Understanding long-term labour market dynamics under deep uncertainty

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in **Engineering and Policy Analysis**

Faculty of Technology, Policy and Management

by

Lucas van Opstal

Student number: 4552555

To be defended in public on August 28 2023

Graduation committee

Chairperson	: Prof. Dr. C.P. van Beers, Economics of Technology and Innovations
First Supervisor	: Dr. ir. W.L. Auping, Policy Analysis
Second Supervisor	: Dr. ir. E. Schröder, Economics of Technology and Innovations

Understanding long-term labour market dynamics under deep uncertainty

A system dynamics approach for a labour market system

Lucas van Opstal — 4552555
Master thesis for: Engineering and Policy Analysis

August 11, 2023

Executive Summary

The Netherlands is subject to labour scarcity across a variety of industries. While labour scarcity has occurred in the past, only now the consequences of higher educational attainment have become prevalent in the Netherlands. Due to a shortage of vocational workers and a population that is getting older, circumstances ask for adequate policy-making. Although policies like increasing the formal retirement age are put in place to negate labour scarcity, it remains difficult to understand the long-term outcomes of policy implications made in the present.

Understanding this long-term behaviour of the dutch labour system is deemed useful to develop strategies that result in a well-functioning labour market. Since the labour market is intertwined with other systems (e.g. population, educational system and the economy) and is subject to uncertain future developments it results in increased complexity of understanding long-term dynamics. This research aims to develop a better understanding of the long-term behaviour of these systems in deeply uncertain futures.

To do so, an highly aggregated system dynamics (SD) model has been developed that represents the labour market and its connection with the population, educational system and economy. The existence of stocks, flows, delays and feedbackloops between these systems make it appropriate to be modelled as an SD model. Furthermore, the model is subject to deeply uncertain futures, to which no probability distributions can be attributed. To enhance the understanding of behaviour in these futures several scenarios were developed with respect to technological development, immigration and the development of GDP.

The model has been used in an exploratory modeling analysis framework to deal with these uncertainties. By sampling thousands of times over the different uncertainties and scenarios the model has been simulated for a total of 15000 times. This resulted in 15000 time series outcomes running from the years 2020 through 2060 for several model outcomes that were determined to be able to asses the quality of the labour market. Then statistical and rule-induction algorithms have been applied to the data to obtain insightful results on which policy strategies could be created.

The results focus on the net FTE scarcity and unemployment rate in the Netherlands. The model showed that the net FTE scarcity may be resolved in the short term if a policy is put in place that results in an increase in the average amount of FTE's a worker supplies. For longer term implications the technological development was mainly driving behaviour in the scarcities. The model suggests that solely increasing the average educational attainment will not resolve the problems. Furthermore an inflow of immigrants in the workforce may slightly resolve some of the scarcities. However, if the inflow of immigrants persists, and these immigrants have difficulties in finding appropriate jobs the unemployment rate will steadily rise.

It was concluded that the increase in educational attainment in itself would not necessarily pose or resolve any problems. Rather, it is wise that policy makers make an effort to design study programmes that are catered towards technological development. This means that both vocational educated people, as well as academically educated people are learnt the adequate skillset to participate in a labour market that is subject to technological development. Then, the unemployment rate increased to high levels in the cases where high immigration levels remained constant. These immigration regimes happened in simulations in which individuals had difficulties in finding or starting a job. This result may help policy makers and local governments to design programmes that aid in

a smooth transition from being unemployed to finding a job.

This has resulted in a potential policy design that may be used as a starting point for developing policy to lower scarcities and unemployment rates when they reach unsustainable high levels. The model showed that a policy that will increase the average FTE per worker would be helpful. This can be incentivised in a few ways. Firstly, individuals receiving unemployment benefits should be permitted to work a few hours per week without it impacting their eligibility for those benefits. Secondly, it should be made fiscally attractive to work for a second employer. And thirdly, allowing people to work more than 40 hours, in which the extra hours are paid out as gross salary. This policy in combination with an immigration regime with short hiring times can help to ease the labour scarcities.

The impact of technological advancement is difficult to asses. However, it is still advised to policy and educational institutions to be an early adopter of those new technologies. By exposing primary and high school students to courses related to technology (e.g. programming), these students can make a better informed decision when pursuing a degree in tertiary education. Then, the mbo, hbo, and wo schools should have a comprehensive understanding of current trends happening in the labour market and adapt their programmes accordingly such that the new supply of workers can resolve some of the scarcities.

Acknowledgements

Finalising this thesis means the end of 7 years of studying in Delft. In the summer of 2016 I started studying the bachelor Systems Engineering and Policy Analysis and Management here at the faculty of Technology, Policy and Management in Delft. I especially liked the modelling and programming courses in the bachelor and hence the choice for Engineering and Policy Analysis seemed trivial. The System Dynamics and Exploratory Modelling Analysis courses within the EPA programme have given me the inspiration to develop a thesis subject related to long term dynamics of an highly aggregated system.

During my thesis I received support and feedback from my supervisors and chair. I would like to thank Prof. Dr. van Beers for his position as chair and providing an unbiased view on the thesis. My gratitude also goes out to my second supervisor, Dr. Schröder. His macro-economic view has helped me better understand concepts that were unknown to me before. His critical view on System Dynamics modelling has helped me reassess the model multiple times and critically review my modelling choices.

Special thanks goes to my first supervisor Dr. ir. Willem Auping. During our weekly meetings he supported me on both a technical level as on a project level. His expertise in SD and EMA have helped me multiple times when I felt stuck during the project. I have appreciated his clear, concise and fair way of communicating throughout the project which has eventually led to me finalising this thesis.

I also want to thank my family and friends for always supporting me throughout my studies. I know it has been a long ride and I do not take the standard routes but I always felt supported.
Finally, I would show my appreciation to Amber. The past 6.5 years she has had my back throughout most of my studies in Delft. Her ambition has pushed me to pursue a double degree, and she is therefore the reason I am writing this thesis. Moreover, I have to credit her for coming up with this thesis topic during a car drive we had last summer. Thank you!

Contents

1 Introduction

Since the outbreak of the COVID-19 pandemic, the Netherlands is struggling with a labour shortage across a variety of industries (CBS, 2022a). Unemployment in the Netherlands is at one of its lowest points in history and the Ministry of Social Affairs and Employment has constructed a plan to ease the labour shortage in the Netherlands (Ministry of Social Affairs and Employment, 2022). The analysis from the Ministry, literature, and news articles mention several causes of the labour shortage. One of these causes is related to the increasing educational attainment in the Netherlands which results in shortages in the vocational labour workforce. The Ministry of Education, Culture and Science forecast that the amount of students applying for a secondary vocational programme (mbo) will decrease with 5.4% by 2028 (Ministerie van Onderwijs, Cultuur en Wetenschap, 2022b) while the number of university students will increase with 13.6% (Ministerie van Onderwijs, Cultuur en Wetenschap, 2022c).

The EU has developed a policy stating that 45% of the 25 through 35 year-olds of every member state should have obtained a tertiary education degree by 2030 (Eurostat, 2022a). The Netherlands has already met these requirements in 2022, since at this point in time, at least half of this age group hold a tertiary education degree (OECD, 2022a). Furthermore, universities in the Netherlands see an influx of international students that have a tendency to stay permanently and participate in the labour market (Elfferich, 2022).

Another cause of the labour shortage is related to societal ageing in the Netherlands. Levanon et al. (2014) found that one of the driving factors for the labour shortage in the United States is the outflow of the 'baby boomer' generation from the labour force. Research institution NIDI confirms this dynamic is also happening in the Netherlands and concludes it will stabilize around 2050 (NIDI, 2021). Another cause that seems to be especially relevant in the Netherlands is the high percentage of part-time versus full-time employees in the labour force (Eurostat, 2022b). A direct result of this is that the total amount of available labour hours is much lower than the maximum labour capacity (e.g. a worker who could work 40 hours a week is only working 20 hours a week).

The societal ageing in the Netherlands has three main causes; a decreasing fertility rate, an increasing life expectancy and the size of the current 'baby boomer' age cohort (Auping et al., 2015; Logtens et al., 2012; Sutrisno & Handel, 2011). In 2020 the fertility rate in the Netherlands was estimated at 1.6 (Worldbank, 2020), which is considered to be far below an 'healthy replacement' level of 2.1 when controlling for immigration and mortality rates (OECD, 2022b). To ease the effect this has on the labour market the Dutch government is steadily increasing the retirement age (Ministry of Social Affairs and Employment, 2022). While this may resolve some labour shortage issues, there exists evidence that questions the effectiveness of this policy in the real world. Rutten et al. (2023) have found that there exists a direct effect between the retirement age of the eldest spouse and the age on which the younger spouse stops participating in the labour market. This implies that the younger spouse will retire before the actual formal retirement age and thus prematurely leave the labour force. Furthermore, they have found that higher income households tend to exploit pension benefits before their formal retirement age.

The causes of the labour shortage in the Netherlands is thus two-fold. Firstly, the ratio of vocational to tertiary educated students is decreasing, resulting in a shortage of skilled workers. Secondly, the pressing effects of societal ageing on the labour market. The actuality and the consequences of a labour shortage make it a relevant problem for policymakers to create robust and effective policy

measures. While the Dutch Ministry has created a plan for tackling these issues, the proposed solutions focus mostly on societal ageing rather than on long-term issues related to the education system. The currently offered solutions in the report of the Ministry include: Stimulating innovation; increase total labour hours; increase wages; utilise underemployed part-time workers and increasing the formal retirement age.
Therfore it seems that there are a vast majority of policy levers that can be operated by different stakeholders. Yet, in the literature, it remains unclear what the long-term dynamics of the system will be when these different policy levers are being used. It is important to understand the long-term implications of policies. Even when short-term effects might reap benefits, the long-term effects may prove detrimental (Forrester et al., 1976).

The ample policy choices central and local governments face make it complex to assess what would happen to the labour market in the short- and long run. However, models can be of help when answering so-called *'what-if'* questions (Oreskes et al., 1994). While a model will not yield answers to what will happen in the future, it may support researchers and policymakers to understand the possible long-term implications of policy choices made in the present.

The context description of the problem results in the following research question: *'How do educational attainment and societal ageing influence the long-term labour market dynamics in the Netherlands?'*

Answering this question will provide insights to the scientific field of macro-economists and policymakers that operate within this field. As mentioned, the problem is EPA-relevant due to the described deep uncertainty, ample policy choices, and the benefits for society if the dynamics of the labour market are better understood. To answer the main research question the following sub questions are proposed:

1. How do the population, educational system, and economy influence the labour market in the Netherlands?

2. How do uncertain drivers in those systems influence the labour market?

3. How can the quality of the labour market be measured and quantified?

4. Which robust policy measures can be imposed to maintain a well-functioning labour market under different uncertain scenarios?

2 Background

This section aims to answer the first research subquestion: How do the population, educational system, and economy influence the labour market in the Netherlands? By covering the different subsystems relevant for this research the structure of the aggregated system can be understood. Thereafter, possible structural uncertainties can be illuminated.

2.1 Dutch population dynamics

The demographics of a country have an influence on the labour market since it will be the basis of the supply of the domestic workforce. As stated in the introduction, people are getting older and hence the ratio of retirees to the working-age population will increase ceteris paribus. The labour force in the Netherlands is defined as the population that currently is working, or has recently been searching for a paid job, aged between 15 and 75 years old (CBS, 2023a). Currently the Ministry of Health, Welfare and Sport have made forecasts available for the life expectancy of males and females in the Netherlands up until 2070. It is expected that in 2070 male life expectancy will be 87.5 years and female life expectancy will be 91.1 years (Ministry of Health, Welfare and Sport, 2022). Due to the increase in life expectancy the Dutch government needs to increase the formal retirement age accordingly. In 2020 the government introduced a law to increase the formal retirement age with 8 months for a one year increase in the general life expectancy (Rijksoverheid, 2020). Note that the formal retirement age in 2023 is set at 66 years and 10 months, which is lower than the required 75 years old to be part of the working-age population (Rijksoverheid, 2020).

The domestic population is not the only factor that influences the supply of the labour force. Immigration will also be a driver for the labour market, especially work-related (seasonal) immigration and international students who study and stay will have an impact. Work related immigration happens across all industries, where lower educated immigrants mostly work in transport and agriculture, and higher educated immigrants will work more often for bigger corporations (ABU, 2022; CBS, 2022d). Furthermore, another inflow of immigrants happens within the educational system, since almost 40% of first-year students are foreign, which tend to stay and join the labour force (Elfferich, 2022; Tweede Kamer der Staten Generaal, 2023).

2.2 The Dutch education system

The educational system will eventually determine the skill set of the population entering the labour market for the first time. The Dutch education system can best be subdivided in three groups: primary, secondary and upper-secondary & tertiary education. Children in the Netherlands are subject to compulsory education from ages 5 through 16. Adolescents aged 16 through 18 with no basic qualification (i.e. a diploma obtained from upper-secondary or tertiary education) will need to continue studying until they have reached the age of 18.

2.2.1 Primary education

Most children in the Netherlands will start primary education at the age of 4 years old (Rijksoverheid, 2023e). In general, students will be following this type of education up until the age of 12 years old. During this trajectory the capabilities of each student is assessed. In the final year of primary school, students will make a final examination test which will give insights in the skill set of each student (Rijksoverheid, 2023b). Based on these results the primary school determines to which level of secondary education the student may go.

2.2.2 Secondary education

The dutch secondary education system is subdivided in three main categories, namely: vmbo (pre-vocational secondary education), havo (senior general secondary education), and vwo (pre university education). While some of the levels can be further subdivided based on specifications of the curriculum, for this thesis the aggregated perspective of just three levels is taken. Vmbo is a four-year programme offering theoretical and practical courses. It prepares students to have an smooth transition to mbo (upper secondary education) (Rijksoverheid, 2023a). Havo will prepare students for higher professional education. Havo will have an increased focus on theoretical aspects of learning compared to vmbo, but this will be to a lesser extent compared with vwo. The havo programme generally takes five years to complete, after which students will mostly go to the hbo (higher vocational education) level. Finally, vwo prepares students for studying at a university. The programme covers more theoretical and complex topics compared to havo. Generally the programme takes six years after which students will go to wo (university) education. (Rijksoverheid, 2023c).

2.2.3 Upper-secondary and tertiary education

After students have finished their high-school they will go to the respective last step of their educational career. While some students may pursue a PhD after wo education, this is assumed to be of little relevance for this thesis. Vmbo students will follow an mbo programme which usually takes four years. This may vary between which specific study programme they chose and if they want to continue studying after having finished their studies. Students from havo will follow a programme of choice at hbo level, and this usually takes four years to complete. Finally, vwo pupils have the possibility to go to a university and follow a programme of their choice. Depending on the curriculum this will take around four to five years, but in some cases (e.g. medicine and dentistry) this will take longer.

Note that for secondary levels and above, students who are performing well have the possibility to upstream during or after completing their studies. For instance, a havo student with high grades does have a possibility to upstream to vwo during their highschool period. Students who are performing poorly will downstream voluntarily or obligatory in some cases. For higher educational levels, that is, upper-secondary and tertiary, it is less common to upstream to another level during studies. This is due to the fact that the curriculum of programmes at mbo, hbo and wo levels, even if it is related to similar industries, is very different. Hence, students have the possibility to start a new study at a higher level, after having completed their current study. Sometimes schools and universities will offer so called 'pre-master' programmes which fast-track students to a master programme instead of doing a bachelors curriculum all over again.

