamous. Three scenarios represented more realistic cases, where a varying degree of rewards are given to mating types (i.e., no rewards to any type, only rewarding homogamy or heterogamy). The thesis aimed to build models with simple rules regarding mating patterns and transmission of educational levels, which allowed a controlled environment to comprehensively explore the impact of mating patterns on educational mobility.

Generally, in the scenarios where heterogamy is prominent or the only marital type allowed in the system (hereafter, heterogamous cases), lower absolute mobility in education was captured compared to the scenarios where homogamy is encouraged or the only martial type allowed in the system (hereafter, homogamous cases). However, the population-wise educational levels throughout cohorts were higher in the heterogamous cases.

While having smaller population-wise educational levels than the heterogamous cases, the homogamous cases demonstrated generally higher and more stable patterns in educational mobility throughout cohorts. This finding aligns with the finding that the population-wise improvement in educational levels was shared with more diverse parental backgrounds in heterogamous cases than in homogamous cases.

Despite the differences between the mating patterns, the impact of mating patterns on educational mobility typically disappeared as time step went further. The difference between the scenarios was the most significant in the first cohort with children, Cohort 3, and its significance became increasingly obscured over time.

When investigating mobility in terms of cultural capital, the difference between the mating patterns was less prevalent. All the scenarios with different mating patterns demonstrated a positive effect of parents' cultural capital on the child's cultural capital (i.e., regression coefficient). A slight difference between the homogamous and heterogamous cases is that the effect size stayed rather constant across generations in the homogamous cases, whereas the heterogamous cases showed a greater decline in the earlier cohorts, indicating improvement in educational mobility. When ranking the cultural capital of the bottom half's children, all the scenarios seemed to show upward trends, signaling upward mobility.

The following summarizes the answers to the research questions:

1. In terms of absolute mobility in educational attainments, higher mobility was captured when people of the same educational group marry than when differently educated people marry. However, in terms of relative mobility in cultural capital, higher mobility was observed for the case where differently educated people marry in the earlier cohorts for a short term.

2. The increase in population-wise educational level did not necessarily reflect higher intergenerational educational mobility. Heterogamous cases, which showed higher population-wise educational levels, had lower intergenerational educational mobility than homogamous cases.

3. Every educational group experiences improvement in population-wise educational levels differently, especially in homogamous cases. Most contribution to the increase in

population-wise educational levels is made by the privileged group whose parents are highly educated. In contrast, in heterogamous cases, more various groups made the contribution.

4.2 Limitations and Suggestions

While this study provides valuable insights into the relationship between educational mobility and educational assortative mating, it is crucial to acknowledge its limitations and explore potential avenues for improvement.

The main limitation of the experiment conducted for this thesis is that the models are simple. While simplicity offers a controlled environment to explore the subject, it does not provide a fully realistic view. Therefore, adding more complexities to the models will be beneficial for future inquiries.

Firstly, with the focus on mating patterns and their influence on educational mobility, the models were limited to simple assumptions about reproduction behaviors. The current models only consider heterosexual couples allowing them to have a maximum of two children, and do not account for divorce or other complexities in family dynamics. Similarly, this thesis considered hypergamy and hypogamy with one concept, heterogamy, to focus on the main research questions and make straightforward models. Therefore, future models can incorporate different marital and family types. Especially considering divorce in the model can be interesting as divorce can affect the child's education both in a positive and negative way.

Secondly, the current models fail to capture the potential correlation between procreation and the educational levels of couples. For instance, in certain countries and periods, highly educated couples may not necessarily have a higher birth rate compared to less educated couples. Therefore, future models can consider the educational level-specific fertility rates to achieve a more realistic model.

Lastly, the thesis had difficulty finding detailed data to base the model on. Therefore, collecting detailed data will enrich system-wise and behavioral rules, which will allow a more realistic model. These data can relate to the system-wise rules or agents' behavioral rules. Specifically, for the system-wise rules, one can delve into the data regarding how accessible the educational system is to each stratum of society to better represent the unfairness factor. For agents' behavioral rules, longitudinal data of educational levels that also indicate detailed information, such as the field of study and the prestige of the school, is desired to enrich the transmission rules for the cultural capital and educational levels. By collecting more data, the study can attain a heightened level of precision, ensuring that the findings and conclusions derived are more relevant and applicable to real-world scenarios.