2.2.4 The dutch government and the education system

To maintain high quality of education and to stay relevant for international students, both the government and schools interact with each other to reach these goals. The past years the Ministry of Education, Culture and Science have seen a decline in the basic skill levels of math and the Dutch language of students in primary and secondary education (Ministerie van Onderwijs, Cultuur en Wetenschap, 2022a; NOS, 2023). Furthermore, the government is planning to ease the threshold for being allowed to continue a study programme at hbo or wo after your first year. Currently there is a threshold in place called the 'Bindend Studie Advies' (bsa) which states that students will need to pass a certain amount of courses to be allowed to continue their study programme. Currently, most hbo and wo institutions ask that students pass at least 75% of the courses to be able to continue next year. However, the Ministry of Education, Culture and Science see that students experience more stress due to the current bsa regulations, and wants to lower the threshold to 50% of the courses (Rijksoverheid, 2023f).

2.3 The Dutch economy

The economy of a country is one of the core systems which can be considered in modelling an aggregated model of a nation. However, for this research the main scope will be on the interplay between the population, educational system and labour market for a longer time horizon. Therefore, it is chosen to neglect the inner workings of the Dutch economy and assume the most important behaviour as exogenous. It is therefore important to at least have some understanding of the output of a country and the different industries that make up the whole economy of the Netherlands. This output will be directly related to the labour market and hence will help to answer the first research subquestion.

For this research the output of the Netherlands, measured by GDP, will be determined as the sum of the output of each industry residing in the economy. That is:

$$GDP = \sum_{i=1}^{N} GAV_i, \tag{1}$$

where GAV_i is the 'Gross Added Value' of industry i, and N gives the number of industries in the Netherlands. To determine the industries the standard industry index (SBI) is being used. This yielded twenty industries and their descriptions can be found in the appendix A in table 5. For a more elaborate explanation of each industry please see Kruiskamp (2022). The exogenous development of GDP will help determine the development of the output of each industry and this will have an impact on labour demand. An important assumption that is being made here is that the industry level output will grow proportionally at the same rate as GDP. This means that the prices within each industry remain the same during the whole time horizon. A more elaborate explanation on the calculations will be given in the section 3.6.

2.4 Connection with the labour market

Now that the needed background context of the subsystems has been sketched, the most important influences on the labour market can be illuminated. It can be concluded that the Dutch population dynamics is a process on its own and can hardly be influenced by exogenous factors. Only immigration and fertility rates will have an impact on the population, and hence the size of the workforce. However, the government can tweak the formal retirement age to a certain degree. Obviously this will be a highly political debate and is a far more complex issue than will be assumed in this thesis.

The Dutch educational system will supply the future labourers and will most importantly determine their skill set and in which industries people may want to work. The government will aim to retain a high quality education system, and a system that will provide sufficient labour for potentially growing industries (e.g. the need for people to work with AI will most likely increase). Furthermore, the Dutch government provides subsidies for retraining purposes, called the 'STAP-budget'

(Rijksoverheid, 2023d). With this policy the government aims to give labourers more opportunities in the job market, especially in industries where the need for labourers is the largest.

Finally, the exogenous development of GDP will determine the development of the output of each of the twenty industries. With this output, in combination with labour productivity it can be determined what the labour hour demand is of each industry. This results in a labour market system where labour demand is determined by the economy and the supply of labour is determined by the population and educational system.

3 Methodology

In this section, the methodology of this thesis is elaborated upon. First in section 3.1 different modelling techniques are considered. Then, in section 3.2 the relevance of System Dynamics (SD) to answer the research question is discussed. In section 3.3 Exploratory modelling Analysis (EMA) for dealing with the uncertainties is explained. Fourth, an highly aggregated conceptual model and its most important relations are discussed in section 3.4. Section 3.5 briefly discusses the used data sources. In section 3.6 the four main submodels are explained. Then, in section 3.7 the validity of the model is discussed. Then, section 3.8 discusses the most relevant variables through an XLRM-framework. Thereafter in section 3.9 the setup for the experiments is discussed. Finally in section 3.10 the methods for scenario discovery are elaborated upon.

3.1 Modelling techniques

As argued in the introduction, the labour market is a complex dynamic system. In this system a lot of individual actors reside which make choices based on individual behaviour. Agent-Based Computing (ABC) is a possible way to asses what the aggregated behaviour of a system in which agents make choices will look like. As first argued by Jennings (1999), ABC can help understand how autonomous agents complete their objectives while residing in a dynamic and uncertain environment. The use of ABC has evolved into the more general Agent Based modelling (ABM) technique and is being widely applied for various domains (see Niazi and Hussain (2011) for a comprehensive visual survey). While ABM may prove useful in modelling some dynamics, as in Neugart and Richiardi (2018), it primarily focuses on the behaviour of *individual* agents in the system. For the scope of this research a more aggregated macro view of the system is chosen, and hence ABM does not seem appropriate.

Another modelling technique that could be considered is Discrete Event Simulation (DES). In a DES model the simulation clock keeps track of the time and an event scheduling method determines what will happen in the model (Fishwick, 1995). The technique is based on discrete events having an influence on the variables in the system. While certain events can be modeled or taken into account (e.g. recession or pandemic) this lies out of the scope of the research question, which looks at the system in a more continuous manner over a time horizon covering several decades.

A third modelling technique which is considered is System Dynamics (SD). This technique was first written about by MIT professor Jay W. Forrester (1961). An SD model uses integral equations to calculate the evolvement of a 'stock' variable during the simulation time. Within the stocks material or information can accumulate over time and leave the stock after a certain delay. Through auxiliary variables, different equations can be defined which form the basis of feedback loops within the described system. Within the described labour market in the introduction it can be argued that there exists delays, accumulation (of people) and feedback loops between other systems (e.g. education and population). Since the research question focuses on an aggregated system over a long time horizon the use of the SD modelling technique is proposed. In section 3.2 this motivation is further elaborated upon.

3.2 System Dynamics

To be able to make changes in this system (i.e. implement policy) it is important for the actors in the system to understand it's dynamics. Otherwise resistance to the proposed policies might arise.

Sterman (1994) argues how ways of 'system thinking' will nullify 'policy resistance'. To help actors with this proposed way of thinking, an SD model will be of great help (Sterman, 2001). While these computer simulated models will provide useful insights, it is of paramount importance to completely understand the 'feedback, stocks, flows, time delays and non linearity' of the system to make informed decisions (Sterman, 2002). Hence, a system is especially useful to model as an SD model if there exists accumulation, delays, and feedback loops related to important variables in the system. These three characteristics can be identified in a labour market system.

Since a labour market consists of several groups of workers, which can be characterised by their educational level and skill set, they can be regarded as a stock variable which is subject to the in- and outflow of workers. The stock variables will increase either by an inflow of newly graduated students, or immigration. Outflow of the workforce may happen through retirement or emigration. A similar line of reasoning can be followed by the educational system in the Netherlands, where there exists an in- and outflow of students in specific educational levels.

Secondly, the labour market is subject to delays in its system. If there exists a shortage of workers in a specific sector it might become attractive for students to follow education catered to these jobs. This will not immediately solve the shortage, but only once the students have finished their education and enter the labour market.

Third, feedback loops play an important role in the system. Students in the education system perceive the importance of higher or vocational education. This will in turn, with a delay, have an impact on the ratio of vocational to academic workers in the workforce. This ratio impacts the demand for specific skills, and hence the perceived importance to study for these skills, and hence implies a balancing feedback loop.

3.3 Exploratory modelling & Analysis

The labour market system described in the introduction can be characterised as wicked and messy (Ackoff et al., 1974; Rittel & Webber, 1973). These kind of problems are subject to deep uncertainty (Bankes, 1993; Lempert, 2003). Deep uncertainty refers to the kind of uncertainty to which no probability distributions or mathematical formulas can be attributed. Certain exogenous drivers of the system, like technological development, will yield different results in different future scenarios. Think of the development of AI tools that will disrupt and create jobs in certain industries. This will have a direct effect on the labour market, and with a delay on the educational system in the Netherlands. Furthermore, immigration can be described as an uncertain exogenous factor in the described system to a certain degree. Climate disasters or disrupting wars in other parts of the world may have an increasing effect on the inflow of immigrants in the Netherlands.

Most real world public decision making is subject to deep uncertainty. This poses challenges when interpreting models which aim to describe a system in the real world. Exploratory Modelling is a developed methodology to aid policy makers in understanding deep uncertainty (Bankes, 1993). The Exploratory Modelling and Analysis workbench is a tool developed to tackle these challenges (Kwakkel, 2017). The workbench is able to aid modelers in discovering certain scenarios which yield unfavorable or unacceptable outcomes if a certain policy is implemented. By running the model hundreds or thousands of times with varying uncertain parameters, the modeller is able to identify which policies are robust when considering certain objectives.

3.4 Conceptual model

An highly aggregated conceptualisation of the system can be seen in figure 1. Key subsystems are depicted by rounded rectangles. Variables in **boldface** are policies, and variables in *italics* are exogenous uncertain drivers in the system. At the base of this model is a population subsystem based on the model by Pruyt and Logtens (2015). The remaining four subsystems either have a direct or indirect influence on each other. Within the demography sub-system different age groups are divided in one year cohorts. Several cohorts will enter (at 15 years old) and leave (at the formal retirement age) the labour market.

Within the labour market sub-system a distinction between twenty industry branches is made. Think of workers entering the healthcare, government or financial industry. Both the retirement age and the demography will affect the workforce pool in each industry. Furthermore, the education subsystem will also determine what parts of the population will go work in which industry.

The education sub-system has a direct impact on the labour market. Within the education system students are categorised by their level of education. This part of the model determines the direct inflow of vocational and academic workers that can supply labour. The sub-system is also subject to government policies which are able to change the inflow of (international) students in both the secondary as tertiary segments.

Within the economy sub-system GDP and sector output are modeled. The GDP determines in which proportion each sector output grows and this will eventually determine in part labour demand. The economy is considered as an exogenous developing system for this research.

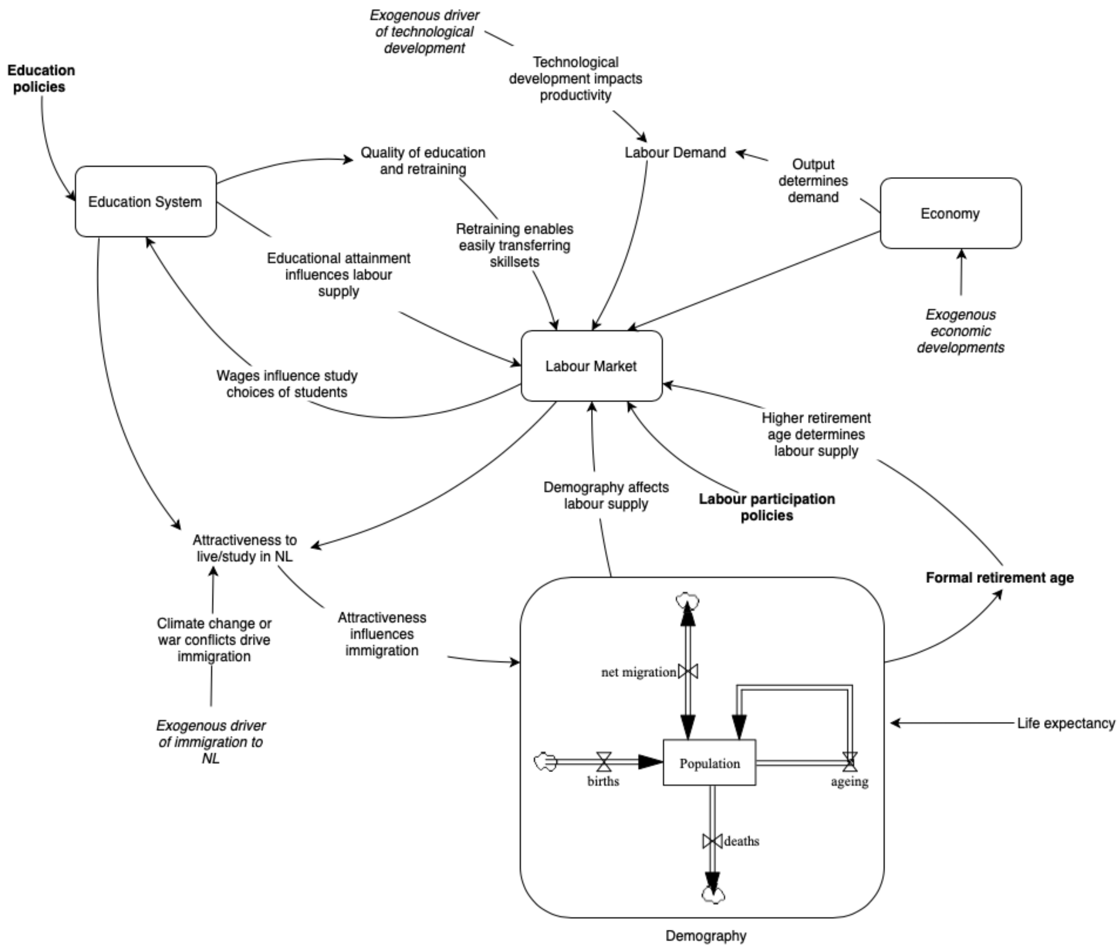


Figure 1 Sub-system Diagram

3.5 Data

To be able to model several submodels it is necessary to collect certain data variables. These will be specifically relevant for the current population, workforce and educational attainment in the Netherlands. A few different databases have been used with in particular the dutch CBS database. This contains accurate values of the initial population, sorted by age cohorts and data relevant for the educational system as well as the labour market. Other data sources that were used can be found on the websites of OECD, Worldbank and ROA. Where used these will be referenced in the text.

3.6 Submodels

The developed sub-system diagram aids with understanding how different parts of the model will interact with one-another. In this subsection the submodels will be further elaborated upon. To be able to understand the equations used in the model a brief introduction to the Ventana Vensim modelling software is given in section 3.6.1. Sections 3.6.2 through 3.6.5 will explain the inner workings of each submodel in more detail.

3.6.1 System Dynamics modelling conventions

This subsection describes the modelling conventions within the System Dynamics paradigm. The explanation is largely based on chapter 1.1.1 of the work of Auping (2018). Within most SD models stocks, flows and delays are modelled as time-series variables over a specified time-horizon. A stock can be specified as as the sum of an initial value and an integral over the inflow and outflow of 'material' or 'information' within this stock. This can be written as:

$$s(t) = s(t_0) + \int_{t_0}^{t} [f(t) - g(t)]\, dt, \tag{2}$$

where $s(t)$ is the value of the stock at time t, t_0 the initial time and $f(t)$ and $g(t)$ the in- and outflow in this stock at time t, respectively. A simple depiction of a stock-flow diagram can be seen in figure 2. In the figure the three constants are are added which are not a function of time. The inflow $f(t)$ uses $s(t)$ and c_1 as an input. The auxiliary variable $a(t)$ uses $s(t)$ and c_2 as inputs. And finally the flow $g(t)$ is a function of $s(t)$, $a(t)$ and c_3. In addition the inputs can be modelled in such a way that they have a delayed effect on the other variables.

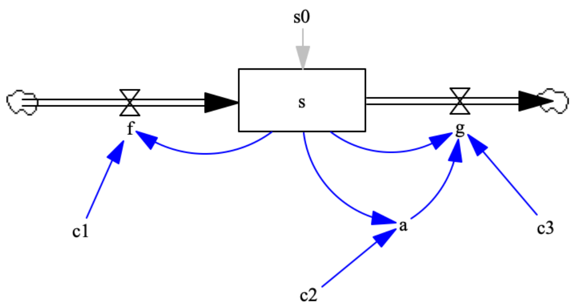


Figure 2 A simple stock-flow SD model, (Auping, 2018)

The combination of multiple of these stock flow systems that are interconnected result in a vast amount of integral equations. A commonly used SD modelling software, Vensim (Ventana Systems, 2010), exploits numerical integration methods to solve these integrals.

Due to the similar structure of the stocks in the submodels, the integrals will always be depicted with lower limit t_0 and upper-limit t. Each stock will also always be accompanied by its initial value $s(t_0)$. To avoid repetitive writing, the explanation of these variables will be assumed to be trivial and hence omitted in the text in the next sections.

3.6.2 Population

The population submodel is an important driver of the other submodels. It is therefore important that the modelling of this part is done in a correct and realistic manner. To reach the needed quality, the submodel of this theses is heavily inspired by the population model created in (Logtens et al., 2012). By using recent data from CBS the model is adapted to work with more recent numbers. The model uses two main stocks of the male and female population which are all subscripted in 106 age cohorts, where the subscript mage0 and fage0 refers to all the males and females with an age between 0 and 1 years old respectively. Persons can enter the system either through birth or through immigration. A person will leave the system upon death or emigration out of the Netherlands. A person will move to the next age cohort through an ageing flow variable. To model this dynamic, two different equations are required, one for age cohort 0 and one for age cohorts 1 through 106. This yields

$$P_0(t) = P_0(t_0) + \int_{t_0}^{t} [births_0(t) - ageing_0(t) - deaths_0(t) + migration_0(t)]\, dt, \tag{3}$$

where the subscript 0 refers to the age cohort, $P_0(t)$ is the population of people aged between 0 and 1, and the other variables are trivial. For the age cohorts 1 through 106 this yields,

$$P_a(t) = P_a(t_0) + \int_{t_0}^{t} [ageing_a(t) - deaths_a(t) + migration_a(t)]\, dt, \tag{4}$$

Where the difference with equation 3 is only the subscript a which can take values 1 through 106. The initial values of each age cohort are obtained from the CBS database (CBS, 2022c).

The flow variable for births is calculated with help of some auxiliary variables. To calculate the births it is assumed that the fertile years of women are between 15 and 50. With these values the total female fertile population can be calculated. Then the fertility rate of women is divided by the total fertile years of women, being $50 - 15 = 35$ to obtain the average number of children per female per year. This results in the flow variable births being

$$births(t) = child(t) * 0.49 * fertile\ population(t) * (1 - death\ rate\ at\ birth(t)), \tag{5}$$

where $child(t)$ refers to the average number of children per female per year, 0.49 is the female fraction of newborn, and the death rate at birth is further elaborated upon in the next paragraphs.

To estimate the mortality rate of people the Gompertz-Makeham law of mortality is used. This mathematical law uses insights on mortality from Gompertz (1825) and Makeham (1860). More recent literature discusses the validity of the Gompertz-Makeham law of mortality for extreme ages (above 106) (Gavrilov & Gavrilova, 2011). The probability density function of mortality yields:

$$f(x) = (\alpha e^{\beta x} + \lambda) \cdot exp[-\lambda x - \frac{\alpha}{\beta}(e^{\beta x} - 1)], \tag{6}$$

Where the variable x denotes the age of a person, and the variables α, β and λ need to be estimated through available data on mortality for a specific country. The authors propose using a logistic model for estimating mortality at extreme ages. For this thesis these models are out of scope and also of little relevance since the oldest age cohort is assumed to be of 106 years old, after which the probability of death is assumed to be 1. Furthermore, an already developed function from the population model of Logtens et al. (2012) in the SD modelling software has been adopted and used for calculations.

The net immigration of people in the population submodel is dependent on several scenarios. A baseline immigration scenario is created with existing data on immigration and emigration (CBS, 2020c). Three uncertain scenarios are created which will be used in the experiments to asses model behaviour if these scenarios might occur. These uncertain scenarios will not have an influence on the base-case model and will be further elaborated upon in section 3.8.

3.6.3 Education

The educational system in the Netherlands is subdivided into three stocks, namely: population in primary, population in secondary and population in upper secondary and tertiary. The latter two aggregated stocks can be divided in subcategories. A person currently following secondary education is assumed to be doing that at 'vmbo, havo or vwo' levels. Here, vmbo loosely translates to: 'pre-vocational secondary education'; havo to 'senior general secondary education'; and vwo to 'pre-university education'. The educational system in the Netherlands works in a way that the level of high-school education determines to which level of upper secondary or tertiary education a student will go. Vmbo will flow to mbo, havo to hbo and vwo to wo. Here mbo translates to upper secondary vocational education, hbo to higher vocational education and wo to university education. For ease of writing all three, mbo, hbo, and wo, are referred to as tertiary education unless stated otherwise.

It is assumed that all 4-year olds will start primary education and stay on average in this stock for 8 years. During those 8 years an inflow of students may happen through immigration. Outflow will occur upon death or when a student prematurely leaves the educational system. These two outflows will influence each of the three aggregated stocks in their own way based on relevant data (CBS, 2020a, 2020f; OECD, 2022a).

After 8 years students will be divided based on their skill level and enter the appropriate secondary education level based on current data (Ministry of Education, Culture and Science, 2020). A student stays in the stock for the amount of years it takes (on average) to complete high-school. For vmbo this is 4 years, havo this is 5 years and for vwo this is 6 years. Students do have the possibility to move to a 'higher' or 'lower' level if their skill set deviates from what is expected from a student in that level. Hence, fractions of the student population will upstream or downstream (CBS, 2020e).

Finally, students will enter mbo, hbo, or wo education. In a similar way as before, students may flow out through deaths or prematurely leave the educational system. The possibility to up- or downstream also exists in this stock, and after several years all students will leave having completed their studies (CBS, 2020b, 2020d). It is therefore assumed that the highest attainable education is wo, and thus PhD's are not considered in the model.

The three main stocks in the educational model have very similar workings as the population stocks described in equation 3. This yields,

$$\begin{gathered} P_{prmy}(t) = P_{prmy}(t_0) + \int_{t_0}^{t} [edu\ inflow(t) - deaths_{prmy}(t) + \\ migration_{prmy}(t) - premature\ leave_{prmy} - prmy\ to\ scnd\ flow(t)]\, dt, \end{gathered} \tag{7}$$

where $P_{prmy}(t)$ is the population currently following primary education, the subscript $prmy$ refers to people in primary education, $edu\ inflow(t)$ is the domestic inflow of 4-year olds in primary education, $premature\ leave_{prmy}$ is the fraction of people that leave primary education without finishing it and *prmy to scnd flow*(t) the flow of people that enter secondary education after finishing primary education.

The stock variables for secondary and tertiary education work in a similar way as the the population in primary stock. People can only enter these stocks through immigration or by flowing from primary to secondary or secondary to tertiary. Within these stocks up- and downstream is also taken into account. This yields,

$$\begin{aligned} P_{scnd}(t) = P_{scnd}(t_0) + \int_{t_0}^{t} [&prmy\ to\ scnd\ flow(t) - deaths_{scnd}(t) + migration_{scnd}(t) \\ &-premature\ leave_{scnd} - scnd\ to\ tert\ flow(t) + net\ up\ down\ streams(t)]\, dt, \end{aligned} \tag{8}$$

where $P_{scnd}(t)$ is the population currently following secondary education, the subscript *scnd* refer to people in secondary education and an extra variable is added to depict the up- and downstream between vmbo, havo and vwo. For tertiary education the exact same structure applies with the only difference that people do not flow to a new educational stock.

Eventually, all persons within the system will have a certain educational attainment. These will be implemented as a subscript within the model. After a person has prematurely left or completed their study their educational attainment will stay fixed. During studies the educational attainment of a person can change. For instance, a 15-year old currently following vmbo-level education will have its educational attainment subscript set at 'primary', since that person has not yet *completed* its vmbo education. If he or she continues studying and enters mbo-level education, its subscript for educational attainment will be updated to 'vmbo', since now he or she has completed its secondary education.

Each default outflow (i.e. non-death or migration outflows), is equal to the value of the stock divided by the average time of that edu system. For example, the default flow from primary to secondary is 12.5% of the current population in primary, since primary education is assumed to be always 8 years for everyone. For upper-secondary and tertiary this time per educational level may deviate and is therefore assumed to be a parametric uncertainty which will be handled in chapter 3.8.1. The tertiary outflow of the educational system is defined as,

$$tertiary\ to\ labour\ market_{tert}(t) = \frac{P_{tert}(t)}{d_{tert}} - deaths_{tert}(t) - premature\ leave_{tert}, \tag{9}$$

where d_{tert} is the average time per educational level of mbo, hbo, and wo. Furthermore, the *premature leave* of each educational system is just simply a fraction multiplied with the current population residing in that stock. Such that,

$$premature\ leave_e(t) = P_e(t) \cdot fraction\ that\ leaves_e, \tag{10}$$

where the *fraction that leaves* is also considered to be a parametric uncertainty in the model, which in general is very low.

To be able to make the connection with the labour market it is important to understand what educational attainment is preferred and is dominated within each industry in the labour market.

The labour market will be subscripted into 20 industries which will be elaborated upon in the next section. Each industry will have a preference for certain workers. For instance, the construction industry will not prefer to have university educated workers dominate the construction site, since it is assumed they will lack specific skills which will only be taught at mbo-levels. On the other hand, industries active in management consultancy will not prefer to have a workforce dominated by mbo educated persons, since it is assumed they do not have the skills needed to assess and valuate the financials of big corporates. See appendix A, table 4 for more information on how the dominant education level per industry is determined. The variable 'dominant edu level per industry' is a vector of twenty values that take on values 1, 2 or 3. This can be seen as a categorical variable representing the preference for a specific educational attainment in each industry. Where '1' refers to a preference of workers with educational attainment mbo or lower. Where '2' refers to educational attainment of hbo, and '3' refers to an educational attainment of wo. This vector will later be used to decide how the allocation of workers will be happening, described in section 3.6.4.

Finally, with the help of the dominant edu level per industry the perceived Return-On-Investment on pursuing a degree for a certain industry can be approximated. Even though it is still possible for someone with an mbo qualification to work in the finance industry, it is assumed that a student wanting to work in that industry will need to 'invest' extra study years and money to obtain a wo-degree which is preferred for these jobs. By calculating the amount of extra years an mbo student needs to study on average to obtain a degree that is preferred by an industry the total costs of studying can be approximated. This yields,

$$C_{i,mbo} = D_{mbo} \cdot 1168 + E_{mbo,i} \cdot 2143, \tag{11}$$

where $C_{i,mbo}(t)$ is the total costs for an mbo student for obtaining a degree which is preferred by industry i. D_{mbo} is the average time spent in mbo education which costs 1168 per year, and $E_{mbo,i}$ the total years of extra studying for mbo students for pursuing a degree for an industry, which costs 2143 per year. Note that these extra years of studying for mbo students will be equal to zero for industries that prefer mbo educated people, and hence the costs will be lower. For hbo students a similar equation holds, except the variables D and E are multiplied with 2143. It is assumed that people with a wo-degree will not need to add extra study years to be able to start a job in other industries. On an occupational level this would not be the case, since a wo-degree in the real world does not offer the correct qualification to become a plumber or car mechanic. Therefore it can be interpreted as if a person with a wo-degree that starts in the construction industry will be more allocated to higher-level functions or management.

To be able to calculate the ROI, both the costs as well as the returns for obtaining a higher degree should also be calculated. The costs were defined in equation 11. For the returns the average wage per month in each industry is used. By multiplying this value with 12 one obtains the total expected yearly income per industry. Since the model assumes that people with an mbo degree will enter the labour market a few years earlier than people with a higher degree, these persons will have more expected working years for specific industries (i.e. industries which only require/prefer an mbo degree). The total working years is the difference between the formal retirement age and the age when one is assumed to start working in that industry. This results in a variable called 'perceived life-time income per industry for mbo'. This variable is then divided by the costs of equation 11 and multiplied with 100 to obtain the perceived ROI per industry. Note that the ROI therefore assumes that there exists only one wage per industry that is the same for everyone, and that a person will always work in the same industry when calculating its perceived ROI.

The average wage within each industry changes throughout time. It is assumed that wages will only increase on average. The key word here is average, since it may be possible that the wage does not move for 5 years, and then suddenly increases with 10% in the 6th year. The model would assume a constant increase in the wage each year that would be equal to a 10% increase over 5 years. However, this average annual increase in the average wage is assumed to be a parametric uncertainty. This yields,

$$w_i(t) = w_i(t_0) + \int_{t_0}^{t} sw_i(t)\, dt, \tag{12}$$

where $w_i(t)$ is the average wage of industry i at time t, and $sw_i(t)$ is the flow variable describing the increase in the average wage. The latter is defined as,

$$sw_i(t) = w_i(t) \cdot x, \tag{13}$$

where x is the uncertain exogenous fractional increase in average wage.

3.6.4 Labour Market

At the base of the labour market submodel lie the two stocks Male/Female Full Time Equivalent (FTE) Employed per Industry. Both these stocks are subscripted to keep track of the amount of FTE's employed in an industry, per age group and per educational attainment. Instead of working with persons employed, it is customary to work with FTE's. Each FTE depicts a full-time work week, which is 40 hours per week in the Netherlands. For example, a person of 16-years old provides only around 0.16 FTE while a person in their 40s provide around 1 FTE. FTE's will leave this stock either permanently (retiring or dying) or temporarily by persons getting fired. In the model each industry is the aggregate of all the companies operating within the industry and hence have an aggregated demand and supply for FTE's. The demand for workers (expressed in FTE's) is calculated as follows:

$$D_i(t) = \frac{GAV_i(t)}{\rho_i(t)}, \tag{14}$$

where $D_i(t)$ are the FTE's demanded for industry i, $GAV_i(t)$ is the gross added value in base prices of industry i and $\rho_i(t)$ is the labour productivity of industry i. The labour productivity is also modeled as a stock variable, which changes over time. The productivity growth is based on the neoclassical assumption that productivity growth in the long-run is driven by technological progress (Solow, 1956, 1957; Stiroh, 2001). For this thesis the labour productivity will be measured in euro per labour hour. Throughout the simulation labour productivity may change due to a flow variable. The growth of this variable is dependent on both the development of GDP and the impact technological advancements has on the labour productivity. Since GDP growth and productivity growth are closely related concepts (see Korkmaz and Korkmaz (2017)) it is assumed that there exists a linear relation between the two. The relationship builds on the fact that investments (in capital and knowledge) are a part of GDP, and is assumed to grow at the same rate as GDP, hence the investments share of GDP remains constant. However, how strong the elasticity of productivity growth is on GDP/investments growth is considered to be a parametric uncertainty. The second influence on labour productivity is related to technological development. A specific constant, called the 'labour productivity technology coefficient' will be influencing how much labour productivity will change under technological advancements. This is also considered to be a parametric uncertainty. The variable can best be compared to the labour input coefficient in a Leontief production function which determines how much labour is needed to reach a specific output if capital remains constant. Therefore, if technological development increases, it is expected that industries become more productive and thus requiring less labour hours. This flow variable is defined as,

$$\Delta\rho_i(t) = \rho_i(t) \cdot r(t) \cdot \eta \cdot TD \cdot LPTC, \tag{15}$$

Where $r(t)$ is the growth of investments which is assumed to be the same as GDP growth, η is the elasticity of productivity growth on the $r(t)$ variable, TD is a variable depicting technological development between 0 and 1, and $LPTC$ is the labour productivity technology coefficient. The $LPTC$ will generally take on positive values such that technological advancement will have a positive impact on the labour productivity. However, in chapter 3.8 it will be shown that it might also be of interest of letting technological development have a negative effect on labour productivity and thus resulting in more labour hours demanded. Note that in standard economic theory the productivity parameter in a Leontief production function can never be negative due to the definition of this production function. Therefore this parameter is not the same as the productivity parameter of a production function. For this thesis it suffices to work with a parameter like this since it is able to model what would happen if labour productivity would decrease, resulting in more human labour demand. This eventually leads to the calculation of the labour productivity as:

$$\rho_i(t) = \rho_i(t_0) + \int_{t_0}^{t} [\Delta\rho_i(t)]\, dt. \tag{16}$$

Now that the labour demand for each industry can be described it is useful to better understand how the supply of workers get determined. Due to differences in labour participation and average FTE per person there were two stocks created that depict the male and female FTE's employed per industry. This resulted in the following equation,

$$E_{i,e,a}(t) = E_{i,e,a}(t_0) + \int_{t_0}^{t} [h_{i,e,a}(t) - f_{i,e,a}(t) - o_{i,e,a}(t)]\, dt, \tag{17}$$

Where $E_{i,e,a}(t)$ is the total FTE's employed, the subscripts i, e, a refer to a specific industry i, educational level e and age cohort a, the flow $h_{i,e,a}(t)$ refers to hiring, $f_{i,e,a}(t)$ refers to firing, and $o_{i,e,a}(t)$ is the permanent outflow of FTE's of the labour market due to deaths or retirement. In this stock variable the flow related to hiring deserves some further details.