To summarize, the agent-based modeling employed in this study provided insights into the long-term impact of educationally assortative mating on intergenerational educational mobility within a controlled environment. On average, greater educational diversity within a married couple appeared to elevate the overall educational level of the population. However, it did not seem to enhance intergenerational educational mobility as fast as when educational diversity within a couple is lower. When considering the broader concept of cultural assets, specifically cultural capital, the influence of mating patterns on relative mobility seemed less pronounced. Nevertheless, heterogamy demonstrated a more substantial decline in the parental effect on children's cultural capital in the short term.

Based on these findings, one may contemplate that diversity facilitates collective progress across all strata of society, albeit with potentially slower growth in absolute mobility compared to a society that promotes uniformity. Conversely, a uniform society may ensure stable and high absolute mobility but at the expense of increased inequality between different strata. While this interpretation is difficult to achieve solely with the experiment conducted in this thesis, it presents an intriguing perspective for future research to explore further.

In conclusion, the study's findings illuminate the complex interplay between educational assortative mating and intergenerational educational mobility. They underscore that population-wise growth does not necessarily reflect upward mobility or equal growth of all the social strata. To achieve a deeper understanding of this disparity, future research could adopt an ergodicity economics perspective with more complex models, which explores the relationship between ensemble averages and time averages (Cf., Peters, 2019; Stojkoski et al., 2022; Farmer and Bouchaud, 2020; Berman et al., 2020). The study's findings also raise important questions about the balance between diversity and uniformity in society when addressing social inequality. Here, introducing an ergodicity perspective on cooperation (Peters and Adamou, 2022) can add another layer of analysis to this discussion, prompting further investigations into the impact of diverse social structures on inequality. Likewise, the findings of this thesis present compelling opportunities for future research endeavors, together with some insights into the intricate dynamics between educational assortative mating and intergenerational educational mobility.

4.3 Data Availability

The main and additional codes for the analysis, as well as the used data, can be found on Github: `https://github.com/SereneKim/MasterThesis_Seorin_Kim.git`

Bibliography

Abraham, K. G. and Mallatt, J. (2022). Measuring human capital. *Journal of Economic Perspectives*, 36(3):103–30.

Arenas, A. and Hindriks, J. (2021). Intergenerational mobility and unequal school opportunity. *The Economic Journal*, 131(635):1027–1050.

Beller, E. and Hout, M. (2006). Intergenerational social mobility: The united states in comparative perspective. *The future of children*, pages 19–36.

Berman, Y., Peters, O., and Adamou, A. (2020). Wealth inequality and the ergodic hypothesis: Evidence from the united states. *Forthcoming in Journal of Income Distribution*.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences*, 99(suppl_3):7280–7287.

Bourdieu, P. (1973). Cultural reproduction and social reproduction. *London: Tavistock*, 178:71–112.

Bourdieu, P. (1986). The forms of capital. In Richardson, J., editor, *Handbook of Theory and Research for the Sociology of Education*, pages 241–258. Greenwood, Westport, CT.

Bourdieu, P. (2002). Pierre bourdieu on marriage strategies. *Population and Development Review*, 28(3):549–558.

Breen, R. (2004). The Comparative Study of Social Mobility. In Breen, R., editor, *Social Mobility in Europe*, chapter 1. Oxford University Press.

Breen, R. (2019). Education and intergenerational social mobility in the us and four european countries. *Oxford Review of Economic Policy*, 35(3):445–466.

Card, D., Domnisoru, C., and Taylor, L. (2018). The intergenerational transmission of human capital: Evidence from the golden age of upward mobility. Technical report, National Bureau of Economic Research.

Conte, R. and Paolucci, M. (2014). On agent-based modeling and computational social science. *Frontiers in psychology*, 5:668.

Coontz, S. (2004). The world historical transformation of marriage. *Journal of Marriage and the Family*, pages 974–979.

Dribe, M. and Lundh, C. (2010). Marriage choices and social reproduction: The interrelationship between partner selection and intergenerational socioeconomic mobility in 19th-century sweden. *Demographic Research*, 22:347–382.