The hiring flow variable in equation 17 will consider a pool of workers that are available and willing to work. This is the unemployed group of available workers, which is subscripted in the age cohorts and educational attainments. This unemployed group is calculated by taking the difference between the total working population per age and edu level (expressed as FTE's) and the total FTE's that are supplied per age and edu level. As described in Boeri and van Ours (2021) the labour force consists of employed persons and unemployed (but willing to work) persons. In the model, first the potential working population is calculated and subscripted to educational attainment. Here, 'potential' means persons that are aged between 15 and the current formal retirement age, which throughout the simulation increases. Data about the labour participation of males and females is used to obtain the total amount of unemployed persons subscripted for educational attainment and each age cohort as follows

$$U_{e,a} = WP_{e,a} * p_a - \sum_{i=1}^{N} S_{i,e,a}, \tag{18}$$

where WP is the working-age population per educational attainment and age cohort, p is the FTE per worker provided per age cohort, N the number of industries in the model, and S the supply of FTE's in each industry per educational attainment per age cohort.

The model then makes use of an internal function called 'allocate available' which allocates a scarce resources (i.e. the unemployed workers) to specific industries demanding them. The allocation function takes the industry preferences for a certain educational attainment into account to determine how much of each group of workers gets allocated to a specific industry. The model structure for the allocation of workers is inspired by previous work found in Logtens et al. (2012). More information on the allocate available function of Vensim can be found in Appendix A.

The model should also take into account that not all these persons that are available to work are actually qualified to work in this industry. Therefore a variable named 'ratios of persons qualified to work in industry' is created. First a baseline ratio is introduced,

$$\phi_i(t) = \frac{\sum_{e=1}^{M} \sum_{a=0}^{J} E_{i,e,a}}{\sum_{i=1}^{N} \sum_{e=1}^{M} \sum_{a=0}^{J} E_{i,e,a}}, \tag{19}$$

where $\phi_i(t)$ gives the baseline ratio of persons currently working in a specific industry, N, M, and J refer to the total amount of industries, educational levels and age cohorts respectively, and $E_{i,e,a}$ can be found in equation 17. The variable $\phi_i(t)$ can be interpreted as the baseline fraction an industry has of the total supplied FTE, and sums to one after adding all the industry ratios. For example, if $\phi_i(0) = 0.05$ for construction, it means that 5% of the total supplied FTE's are active in construction. However, this does not mean that only 5% of the population is qualified to have a job in the construction industry. This value may be far above the 5% implying that the total sum of all the fractions of persons *qualified* to work in an industry can exceed 1.

The model assumes at the start a baseline according to equation 19. If the unemployment rate is above a certain threshold for a certain amount of time, this baseline ratio will increase due to people retraining which results in an increased ratio of people qualified to work in an industry. This finally results in the variable 'ratios of persons qualified to work in industry' written as

$$\psi_i(t) = \phi_i(t) \cdot m, \tag{20}$$

where m is a multiplier depicting how much the baseline ratio increases as a consequence of retraining after high unemployment rates. The variable is smoothed over time and ranges between 1.05 and 1.3 and is considered to be a parametric uncertainty. Further elaboration on the variable m is given in Appendix B, equation 33.

Finally the model has been adapted to take into account the perceived ROI for working in each industry described in section 3.6.3. To model this a variable called 'impact of ROI on extra hiring' was created with the following formula,

$$\zeta_{i,e}(t) = \beta \cdot \Delta ROI_{i,e}, \tag{21}$$

where $\zeta_{i,e}(t)$ is the impact of ROI on extra hiring, β is the elasticity of hiring on expected earnings and is considered a parametric uncertainty, and $\Delta ROI_{i,e}$ is the percentage change in perceived ROI per industry. For example, if $\zeta_{i,e}(t) = 1.05$ it means that hiring in a certain industry for a specific educational attainment will be increased with 5% if there is available supply.

All these concepts tie in together in the hiring flow variable from equation 17. This gives,

$$h_{i,e,a}(t) = \frac{A_{i,e,a}(t) \cdot \psi_i(t) \cdot (1 + \zeta_{i,e}(t))}{\textit{hiring time}}, \tag{22}$$

where $A_{i,e,a}(t)$ is the allocation of workers from the allocate available function, and *hiring time* is the average time it takes for workers to get allocated to a job.

With the labour demand and supply the relative FTE scarcity can be calculated. As defined by EurWORK (2022), labour shortage/scarcity 'arises when the demand for workers in an occupation exceeds the supply of workers available who possess the required skills and are willing to work at a specific wage rate and in specific working conditions in a particular place and point in time'. For this thesis this quite elaborate definition has been simplified. The specifics related to wage rates and working conditions are not taken into account. The supply of worker is however distributed into specific industries and educational attainment. This yields,

$$\textit{relative FTE scarcity} = \frac{D_i - \sum_{e=1}^{M} \sum_{a=0}^{J} E_{i,e,a}}{\sum_{e=1}^{M} \sum_{a=0}^{J} E_{i,e,a}}, \tag{23}$$

An industry will fire people if demand exceeds supply and thus when the relative FTE scarcity becomes negative. See equation 34 in Appendix B.

In the labour market there exists a general minimum wage (as set by the government). This is initially set at €1934 per month (Rijksoverheid, 2021). The average hourly wage within each industry varies and hence this variable is subscripted per industry. These variables are used to determine the ROI for working in a specific industry from the perspective of someone following tertiary education. In the real world unions will participate in wage setting by putting pressure on industries. This process is empirically difficult to measure and could be considered a study in itself (see for instance Binmore et al. (1986) and Osborne and Rubinstein (1990)). Therefore the wage setting is simplified as an exogenous process that increases at a certain rate. This rate is a parametric uncertainty in the model which will be handled in Section 3.8.

Finally the labour market produces several outcomes of interest which will be covered chapter 3.8.3. Two of them are particularly important and are hence given in an equation. The first one is the unemployment rate,

$$unemployment\ rate = \frac{\sum_{e=1}^{M}\sum_{a=0}^{J} U_{e,a}}{\sum_{i=1}^{N}\sum_{e=1}^{M}\sum_{a=0}^{J} E_{i,e,a}}, \tag{24}$$

where the variables are defined in equations 17 and 18. The second one is the net FTE scarcity. For this the total average FTE scarcity is calculated in the Netherlands as a whole. This is just taking a simple average of equation 23 over all the 20 industries. This yields,

$$net\ FTE\ scarcity = average\ FTE\ scarcity - unemployment\ rate. \tag{25}$$

The model also gives the FTE scarcity for low, mid and high educational attainment, by taking the average of the relative FTE scarcities of the industries that prefer that specific educational attainment.

3.6.5 Economy

The economy submodel is the simplest submodel in the system. After several modelling iterations it was decided that GDP growth could best be modeled exogenously. It is assumed that the 20 industries (see appendix A table 5 for details) make up the whole economy. This assumption is deemed reasonable since exactly those 20 industries are used across several databases related to more granular economic output data (CBS, 2022b; Kruiskamp, 2022; ROA, 2020). The GDP will grow at a specific rate and the industries will grow proportionally at the same rate. It is therefore assumed that prices will stay constant and not react to other variables in the system. This assumption was being made due to the scope and time frame of the thesis, and it is expected that model output can still be interpreted with this assumption in mind.

3.7 Model validation

This section aims to asses the model validity for answering the main research question. In other words, is the model fit-for-purpose? What is important to understand is that the concept of validity in this context is a relative subject. This means that validity can only be assessed relative to the purpose of the model in mind (Senge & Forrester, 1980). Due to the structure of EMA the model is already subject to several validity tests Auping (2018). First, the purpose of this research and the model is discussed. This will guide the reader in understanding when the model is valid for this thesis. Then, the structural validity of the model will be discussed. Finally, the behavioural validity of the model will be shed light upon at the end of this section.

3.7.1 Purpose of the model

The goal of this thesis is to understand long-term labour market dynamics in the Netherlands. Long-term for this context means a time horizon of several decades. It is important to understand that this research is not conducted in the fashion of economic or econometric modelling or forecasting. If that were the case, the time horizon would most likely be far shorter and the use of probability density functions and confidence intervals would become more prevalent. This research is meant to ask 'what-if' questions related to a large dynamic real world system (Oreskes et al., 1994). Hence, it is not the purpose of the model to make highly confident forecasts of the unemployment rate two, three or five years from now. It is rather meant for exploring behaviour of the system under an ensemble of deeply uncertain futures, and how the system behaves under different policy measures.

To test if the model is valid for this specific purpose the structural validity as well as the behavioural validity are considered in the next subsections.

3.7.2 Structural validity

Senge and Forrester (1980) propose several tests to test the validity of a system dynamics model. This paper will be used as guidance throughout this subsection. One of these tests is called the 'structure-verification test' (SVT). This test is meant to verify if the structure of the model coincides with what the real world looks like. With the conducted literature study, described in the background chapter 2, the submodels are created to be close to reality. However, certain concepts are assumed to be of little relevance when the submodels were created (e.g. PhD students are not considered). While making an univariate assumption might have little effect, the effect of having multiple smaller assumptions all at the same time might result in implausible behaviour. However, the structure of the model has been discussed multiple times with supervisors of this research. Validating the structure with them aided with the SVT and resulted in the 'boundary-adequacy structure test' (BAST) (Senge & Forrester, 1980).

The BAST is needed to check if important structures are not left out by accident or even by design. In the basis BAST will be the process of developing hypotheses that certain model structure needs to be included to address certain model issues. In most cases, BAST is conducted with help of SVT's. For instance, for this research the extensiveness of the economy submodel was put to question. This led to the hypothesis that the inner workings of the economy (e.g. imports, exports & domestic consumption) will play a part in understanding long-term labour market dynamics in the Netherlands. For this a simple economy submodel was developed, and it was concluded that only the development of GDP was needed to be able to answer the research question. Aspects as imports and exports fell outside the scope of the research as they were deemed for little relevance for the purpose of the model. It is therefore important to perform a BAST to understand how far the model boundaries should reach to still be able to answer the research question, while keeping in mind that the model will not grow indefinitely.

Another test that is applied is called the 'parametric-verification test' (PVT). Within the model the majority of parameters is modeled endogenous and depends on values of other parameters. However, the model needs certain values as initial inputs as well as values for the exogenous modeled parameters. As explained in subsection 3.5, Data, the initial values are taken from trusted databases and sources and are considered to be a close representation of the real world. The exogenous parameters are determined through existing literature and empirical estimates. Where research is non-existent or a parameter is difficult to measure, assumptions have been made. These are further discussed in section 3.8.1 about model uncertainties.

The described PVT test can be extended with an 'extreme-conditions test' (ECT). This test is especially designed to test model behaviour in the case of extreme values of particular uncertainties. The ECT is part of EMA, where the model is ran several thousands of times under a wide range of changing values for the parametric uncertainties. This test will be part of the results chapter, chapter 4.

Finally, the 'dimensional consistency test' (DCT) for structural validity was also included. Within the software used for modelling the problem there already exist a way to perform this test. Since it is customary to define the units of each variable, the software is able to detect any inconsistencies when units in equations do not match up. This made it possible to make sure that a flow variable is always taken yearly and a stock variable is just a value changing over time. The DCT has helped to find any ommited variables which may have forgotten in the design of the model.

3.7.3 Behavioural validity

To test the model's behavioural validity several runs are performed within the Vensim modelling software. By checking if certain variables of interest behave as one would expect them to behave in the real world the validity of the model can be enhanced. In figure 3 the development of educational attainment of 46 through 55 year olds is depicted for all seven educational levels. What can be seen is that the fraction of higher educated people (hbo and wo) will increase throughout the simulation. Also, in these age groups it is more common to observe people that have not continued studying after high school education or even primary education. This shows that the model is capable of capturing the changing educational attainment in the Netherlands in which more people will start a higher educational level. The picture shows the delayed effect of having a different educational attainment distribution in the younger cohorts, which will result in a changin distribution in the older cohorts over time.

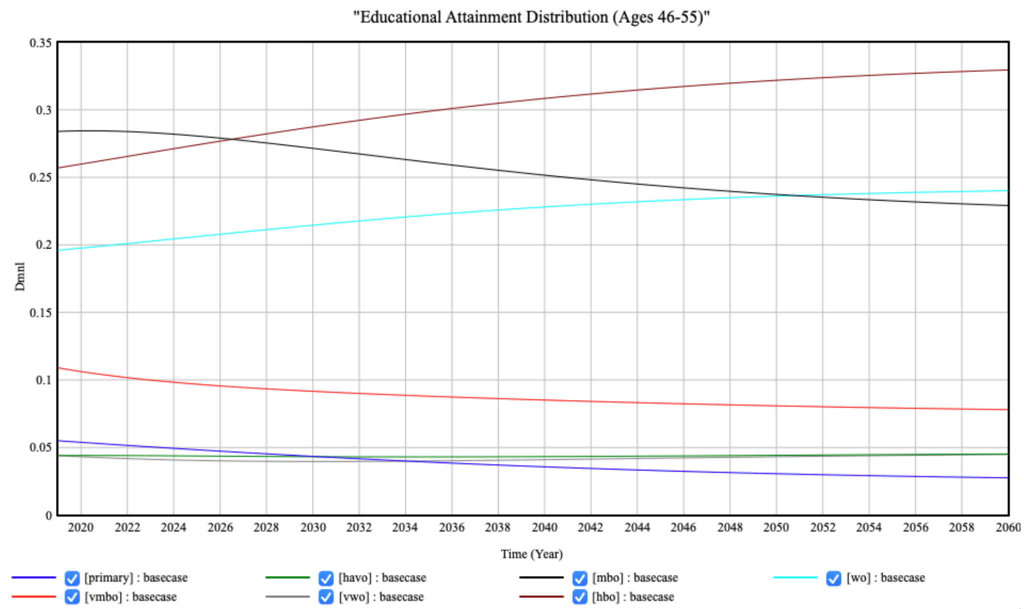


Figure 3 Educational attainment 46-55 year old

Then, in figure 4 the forecasts of the CBS are given CBS (2023b). The dotted blue line is the forecast of CBS, the dark green shade covers the 67% confidence interval and the light green covers the 95% confidence interval. The dark orange line is what the model outputs for the total population. The output of the model falls within the 95% confidence interval of the CBS forecasts. The population output is sensitive to the fertility rate, which is a constant variable in the model. In the real world this may vary overtime which results in different demographics. In any case, the model can be calibrated by setting a higher fertility rate to remain coherent with other forecasting models if these change over time.

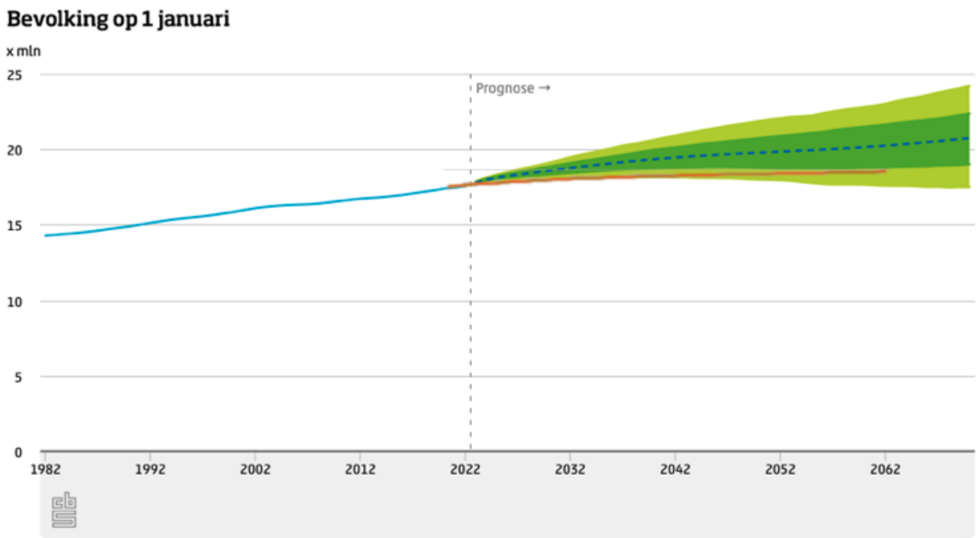


Figure 4 Population forecast and output (CBS, 2023b)

3.8 Experimental setup

The experimental setup commences with the Robust Decision Making approach (RDM) (Lempert et al., 2006). RDM can be considered as a quantitative multi-scenario simulation approach, which aids modelers in dealing with deep uncertainty. Recent literature suggest the following steps within the RDM framework (Auping et al., 2015; Hall et al., 2012):

1. Scope the problem with help of the XLRM framework
2. Run the model with a candidate policy and identify in which scenarios this policy performs poorly
3. Redesign or reidentify policies and iterate through step 1 and 2 again

To run the model, Vensim DSS Version 9.4.2 software by Ventana Systems is used. Then Python version 3.11.4 64-bits is used in combination with the EMA workbench version 2.4. Furthermore for the analysis it is important to downgrade the pandas library to a version of the form 1.x.x. Since a certain functionality used by the EMA workbench is not available in pandas version 2.x.x. This section will continue as follows. First in section 3.8.1 the model uncertainties are discussed. Then, in section 3.8.2 the several assumed scenarios are discussed. Then, in section 3.8.3 the outcomes of interest are elaborated upon. Finally the model's policy levers are described in section 3.8.4.

3.8.1 Model uncertainties

Within the model structure certain variables are modeled exogenous and are difficult to base on data. Hence these variables can be considered as parametric uncertainties in the model. Within this set of uncertainties there reside three scenario variables related to immigration, GDP growth and technological development. The uncertainties are reported in table 1. Each uncertainty is parametric of nature and will be used to sample over in the EMA. The equation in which the uncertainty plays a role can be seen in the 6th column.

Table 1 Model uncertainties

Name	Unit	Min	Max	Type	Equation	Source
Annual change in preference of workers	Dmnl	0.0	0.1	Real	29	Assumption
Annual change in FTE per worker	Dmnl	0	0.1	Real	31	Assumption
Labour productivity coefficient	Dmnl	-0.5	1.5	Real	15	Assumption
Elasticity of productivity to investment growth	Dmnl	0.4	0.57	Real	15	Verdoorn (2002)
Threshold number of years long unemployment	Years	0.5	3	Real	33	Assumption
Increase of skill multiplier	Dmnl	1.05	1.3	Real	20	Assumption
Hiring time	Years	1	5	Real	22	Blázquez Cuesta (2005)
Firing time	Years	2	15	Real	34	CBS (2018)
Time per edu level[mbo]	Years	4	6	Real	9	Assumption
Time per edu level[hbo]	Years	4	7	Real	9	Assumption
Time per edu level[wo]	Years	4.5	8	Real	9	Assumption
Fraction that prematurely leaves edu level[mbo]	Dmnl	0	0.1	Real	10	CBS (2020a)
Fraction that prematurely leaves edu level[hbo]	Dmnl	0	0.1	Real	10	CBS (2020f)
Fraction that prematurely leaves edu level[wo]	Dmnl	0	0.1	Real	10	CBS (2020f)
Exogenous fractional increase in average wage	Dmnl	0	0.1	Real	13	OECD (2019)
SWITCH GDP growth scenario	Dmnl	0	S_{econ}	Categorical	N/A	N/A
SWITCH immigration scenario	Dmnl	0	S_{imm}	Categorical	N/A	N/A
SWITCH technological development scenario	Dmnl	0	S_{tech}	Categorical	N/A	N/A

Note: Here the value S_x, where the subscript x refers to a specific variable, is the total amount of scenarios existing for the corresponding uncertainty.

3.8.2 Scenario's

Since the model will be operating in an environment subject to deep uncertainty several scenarios are developed. The scenarios range from rather possible, all the way to more extreme scenarios to test model behaviour under more extreme conditions.
The immigration scenario consists of three possible future immigration flows which entail the following:

- Immigration basecase: Immigration does not change throughout the time horizon
- Immigration scenario (1): An overall 5% increase of immigration across all age groups

- Immigration scenario (2): A 5% increase of immigration of ages 21 through 45. This scenario will only come in play if there is a labour shortage, and hence more need for external workers
- Immigration scenario (3): A 10% increase of immigration of ages 21 through 45. This scenario will only come in play if there is a labour shortage, and hence more need for external workers

Note that while the aforementioned points are labeled as scenarios, they can be classified as a policy to a certain extent. For instance, the government can design a policy which aims to attract more foreign workers to the Netherlands in the case of a labour shortage. This specific policy design however, lies outside the scope of this research.

The technological development scenario will have as output a number between 0 and 1. Where 0 equals current technological development, and 1 will imply that technological development has reached its theoretical maximum level. In each scenario this value will either slowly or quickly increase from 0 to a higher value. Then the value is used in equation 15. The scenarios are structured as followed:

- Technological development basecase: There is no increase in technological development compared to the start of the simulation.
- Technological development scenario (1): Slowly increasing technological development to maximum levels. In 2050 the variable equals 0.5 and in 2060 it equals 1.
- Technological development scenario (2): Slowly increasing technological development to half of maximum levels. In 2050 the variable equals 0.25 and in 2060 it equals 0.5.
- Technological development scenario (3): Rapid technological development to maximum levels. In 2033 the variable equals 0.5 and in 2040 it equals 1.

A visual representation of the variable is given in figure 5.

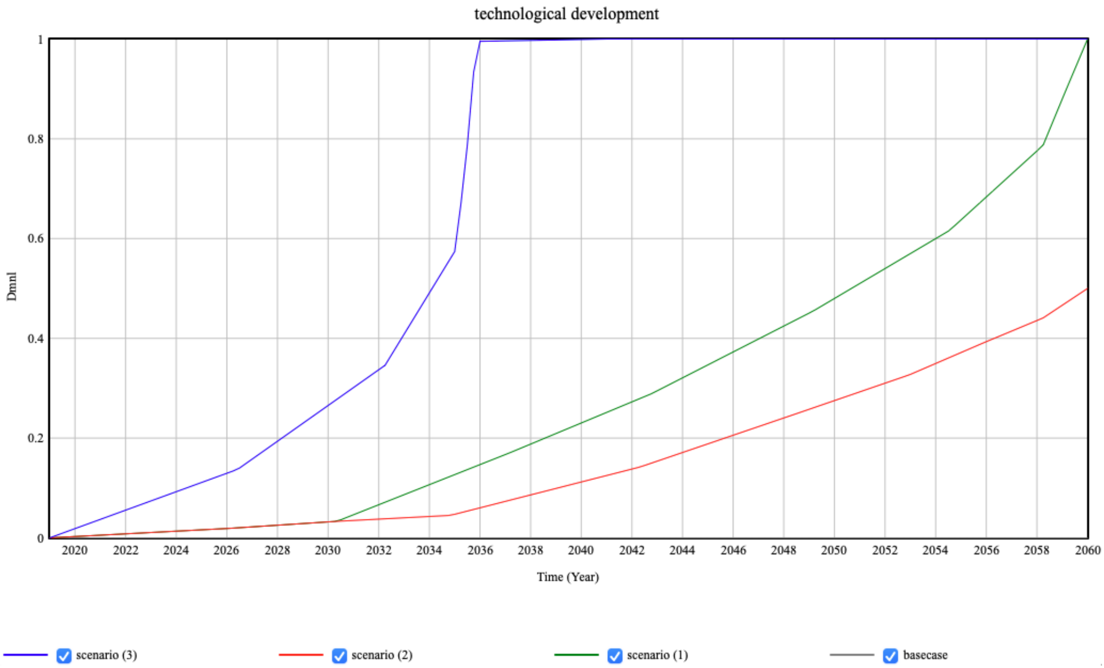


Figure 5 Technological Development Scenarios

Finally, the development of GDP in the Netherlands may be subject to three different scenarios outside the base-case. Based on forecasts of the GDP from PWC (2017), it is assumed that GDP will increase steadily with 2% per year in the basecase. However, the economy may be subject to business cycles and recessions. Based on econometric data (Filardo & Gordon, 1998) and choosing values such that more extreme conditions can be explored the following scenarios were developed:

- GDP growth basecase: GDP grows with 2% per year.
- GDP growth scenario (1): Baseline GDP growth with three short 1-year recessions (-4%) in 2020, 2035 and 2050.
- GDP growth scenario (2): Baseline GDP growth with three long 4-year recessions with lows of (-4%) in 2020, 2035 and 2050.

The effect of the different growth scenarios on GDP can be seen in figure 6.

3.8.3 Outcomes of interest

After having identified key uncertainties, reported in table 1, different scenarios can be generated with help of Latin Hypercube sampling, which is available in the EMA workbench tool. The sampled scenarios can be ran as experiments which will yield outcomes of interest. Thereafter, with help of rule-induction algorithms like Patient Rule Induction Algorithm (PRIM) (Friedman & Fisher, 1999), it is possible to identify experiments in which outcomes yield unfavorable or unacceptable values. To be able to asses unfavourable model output, it is necessary to identify output variables of interest within the system. These are depicted in table 2.

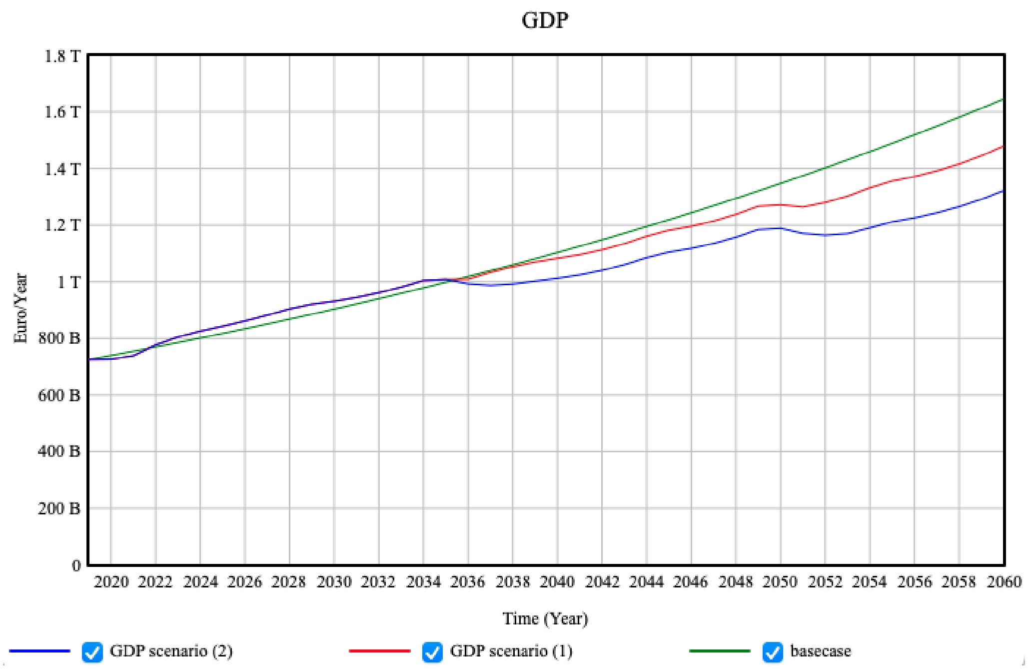


Figure 6 GDP Grwoth Scenarios

3.8.4 Policy levers

In this section the possible policy levers will be discussed. In the model, these are the variables which actually, at least partially, can be influenced by policy measures. They are depicted in table 3. The policy 'Increase FTE per worker' can be switched on or off in the model and an extensive explanation of the variable can be found in Appendix B equations 30 through 32. If it is switched on the policy will target all ages above the variable 'Starting age for targeting FTE per worker increase' and will gradually increase the average FTE's per worker for those age groups. This policy will start having an effect after the year that is specified by the variable 'starting year increase FTE per worker policy'.
The policy lever 'secondary school level flow' is a categorical variable in the model and can best be compared to a scenario variable. The policy has as goal to increase the educational attainment of students starting high-school. This will impact the final educational attainment that exists in the Netherlands. Since the EU has set certain goals for member states to have at least a certain amount of highly educated people, it is assumed that this variable is (in part) a policy. The variable can take four different values resulting in four different outputs:

- Secondary school flow basecase: No changes in the distribution of vmbo/havo/vwo student inflows
- Secondary school flow policy result (1): Vmbo will decrease with 1 percentage point (pp) per 10 years and havo will increase more sharply (0.6 pp) than vwo (0.4pp) per 10 years
- Secondary school flow policy result (2): Vmbo will decrease with 2 pp per 10 years and havo will increase more sharply (1.2pp) than vwo (0.8pp) per 10 years

- Secondary school flow policy result (3): Vmbo will decrease with 2 pp per 10 years and havo and vwo will increase at the same rate (1pp) per 10 years

Table 2 Model outcomes of interest

Name	**Explanation**
Total average FTE scarcity	Depicts the relative scarcity in FTE's averaged out over all the 20 industries
FTE scarcity for low edu	Depicts the relative scarcity in FTE's in the industries that prefer lower educated workers
FTE scarcity for mid edu	Depicts the relative scarcity in FTE's in the industries that prefer middle educated workers
FTE scarcity for high edu	Depicts the relative scarcity in FTE's in the industries that prefer higher educated workers
Unemployment rate	The ratio between total persons unemployed (and willing to work) and the total persons employed
Net FTE scarcity	Total average FTE scarcity minus the unemployment rate
Total FTE's demanded	Sum of all FTE's demanded across industries
Total persons employed	-
Total persons unemployed	-
Current percentage mbo/hbo/wo attainment	Gives the current percentage in the educational system of a certain educational level

Table 3 Policy levers

Name	**Explanation**
Increase FTE per worker	This variable can be switched on or off in the model. It will steadily increase the FTE per worker of persons above the age of a certain starting age.
Starting age for targeting FTE per worker increase	This variable is used for the Increase FTE per worker policy. Ranges between 40 and 55.
Starting year increase FTE per worker policy	This variable is used for the Increase FTe per worker policy. Gives after how many years the policy will start having an effect. Ranges between 1 and 20 years.
Secondary school level flow	A semi-policy lever that determines how quickly the general eductional level of high-school students shift from vmbo to higher levels.
Percentage change inflow hbo/wo students	This lever can be operated to reduce the influx of foreign hbo/wo students to 80% or increase it up to 120% per year
Policy increase retirement age	Determines how much the formal retirement age will increase per one year of increased life expectancy. Variable ranges between 0 and 0.75 years.
Average percentage yearly increase in industry wages	Semi-policy lever that gives the *average* yearly increase in industry wages during the whole time horizon.

The defined model uncertainties and model outcomes can be summarised in an XLRM-framework. Together with the conceptual model and the policy levers this results in the XLRM diagram depicted in figure 7.
To summarise important information needed for EMA, the use of an XLRM-framework is customary (Lempert, 2003). Where X are exogenous uncertainties, L refers to policy levers, R depicts the relationships to describe the system and M is used to define performance metrics. For this research the relationships in the system will be modeled with an SD model in which the other three parts of the XLRM framework reside.