Erát, D. (2021). Educational assortative mating and the decline of hypergamy in 27 european countries. *Demographic Research*, 44:157–188.

Erikson, R. and Goldthorpe, J. H. (2002). Intergenerational inequality: A sociological perspective. *Journal of Economic Perspectives*, 16(3):31–44.

European Social Survey European Research Infrastructure (ESS ERIC) (2021). ESS9 - integrated file, edition 3.1. [Data set].

Farmer, R. and Bouchaud, J.-P. (2020). Self-fulfilling prophecies, quasi non-ergodicity & wealth inequality. Technical report, National Bureau of Economic Research.

Goldthorpe, J. H. (2014). The role of education in intergenerational social mobility: Problems from empirical research in sociology and some theoretical pointers from economics. *Rationality and society*, 26(3):265–289.

Grow, A. and Van Bavel, J. (2015). Assortative mating and the reversal of gender inequality in education in europe: An agent-based model. *PloS one*, 10(6):e0127806.

Grow, A. and Van Bavel, J. (2017). *Agent-based modelling in population studies*. Springer.

Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using networkx. In Varoquaux, G., Vaught, T., and Millman, J., editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA.

Jäger, G. (2019). Replacing rules by neural networks a framework for agent-based modelling. *Big Data and Cognitive Computing*, 3(4):51.

Jäger, G. (2021). Using neural networks for a universal framework for agent-based models. *Mathematical and Computer Modelling of Dynamical Systems*, 27(1):162–178.

Jerrim, J. and Macmillan, L. (2015). Income inequality, intergenerational mobility, and the great gatsby curve: Is education the key? *Social Forces*, 94(2):505–533.

Kazil, J., Masad, D., and Crooks, A. (2020). Utilizing python for agent-based modeling: The mesa framework. In Thomson, R., Bisgin, H., Dancy, C., Hyder, A., and Hussain, M., editors, *Social, Cultural, and Behavioral Modeling*, pages 308–317, Cham. Springer International Publishing.

Kelton, W., Sadowski, R., and Zupick, N. (2015). *Simulation with Arena*. McGraw Hill LLC, sixth edition.

Lee, K., Rucker, M., Scherer, W. T., Beling, P. A., Gerber, M. S., and Kang, H. (2017). Agent-based model construction using inverse reinforcement learning. In *2017 Winter Simulation Conference (WSC)*, pages 1264–1275. IEEE.

Lewis, F., Butler, A., and Gilbert, L. (2011). A unified approach to model selection using the likelihood ratio test. *Methods in ecology and evolution*, 2(2):155–162.

Macal, C. M. and North, M. J. (2009). Agent-based modeling and simulation. In *Proceedings of the 2009 winter simulation conference (WSC)*, pages 86–98. IEEE.

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., and Hines, J. E. (2018). Chapter 3 - fundamental principals of statistical inference. In MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., and Hines, J. E., editors, *Occupancy Estimation and Modeling (Second Edition)*, pages 71–111. Academic Press, Boston, second edition edition.

Mudd, A. L., Groeniger, J. O., Bal, M., Verra, S. E., van Lenthe, F. J., and Kamphuis, C. B. (2023). Testing conditionality with bourdieu's capital theory: How economic, social, and embodied cultural capital are associated with diet and physical activity in the netherlands. *SSM-Population Health*, 22:101401.

Mueller, C. W. and Parcel, T. L. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Child development*, pages 13–30.

Nomes, E. and Van Bavel, J. (2017). Education and marriage: The shift from female hypergamy to hypogamy in belgium, a 20th century cohort analysis. *Quetelet Journal*, 5(2):37–67.

OECD (2022a). *Education at a Glance 2022*. OECD Publishing. Retrieved from https://www.oecd-ilibrary.org/content/publication/3197152b-en.

OECD (2022b). Oecd education gps. Retrieved from https://gpseducation.oecd.org/. on 25-July-2023.

OECD (2022c). Oecd education gps: Belgium. Retrieved from https://gpseducation.

oecd.org/CountryProfile?primaryCountry=BEL&treshold=10&topic=EO. on 07-November-2022.

OECD (n.d.). Understanding social mobility. Retrieved from https://www.oecd.org/stories/social-mobility/. on 21-April-2023.