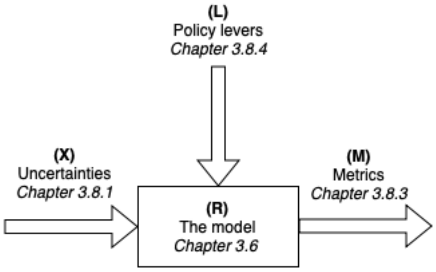


Figure 7 XLRM framework

3.9 Performing Experiments

Now that the three types of variables and their possible values have been designed, the model can be ran over an ensemble of these values. The model will be ran a 1000 times for 15 different policies. Here, a policy refers to a certain set of values the policy levers will take. The EMA workbench uses full factorial sampling over the policy values to create a diverse set of different policies. The 1000 scenarios refer to 1000 uncertain futures of the system, and hence will take 1000 different sets of values for all the uncertainties. The workbench exploits Latin Hypercube Sampling (LHS) for this purpose. This will eventually result in 15000 experiments, which may have in some cases the same outcome. Depending on the setup the model takes 32 hours to run on a 2018 MacBook Pro with four Intel i5 processors. To save time, a more powerfull desktop has been used on which the model ran roughly 8 hours. Eventually the model has generated 106 time steps per outcome, resulting in a dataframe of 15000 by 106 per outcome.

3.10 Scenario discovery

The main analysis that will be performed is scenario discovery. Scenario discovery was first written about by Bryant and Lempert (2010) and has as purpose to aid policy makers in the public domain to develop robust strategies by analysing futures considered most important. To do so rule induction algorithms like PRIM & Classification and Regression Trees (CART) can help identify in which futures unwanted model behaviour exists. Based on the output of these analysis it is possible to further develop or recommend policy strategies that will negate undesirable outcomes.

4 Results

In this section the results of several analyses will be discussed. First in section 4.1, the key drivers in each outcome's behaviour are discussed. Then section 4.2 reveals outcome behaviour of two important variables under different scenarios. In section 4.3 the temporal behaviour of important metrics is discussed. This chapter concludes with sections 4.4 where a rule-induction algorithm is applied to identify statistically significant uncertainties and policy levers that determine outcomes of interest. The GitHub repository containing the model and analysis jupyter notebooks can be found here or via this url https://github.com/lvanopstal/MasterThesisEPA.

4.1 Identifying key drivers in the system

To obtain some insights in what are the most important drivers in the system, a feature scoring algorithm has been applied. To do so, the algorithm will only take the last values of each outcome dataframe, meaning at the end of the simulation. It then determines for each outcome which variable drives most of the variance within the outcome. The feature scoring is visualised in a heatmap in figure 8.

From the figure it can be deduced that the unemployment rate will be mainly driven by two variables. Namely, the immigration scenarios and the hiring time (i.e. how long does it take to find a job). Furthermore note that the variance in the net FTE scarcity can be explained largely by changes in the GDP scenarios, elasticity of productivity to investment growth, the technological development scenarios, and the labour productivity coefficient. Also note that the net FTE scarcity can be subdivided in scarcity of workers of three different educational levels. Where 'low' refers to primary through mbo, 'mid' refers to hbo, and 'high' refers to wo. The outcomes are depicted in columns six through eight.

Finally, the variance in the six education variables (last six columns) is mostly explained by the expected uncertainties. The fractions of persons that prematurely leave the educational system will explain around 78% of the variance in the current educational attainment within the school system. This will hence determine the distribution of educational attainment that exists in the labour market. Also, the educational attainment is deemed quite insensitive to the change in average wages in each industry (i.e. 'exogenous fractional increase in average wage').

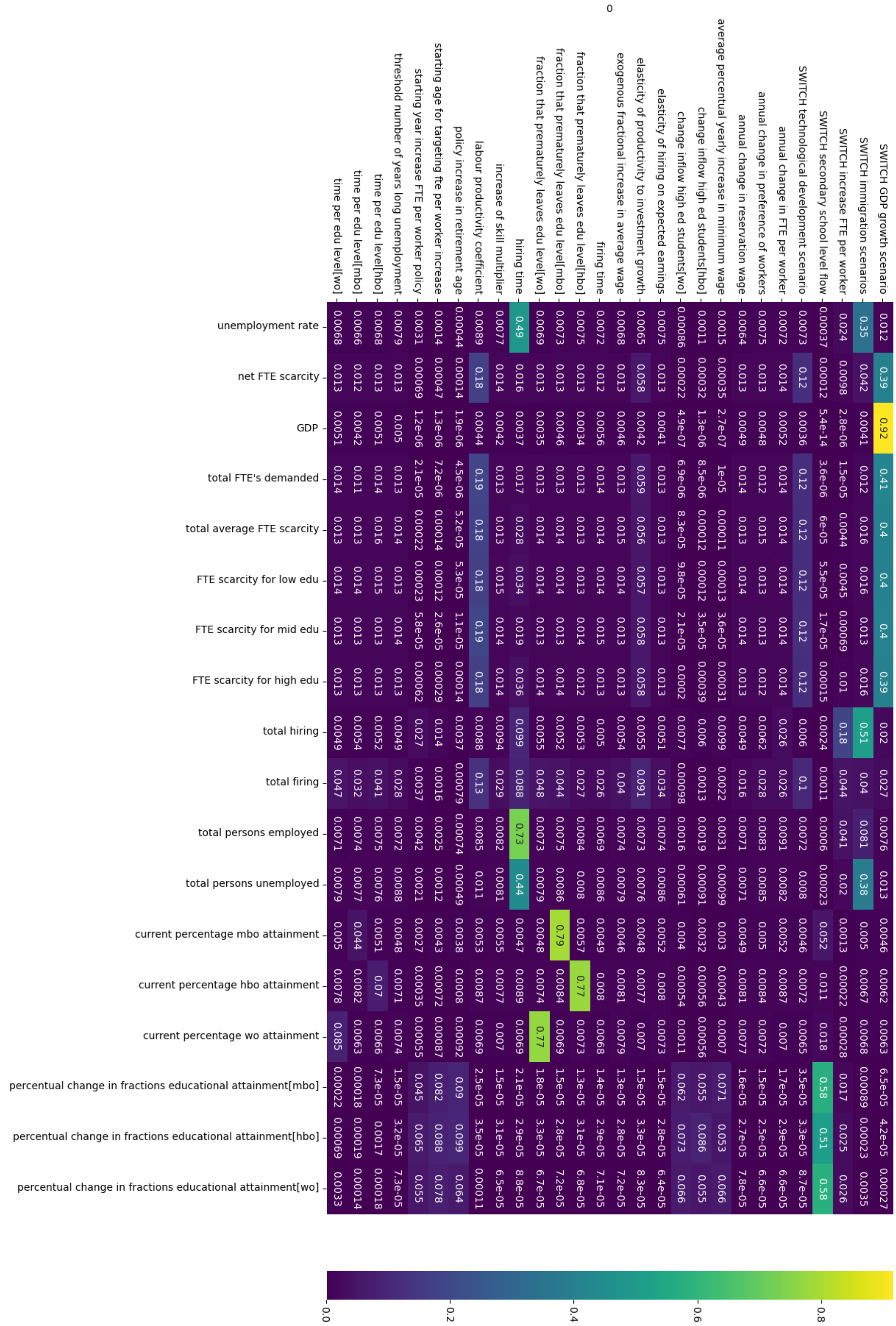


Figure 8 Feature scoring

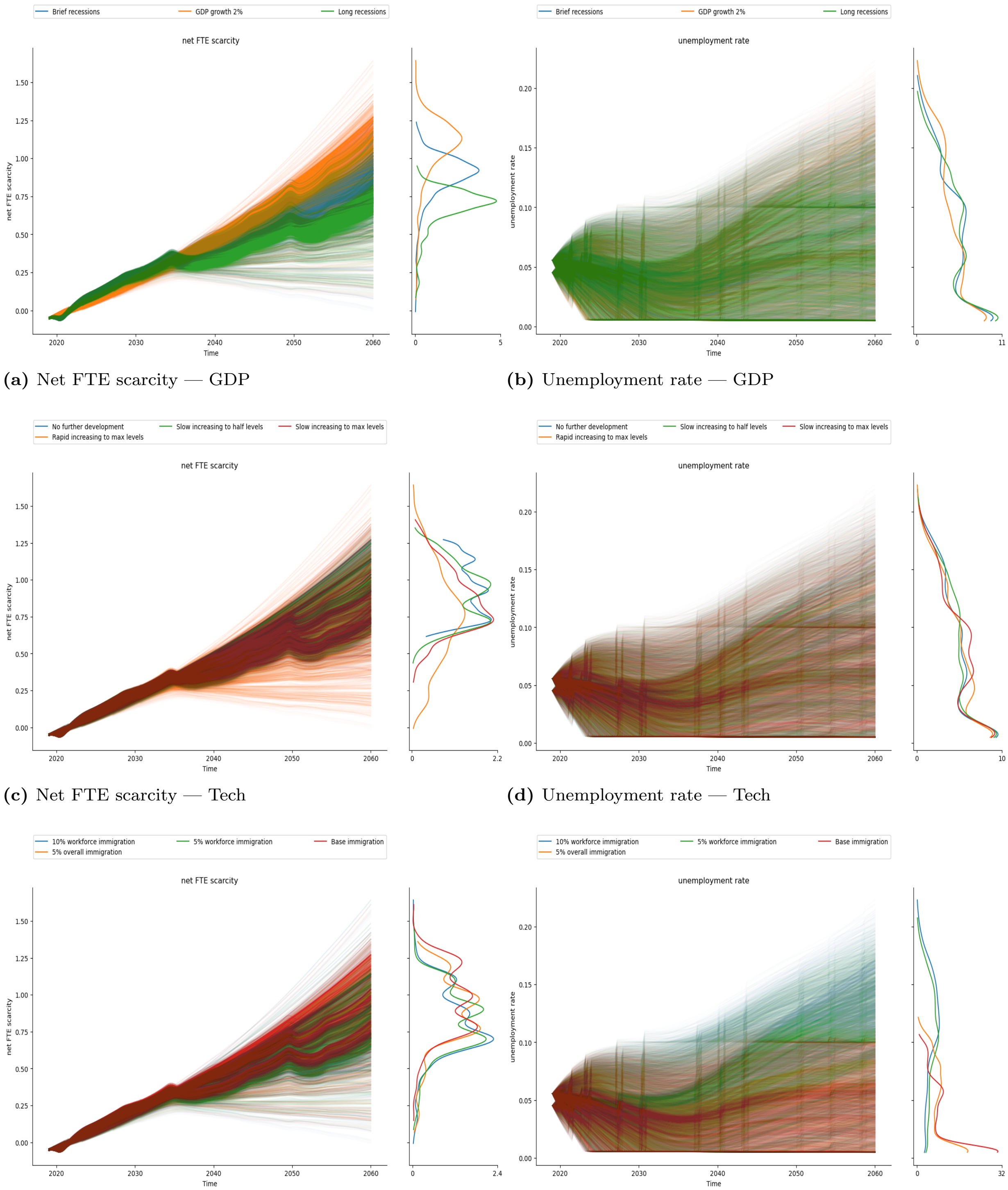
4.2 Net FTE scarcity and unemployment rate

For further analysis the focus shifts to the two variables: net FTE scarcity and the unemployment rate. The reason for this is the fact that the variable 'total FTE's demanded' is behaving very similar to the net FTE scarcity, resulting in interchangable conclusions. The FTE scarcities that are split out between the different educational attainments are also portraying similar behaviour and could hence be summarised in the variable 'net FTE scarcity'.

To better understand the behaviour of two key outcomes, the time series are grouped by several uncertain futures in figure 9. The three different GDP scenarios have different effects on the net FTE scarcity as can be seen in figure 9a. The orange line represents the base case GDP scenario in which GDP will grow on average 2% per year. As expected, this scenario will create the highest scarcities due to the increase in FTE demand as a result of GDP growth. Then, The other two GDP scenarios take into account that GDP may grow at a lesser rate due to recessions, which results in a negative GDP growth for several years. As a result the net FTE scarcity will be lower in those cases due to the lower demand for FTE's. The unemployment rate in figure 9b is not driven by the different GDP scenarios, as is also suggested by the heatmap in section 4.1. This is due to the fact that in all simulations there exists the need for workers and in theory most people could start a job. The increases in the unemployment rate is hence driven by other factors which will be discussed.

The four scenarios related to technological development portray similar behaviour for the net FTE scarcity as shown in figure 9c. What can be noted is that the density of the outcomes is more spread out across the whole range for the orange and red lines, where the technological development will eventually reach its theoretical maximum level. For the base case, in which no further technological development will occur, most of the simulations end up with a scarcity centered around 0.8. Then, the unemployment rate in figure 9d behaves in similar ways for the four different scenarios. Just as before, the unemployment rate seems not necessarily to be driven by the technological development because in all simulations there exists a scarcity for workers in each industry.

Finally the four different immigration scenarios were considered for both variables in figures 9e and 9f. In this case the net FTE scarcity behaves similar for the different scenarios. The unemployment rate however shows different behaviour for the different scenarios. The model suggests that unemployment rates will rise when there is an increase in immigration of people aged between 21 and 45 years old. This seems counter-intuitive at first sight, but will be elaborated upon in the following sections. Furthermore, the unemployment rate is slightly overvalued due to data limitations in the system. These are discussed in chapter 5.2.



(a) Net FTE scarcity — GDP **(b)** Unemployment rate — GDP

(c) Net FTE scarcity — Tech **(d)** Unemployment rate — Tech

(e) Net FTE scarcity — Immigration **(f)** Unemployment rate — Immigration

Figure 9 Behaviour under uncertain futures

4.3 Temporal behaviour of net FTE scarcity and unemployment rate

The kernel density plots from section 4.2 show only the densities at the end of the simulation. Since we are dealing here with time-series outcomes, it is interesting to investigate the behaviour of outcomes over time. In figure 10 a temporal feature scoring heatmap for the net FTE scarcity outcome is depicted. On the x-axis a specific point of time is depicted within the simulation. For each of these specfic time points the feature scoring looks at the 5 variables that explain most of the variance for that specific outcome. Then the algorithm continues to the next time step and determines the following 5 variables that explain the most variance.

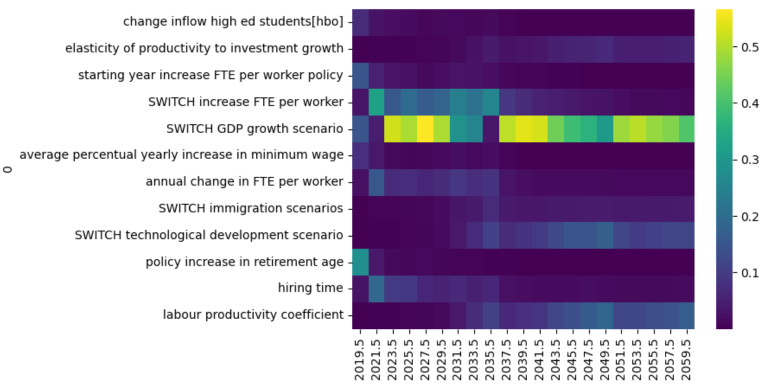


Figure 10 Temporal feature scoring — Net FTE scarcity

From the figure it can be seen that there is a slowly shifting explained variance over time. During the first 15 years the outcome behaviour is mostly driven by the GDP scenarios, the increase in FTE policy and the hiring time. Around 2035, which is chosen to be a year of recession in two GDP scenarios, it can be seen that GDP growth does not really influences the net FTE scarcity much. From that year a shift can be noted to the technological development scenarios and the related labour productivity coefficient. Throughout the remaining years, most variability in outcome behaviour can still be explained by the GDP growth scenarios.

In figure 11 the temporal feature scoring map for the unemployment rate is visualised.

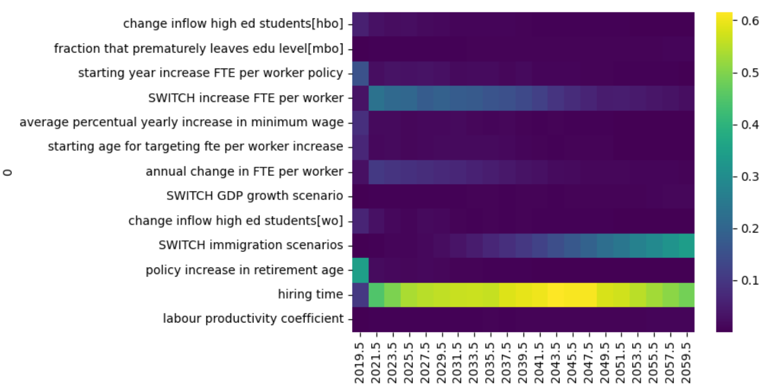


Figure 11 Temporal feature scoring — Unemployment rate

For the unemployment rate no visible switch in drivers can be noted. Rather, the heatmap shows a smooth transition from one variable to another. During the first years of the simulation the increase in FTE policy in combination with the annual change in FTE per worker variable play an important role. After around 15 years the impact of constant immigration is showing its consequences with an increasing impact over the time horizon. Throughout the simulation the hiring time remains the most important uncertainty that explains the variability in the unemployment rates.

4.4 PRIM

Now that general model behaviour over the uncertainty and lever space has been explored, the following step is to perform scenario discovery through PRIM. The algorithm has been applied to the variables 'net FTE scarcity' and 'unemployment rate'. Even though it is implausible that net FTE scarcity will increase in the real world almost 100% of the time, it is still interesting to understand what strategies (i.e. a combination of policies) causes the scarcity to grow at a higher rate compared to other strategies. Hence, we are interested in the cases where the scarcity goes beyond a high level of net FTE scarcity in the 15000 experiments. It uses a threshold of net FTE scarcity higher than 0.9. The boxes that satisfies this threshold are visualised in figure 12.

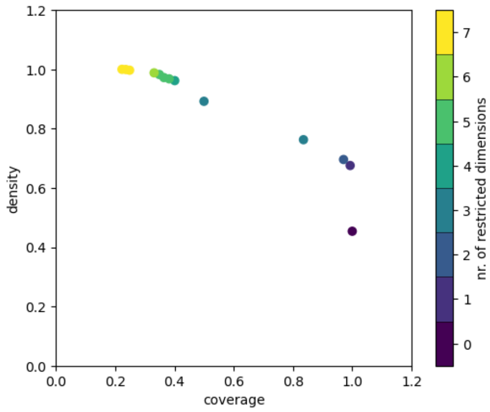


Figure 12 PRIM — net FTE scarcity first trade-offs

With this visualisation it is possible to make a tradeoff between density (i.e. how many of the cases in the box are of interest), coverage (how many of the cases of interest do we cover in the box?) and the interpretability (how many dimensions need to be restricted?). It is customary in EMA to aim for at least a density of roughly 80% or higher. Therefore it was chosen to inspect the 5th box counted from the lowest right point. The results of this box are depicted in figure 13.

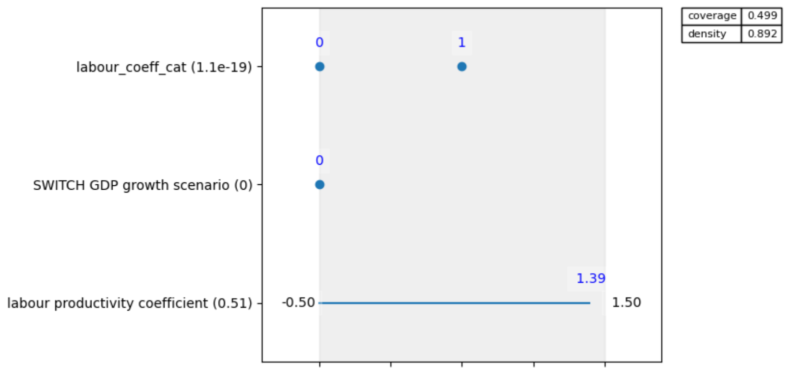


Figure 13 PRIM — net FTE scarcity box 5

In the top right of figure 13 it can be seen that around 50% of cases of interest are covered by the box and that it has a density of 89.2%. The three restrictions are given on the y-axis and the value in the parentheses is a so-called quasi-p-value. Just as normal p-values, it will tell if a restriction is statistical significant. If this value is lower than 0.05 then the restriction is statistically significant at a 95% confidence interval. The grey area gives all the possible values the restrictions can take, and the blue line takes the subrange of these values to end up with the box found by PRIM. Note that two out of three restrictions are statistically significant in causing an net FTE scarcity higher than 0.9. The restriction 'labour productivity coefficient' is dropped from the analysis. The variable called *labour coeff cat* is a categorisation of the labour productivity coefficient variable. Where a negative labour productivity coefficient is categorised with a 0, a coefficient between 0 and 1 is categorised with a 1, and a labour productivity coefficient higher than 1 is labeled as a 2.

The chosen box just covers about 50% of the cases. PRIM was applied to the data again to find a box that could explain the remaining 50%. By applying the algorithm again PRIM was able to find a new rule set that could explain some of the remaining cases. Again, it is customary to look for a box that has a density of at least 80%. The trade-offs for for the remaining cases in which net FTE scarcity was higher than 0.9 is depicted in figure 14.

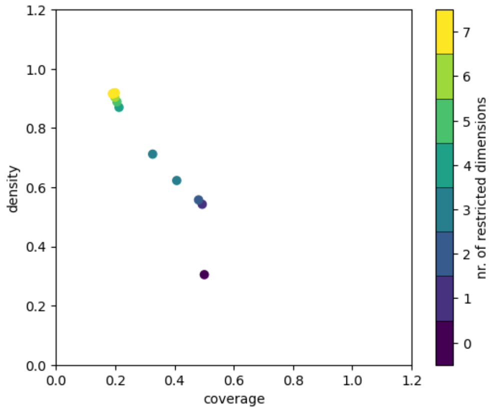


Figure 14 PRIM — net FTE scarcity second trade-offs

Based trade-off figure the the sixth box counted from the bottom right was chosen. This box has density of 87% and coverage of 21.3% of the cases. Meaning that with these two boxes and rule sets around 71% of the cases of interest can be explained, while still maintaining a density of at least 87%. The rule set is given in figure 15.
The figure indicates that the last restriction related to the labour productivity coefficient may be dropped from the analysis due the fact that it is not statistically significant. The joint set of restrictions are related to simulations in which no policy related to an increase in FTE scarcity occurs, technological development does not occur or very slow, and there is constant GDP growth with either no or brief recessions. The full peeling trajectory of the prim algorithm can be found in the Appendix D figures 36 through 39.

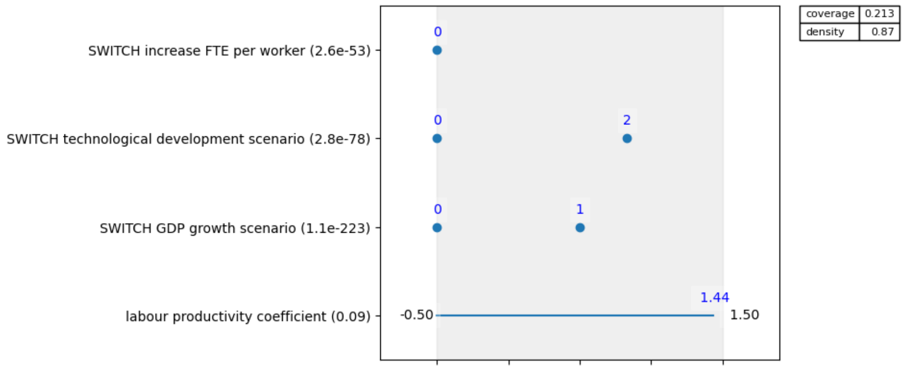


Figure 15 PRIM — net FTE scarcity box 6

A similar analysis has been performed for the 'unemployment rate'. For this a threshold of 0.1 is chosen. The PRIM algorithm will try to find cases for which the unemployment rate will be relatively high compared to the majority of the final values of the outcomes of the unemployment rate. The boxes the algorithm found for this outcome of interest are depicted in figure 16.

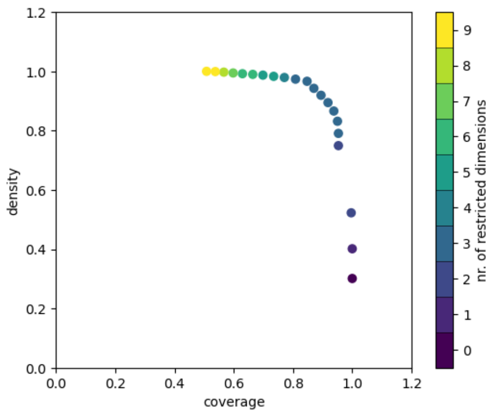


Figure 16 PRIM — Unemployment trade-offs

From the trade-off visualisation the 6th box is chosen. This box has a coverage of 95% and a density of more than 83%. This point can be identified with just three restrictions of which two are significant, as can be seen in figure 17.

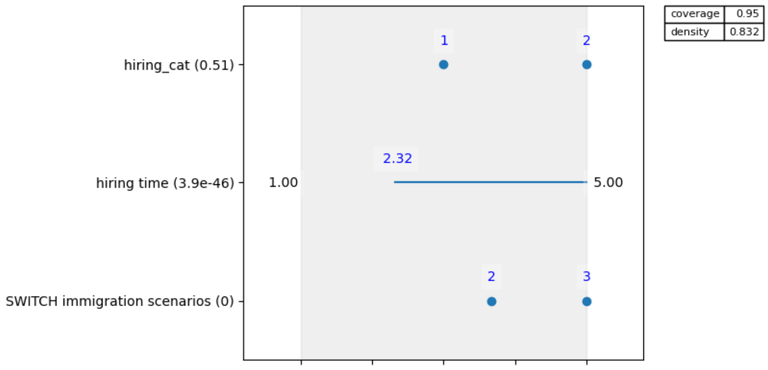


Figure 17 PRIM — Unemployment box 6

Almost 83% of the cases for which the unemployment rate is higher than 10% can be explained by just two uncertainties. The uncertainty called *hiring cat* is not statistically significant and is therefore dropped. As before this was a categorisation of the hiring time variable. Both the joint restrictions of relatively long hiring times and high workforce immigration rates are causing increases in the unemployment rate. PRIM was performed again to find a rule set that might explain the last 5% of boxes, but this was not found. The full peeling trajectory can be found in Appendix D figures 40 and 41.

5 Discussion

In this chapter model output and model assumptions are discussed and reflected upon in section 5.1. This chapter is meant to put the model output into perspective and aid readers in interpreting the results. In sections 5.2 and 5.3 the data limitations, model scope, and time horizon are discussed. The chapter concludes with section 5.4, explaining the scientific relevance of the research.

5.1 Model assumptions and improvements

The model of this research was subject to several assumptions. These assumptions have led to possible limitations and hence suggestions for further research. One of the most prevalent assumptions that has also exposed itself through model outcome is related to the submodel of the economy.

It was shown in chapter 4 that in almost all the simulations the different FTE scarcities grew or remained positive. This result is related to the fact that the economy submodel was modeled exogenously. In all the different GDP scenarios the GDP grew at a base rate of 2% based on forecasts of PWC (2017). In some scenarios GDP decreased due to recessions, but the GDP has increased in 100% of the simulations compared to its starting value. The GAV of each industry grew at the same rate due to the fact that it was assumed that prices remained constant. This assumption has lead that the labour demand in each industry kept growing, and that the increase in population and technological development could not keep up with the growth rate of labour demand. Eventually this has lead to FTE scarcities responding in a similar way across the different industries.

Another aspect to consider regarding labour scarcities is the impact on the development of wages across different industries. In the current modelling approach, wages are assumed to follow an exogenous process that is similar for all industries. However, this assumption overlooks the connection between the scarcity of FTE workers in an industry and the corresponding wage dynamics. In reality, when labour becomes scarce in a particular industry, employers often respond by adjusting wages to attract more workers (Boeri & van Ours, 2021; Osborne & Rubinstein, 1990). To capture this dynamic relationship, it would be beneficial to couple the development of wages with the labour scarcity experienced by each industry. By introducing this coupling, a more realistic scenario can be modeled, where employers actively respond to labour scarcities by increasing wages. This adjustment creates a feedback loop wherein higher wages act as an incentive for workers to migrate or seek employment in industries facing labour shortages. Consequently, the increased labour supply helps alleviate the scarcity and restore a balance in the labour market. Incorporating this wage dynamic into the model enables a more accurate representation of the interplay between labour supply and demand and this is done before but rather in an agent-based setting (Fagiolo et al., 2004). It recognizes the fundamental economic principle that scarcity drives market forces, prompting employers to compete for workers by offering more attractive compensation packages. Such an approach can lead to a more nuanced understanding of how labour scarcities impact different industries and how wages respond to alleviate imbalances within the labour market. Furthermore, wage-setting in each industry may have an effect on workforce immigration in that particular industry (Brücker & Jahn, 2011).

A third limitation within the current model pertains to the study choices made by students. Currently, the assumption is that the distribution of students interested in a particular industry, such as healthcare, remains relatively constant throughout the entire simulation. However, understanding the factors that drive the decision-making process for individual students is a rather complex econometric exercise and could be considered a research topic in itself (Agrey & Lampadan, 2014;

Drewes & Michael, 2006). Although the aggregation of the entire student population assumes a relatively stable distribution, this may not accurately reflect real-world dynamics. The role of average wages within an industry is considered to have a minor influence on changes in educational attainment within the model. However, in reality, the impact of average wages on students' study choices may be more significant, especially males seem to be more sensitive for expected earnings (Rapoport & Thibout, 2018). The decision-making process involved in selecting a specific educational level or industry to work in involves multiple latent factors that interact in complex ways. The model's simplification of measuring these latent factors limits its ability to capture the full complexity of real-world dynamics. To address this limitation, a more nuanced approach could be taken to better understand the underlying drivers of student choice behavior. This may involve incorporating additional variables or factors that influence students' decision-making processes, such as personal interests, career aspirations, socio-economic background, and external influences like media or societal trends. By considering a broader range of influences, the model could provide a more comprehensive analysis of how study choices are made and how they may evolve over time, thereby enhancing its accuracy and applicability in real-world scenarios.

Finally, a critical driver within the system is the influence of technological advancements on the labour productivity. Currently, this dynamic is modeled in a manner that affects the labour market as a whole, without distinguishing between industries or educational levels. However, it is essential to recognize that the impact of technological advancements can vary significantly across industries and educational backgrounds (Webb, 2019). Incorporating a more nuanced approach to capture this variability would be valuable. By considering different scenarios that highlight the varying degrees of susceptibility to technological substitution across industries, a more comprehensive understanding of the labour market dynamics can be achieved. For instance, certain industries, such as construction, may prove to be less easily substitutable by technology compared to industries like business services, which may involve tasks such as financial analysis. Moreover, this differentiation also suggests that the impact of technological advancement on lower-educated individuals might differ from that on higher-educated individuals. While higher-skilled roles may be more resilient to technological substitution, lower-skilled positions could potentially face greater challenges as automation and AI technologies continue to advance. It is important to acknowledge that the long-term impact of technological advancement on the labour market remains uncertain, particularly when considering longer time horizons. Deep uncertainty surrounds the extent and nature of these advancements and their subsequent effects on employment (Frank et al., 2019). Therefore, incorporating further research and scenario development that specifically explores the potential ramifications for industries that are more susceptible to technological substitution can offer valuable insights into the future dynamics of the labour market. By recognizing the differentiated impacts across industries and educational levels, the model can provide a more comprehensive analysis of how technological advancements shape the labour market, enabling policymakers and stakeholders to better anticipate and respond to the evolving challenges and opportunities presented by these advancements.