OECD.Stat (2019). Education and earnings. Retrieved from https://stats.oecd.org/Index.aspx?datasetcode=EAG_EARNINGS. on 17-November-2022.

Parunak, H. V. D., Savit, R., and Riolo, R. L. (1998). Agent-based modeling vs. equation-based modeling: A case study and users' guide. In *International workshop on multi-agent systems and agent-based simulation*, pages 10–25. Springer.

Peters, O. (2019). The ergodicity problem in economics. *Nature Physics*, 15(12):1216–1221.

Peters, O. and Adamou, A. (2022). The ergodicity solution of the cooperation puzzle. *Philosophical Transactions of the Royal Society A*, 380(2227):20200425.

Reichelt, M., Collischon, M., and Eberl, A. (2019). School tracking and its role in social reproduction: Reinforcing educational inheritance and the direct effects of social origin. *The British Journal of Sociology*, 70(4):1323–1348.

Ruggera, L., Kilpi-Jakonen, E., and Erola, J. (2023). Intergenerational social mobility of cohorts born from 1951 to 1980 in finland. *SocArXiv*. Retrieved from https://osf.io/u8qsb/download.

Sieben, S. and Lechner, C. M. (2019). Measuring cultural capital through the number of books in the household. *Measurement Instruments for the Social Sciences*, 1(1):1–6.

Špaček, O. (2017). Measuring cultural capital: Taste and legitimate culture of czech youth. *Sociological Research Online*, 22(1):113–129.

Statbel (2021). Statistique des naissances et de la fécondité. Data set, Retrieved from https://statbel.fgov.be/en/themes/population/natality-and-fertility#figures. on 13-July-2023.

Stojkoski, V., Jolakoski, P., Pal, A., Sandev, T., Kocarev, L., and Metzler, R. (2022). Income inequality and mobility in geometric brownian motion with stochastic resetting: theoretical results and empirical evidence of non-ergodicity. *Philosophical Transactions of the Royal Society A*, 380(2224):20210157.

Tang, Y., Parsons, S., and Sklar, E. (2006). Modeling human education data: From equation-based modeling to agent-based modeling. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*, pages 41–56. Springer.

Torche, F. (2015). Analyses of intergenerational mobility: An interdisciplinary review. *The ANNALS of the American Academy of Political and Social Science*, 657(1):37–62.

Torche, F. (2019). Educational mobility in developing countries. *UNU-WIDER*, 2019. Retrieved from https://doi.org/10.35188/UNU-WIDER/2019/724-8.

Turgut, Y. and Bozdag, C. E. (2023). A framework proposal for machine learning-driven agent-based models through a case study analysis. *Simulation Modelling Practice and Theory*, 123:102707.

Van Bavel, J. (2012). The reversal of gender inequality in education, union formation and fertility in europe. *Vienna Yearbook of Population Research*, pages 127–154.

van der Weide, R., Lakner, C., Mahler, D. G., Narayan, A., and Ramasubbaiah, R. (2021). Intergenerational mobility around the world bank policy research working paper. Technical Report 9707, World Bank.

Van Leeuwen, M. H. and Maas, I. (2005). Endogamy and social class in history: An overview. *International review of social History*, 50(S13):1–23.

Verbeke, G. and Molenberghs, G. (2013). Generalized linear mixed models–overview. *The Sage handbook of multilevel modeling*, pages 127–140.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442.

World Bank (2023). Adolescent fertility rate (births per 1,000 women ages 15-19) (sp.ado.tfrt). Data set, Retrieved from https://datacatalog.worldbank.org/public-licenses#cc-by. on 13-July-2023.

World Health Organization (2023). Total fertility rate (per woman). Data set, Retrieved from https://www.who.int/data/gho/indicator-metadata-registry/imr-details/123. on 21-April-2023.

Yang, X. and Zhou, P. (2022). Wealth inequality and social mobility: A simulation-based modelling approach. *Journal of Economic Behavior & Organization*, 196:307–329.

AFDELING
Straat nr bus 0000
3000 LEUVEN, BELGIË
tel. + 32 16 00 00 00
fax + 32 16 00 00 00
www.kuleuven.be