5.2 Data limitations

In subchapter 3.5 the used databases for this research were briefly mentioned. While these databases are considered highly accurate, there were some cases for which detailed data did not exist and assumptions were made that influenced model behaviour. An example of this is to find the amount of males and females employed in a certain industry for a specific age group. CBS provides this data for all 20 industries for both males and females, but only for 10-year cohorts (e.g. 25 through 35 years old). Therefore the assumption was made that the number of (fe)males working in a industry could be divided by 11 to obtain an estimate for single year cohorts. However, for the 10 year cohort of 15 through 25 years old this posed some problems. The reason for the problems is due to the fact that a major part of this 10-year cohort is still participating in the educational system. For an industry like 'business services' where consultants and financial planners are active, it is assumed unlikely for a 15 year old to have a job in that specific industry. Therefore, these values were set to zero in specific cases to correct for this uncertainty. For the other 11-year cohorts the even distribution of people over all the single year cohorts resulted in the under- and overrepresentation of people actually working for that specific age group. This matrix of 20x106 (i.e. industries by age cohorts) hence contains slightly inaccurate values.

Once the initial values of the amount of employed (fe)males per age cohort have been calculated, the distribution of educational attainment has yet to be added. This meant that two matrices needed to be multiplied with each other to have an estimate of the amount of (fe)males working in a specific industry, with a specific age, with a specific educational level. As a result, certain points in this three dimensional dataframe will be under or overrepresented. This did not have an impact on general model behaviour and outcomes, since these were highly aggregated. The model did need some calibration to correct for the underrepresented groups and end up with more real-world initial values. Yet, the unemployment rate was about one percentage point higher than the real world value due to the under- and overvaluation of a certain amount of workers.

Finally, the data used to determine the GDP growth comes with uncertainty. The forecasting of GDP is a hot topic in econometric literature and most of these studies conclude that even 12 or 24 months ahead forecasting of GDP is a complex task (Stock & Watson, 2011). For this research it is assumed that baseline GDP will grow with 2% per year for over 40 years in total. With the scenarios the model aims to explain what happens in the case this steady growth declines for longer periods of time.

5.3 Model scope and time horizon

In subchapter 3.7 the structural validity of the model was put to discussion. The model scope was deemed appropriate for making the analysis in the long term. Although the model is able to provide more insights in long term labour market dynamics the model scope could be more extensive. The growth of GDP or technological development could be further elaborated upon and modeled more extensively. These two factors in combination with the time horizon of roughly 40 years could be put to question. Due to the long time horizon some dynamics or policies that might occur after certain years could not be taken into account. For example, what kind of policies would come into place if indeed technological development would result in the loss of 20% of jobs? Since the purpose of this study is to *explore* model *behaviour* rather than forecasting outcomes, a short time horizon is not deemed appropriate. However, the choice of 40 years is arbitrary and one could opt to choose for instance 30 years or even 60 years, depending on the context.

5.4 Scientific relevance

This research is meant to add to the literature in the field of labour market economics and policy related fields. The developed model has provided insights and analyses but it also may serve as a tool for policy makers seeking to navigate the complexities of long term labour market dynamics. Additionally, it can provide a solid foundation for further model development. This research aims to add to the literature due to the fact that it describes an highly aggregated labour market system, with dynamics over a long time-horizon, and its behaviour under highly uncertain futures. While in recent years there have appeared papers describing labour market dynamics, most do not cover the highly aggregated uncertain view that has been presented in this thesis. Aiyetan and Das (2018) analyze, through an SD model, the staff shortage of skilled labour workers in the south African construction industry. They were able to establish factors contributing to the shortage of skilled labours (i.e. investments, talent management, wage challenges) but conclude that their findings are difficult to generalize. In a same way both Devadoss and Luckstead (2018) and Leitner (2022) only analyze parts of the complete system of interest of this thesis, where both models have a focus on immigration in a quantitative way. While these papers do not model the whole system, they do provide insights which can be used to model relationships between variables relevant for an aggregate model (e.g. workforce population growth of foreign countries). In a similar manner *Labour Market Intelligence* (LMI) is a growing field of interest in the machine learning literature (Altun et al., 2017) (Dawson et al., 2020) (Giabelli et al., 2020), which takes a rather quantitative approach to model only specific relationships within the whole system.

Furthermore the temporal dimension enhances the model's applicability in capturing and analysing trends, patterns, and policy effects that unfold in highly uncertain futures. By incorporating uncertainties and exploring potential policy responses, the model contributes to the existing literature by offering insights into the potential outcomes and implications of policy decisions in an ever-changing labour market landscape.

6 Conclusion and recommendations

This chapter aims to answer the fourth and final subquestion: 'Which robust policy measures can be imposed to maintain a well-functioning labour market under different scenarios?' This chapter will continue by providing a conclusion based on the results, and then some recommendations are proposed. In section 6.3 a policy design is presented. Finally in section 6.4 the main research research question will be answered and the thesis will be concluded with several suggestions for future work.

6.1 Conclusion

The created model has had as purpose to understand long term labour market dynamics. The open exploration methodology has shed light upon possible sensitivities of the model to uncertain variables and policies. With help of these analyses policy makers can be aided in developing strategies that are robust and will hopefully lead to desirable system behaviour.

The results show that the net FTE scarcity is particularly sensitive to the GDP growth scenario. This makes sense since the growth of the GDP determines the growth of the gross added value of each industry, which in turn will have an increasing effect on the labour demand when controlling for constant labour productivity. While the net FTE scarcity remained sensitive for this variable throughout the whole time horizon, it was shown that scarcities are sensitive to the policy that aims to increase the FTE per worker during the first 15 years of simulation. The outcome becomes more sensitive to technological development and the labour productivity coefficient during the later phases of the simulation. However, this sensitivity does not provide us with a direction the net FTE scarcity will move if for instance an increase in FTE policy is adopted. PRIM has proved to be helpful in answering this question.

PRIM found two different rule sets that were able to explain more than 70% of the cases in which the net FTE scarcity was higher than 0.9. For the first rule set this happened when GDP kept growing at a steady rate, while the labour productivity did not grow at the same rate or was even decreasing in some cases due to a labour productivity coefficient being between -0.5 and 1. For the second rule set scarcities grew to extreme values when no FTE policy was adopted, technological development does only slowly or not occur at all, and GDP grows at a steady rate with only three potential brief recessions. These two rule sets imply two important relationships between GDP growth and technological development. Consider two cases, one in which labour productivity is negatively related or insensitive to technological development, and one in which labour productivity *is* positively related to technological advancement. The first rule set indicates that when GDP grows at a steady rate, and technology advances but industries do not become more productive as technology advances, this will eventually lead to more labour demand. For the second case, GDP keeps growing at a steady rate and technological development stagnates. This in turn will result in a labour productivity that also lags behind, which will eventually also increase the scarcities. Furthermore, if policy makers then also opt to not choose to apply the increase in the FTE per worker policy, net FTE scarcity will grow to extreme levels.

Although these relationships might state the obvious, the results can still be of value. By also considering the variables for which net FTE scarcity is insensitive it is still possible to steer policy makers in the right direction. The fact that inflows in higher educational attainment have little effect on changes in net FTE scarcity, will aid policy makers in shifting their focus to other policy levers. The insensitivity of educational attainment on net FTE scarcity therefore implies that policy makers should not focus on getting people solely a 'higher' degree, but rather on learning students specific skills such that they are not easily substituted by technology. This gives the opportunity for stakeholders to asses if the Netherlands is currently subject to FTE scarcity, and then develop policies to steer the labour market in the desired direction.

As for the unemployment rate in the Netherlands some things can be concluded based on the temporal heatmap and PRIM. It was clear that the majority of the variance in the unemployment rate is explained by changes in the hiring time and immigration scenarios/policies. The temporal heatmap showed that the negative effects of constant immigration will only happen from 2040 onwards. Also here the policy to increase FTE per worker determines behaviour in the unemployment rate during the first 15 to 20 years. From PRIM it can be concluded that unemployment rates above 10% will occur when there is a high yearly increase of immigration in the workforce *combined* with relatively long hiring times. Since the immigration scenario is active during the whole simulation period the inflow of new workers keeps steadily increasing. But the relatively high 'hiring time' is not able to keep up with this increased inflow. Therefore, the model shed light upon the importance of job-searching dynamics. How long does it take for someone joining the labour market for the first time to find a job? These two factors will let the unemployment rate rise. However, the model suggests that the increase in FTE per worker might negate this effect for the short term.

When interpreting these results it is important to take into account all the points mentioned in the Discussion chapter 5. It is very questionable that when immigration occurs because of high net FTE scarcity, that those people will come live in the Netherlands but will find difficulties in finding a job. However, the purpose of this model is not to make accurate forecasts on macro-economic variables, it is rather meant to explore behaviour. The analyses have provided some useful insights in behaviour that might have been overlooked at first glance. The model has proven to be a decent basis for conducting initial policy and strategy exploration. Furthermore, the model provides several opportunities for improvement to be able to better understand other parts of long-term labour market dynamics.

Finally to answer the main research question of this research it is restated: *How do educational attainment and societal ageing influence the long-term labour market dynamics in the Netherlands?* What can be concluded is that societal ageing will have a long term impact on the labour market. The population getting older will mean that the current workforce need to carry the burden and labour demanded by a bigger population. It is important for policy makers to further understand the effects of increasing the FTE's supplied by the labour force. Either through immigration, through increases of the formal retirement age, or by creating policy to stimulate working more hours. Even though the model suggest that increases in the retirement age will have little effect, it is assumed that in the real world this effect will be visible. According to the model the educational attainment in the Netherlands will be of little effect. Rather, the skill set across all educational levels is important. It is therefore advised to adapt policy such that the workforce is made up of workers that can fulfill jobs created by technological development and the extra labour demand that the future will bring.

6.2 Recommendations

Now that conclusions have been drawn based on the results, some recommendations for policy makers will be proposed. This research and model have been subject to deep uncertainty, and therefore it is advisable that the model is only used as a tool to understand labour market dynamics, rather than forecasting the outcomes of policies. Policy makers may use the insights of this research for iteratively adapting policy for cases where the future development of technology (e.g. AI) will create more jobs, rather than it destroys. If this may happen, and the demand for labour will steadily increase, it is important that the educational system will quickly react to these developments. By catering study programmes to specifics of technological development and labour demand, the Netherlands will create an adequate supply of skilled workers (e.g. AI developers) that can satiate this demand. It is not recommended to solely focus on the educational attainment of the country, but rather on the contents of the different programmes. By having a future workforce that is both practically educated (mbo or lower) and academic educated (wo), the rises in net FTE scarcity will be partially negated.

To further help lower the net FTE scarcity an increase in immigration in the labour workforce may slightly help. This increase in immigration can both be seen as a scenario as a policy since only 21 through 45 year-olds will immigrate and actively participate in the labour market when they can. However, it is important that those people coming to the Netherlands are quickly starting a job. Since the unemployment rate will rise when it takes a relatively long time for these workers to find an appropriate job or start their job. If the government actively chooses to pursue workforce immigration it is advised to create programmes or policy that help these workers to quickly start their new job.

These policies should be occurring in combination with a focus of increasing the FTE per worker in the Netherlands. The model showed that the policy will have most effect if it is been put in place between now and 15 years from now with a focus of persons older than 40 up until the formal retirement age. Due to the design of the technological development scenarios the positive effect of the FTE policy will be negated once technological development starts to take off. At that moment the scarcities are dependent on how labour productivity and thus labour demand will react on technological development.

6.3 Policy design

This section proposes a policy strategy that can be possibly adopted in real world decision making. Note that this design is based on the model outcomes and is therefore very sensitive to how specific future scenarios will play out. However, it is still possible to use the proposed design as a starting point which then can be changed and adapted as time passes, which is a standard practice in political decision making. The policy design is two-fold, first focusing on short to midterm implications (i.e. within 15 years from now), and secondly on longer term implications (i.e. 20 years from now).

6.3.1 Short- to midterm policy design

Currently the dutch government is working on a policy to encourage working full-time, and hence increasing the average FTE per worker. By providing part-time workers with a financial bonus the government hopes to get people to put in more hours of work (Blotenburg, Suzanne, 2022). However, it remains the question if this is legally feasible, since currently no financial bonus will be granted for persons that already are supplying 40 hours of labour per week. Furthermore, the question remains which demographic group should be targeted with this policy. Consider for instance the group of part-time workers that are mothers with young children. It is very possible that this group does not have a lot of flexibility in working more hours even if a financial incentive were put in place.

Therefore it might be wise to shift the target group to older age cohorts who will, in general, have older or no children at all. The model showed that an FTE increase of people older than 40 years old would be effective. How then, can this be achieved?

First, the government could provide the possibility for persons that receive unemployment benefits to do paid work for a few hours per week such that this has no impact or consequences on their unemployment benefit. A possible result of this could be that the distance those people have from the labour market may decrease. Secondly, the government could make it fiscally more attractive for workers to work for multiple employers. Finally, more hours could be drawn from person already providing a full FTE. This could be done for instance by saying that for every extra hour worked, their gross salary will be their net salary. This policy could be augmented with a directed immigration regime. It is called a regime since immigration can partially be policy or a scenario. By having a (seasonal) immigration regime across 21 through 45 year olds, part of the net FTE scarcity can be resolved. However, it is of great importance that the allocation of those extra workers happens in a quick fashion, such that the unemployment rate will not rise.

6.3.2 Long-term policy design

As for long-term policy design it is advised for the government and educational systems to prepare and adapt to the potential of technological advancements. As of now, it is difficult to asses if technological advancement, and specifically AI, will create or destroy jobs. The model suggests that in the first case, scarcities will grow, and thus implies the need for workers that have the qualified skills related to these technological advancements. Therefore it is advised that students at primary and high school are learnt the adequate subjects such that they can make an appropriate study choice, in for instance ICT. While schools are introducing their students to programming from a young age, it is not yet part of the standard curriculum (Witlox, Malini, 2022).

The main objective of this policy should be to ensure a sufficient supply of skilled workers from the educational system to meet the labour market's demand caused by technological advancements. By offering subsidies to primary and secondary schools, students can receive 1 or 2 hours of programming education per week. This initiative aims to enhance students' understanding of various subjects, enabling them to make informed choices and potentially increasing enrollment in AI/tech-related studies. Additionally, mbo, hbo, and wo schools should have a comprehensive understanding of current trends happening in the labour market. They should adapt the curricula of specific programmes or consider creating new programmes, as long as it is practical and feasible.

6.4 Suggestions for future work

The suggestions for future work will be made with the purpose of the model in mind. So, how can the model and research be improved to better fit the purpose of understanding the effect of educational attainment and societal ageing on long term labour market dynamics. To better understand the effects of educational attainment it would be advised to find methods to measure latent factors that determine study choices. If these measures reveal latent factors that can be coupled to the labour market this can be explicitly modeled in an SD model to add an extra feedbackloop. Moreover, gaining a better understanding of wage-setting structures may be helpful in improving the model. By delving deeper into the mechanisms that govern wage determination, such as negotiation processes, collective bargaining, or market forces, a more realistic portrayal of wage dynamics can be achieved. This refinement would allow for a more accurate representation of how wages respond to changes in labour supply and demand, providing insights into how wage adjustments may influence labour market outcomes.

Furthermore, an extensive policy and stakeholder analysis may be performed to develop better policy variables which can be included in the model such that it becomes closer to the real world system. By closely examining the policy landscape and engaging with relevant stakeholders, a more comprehensive set of policy variables can be identified and incorporated. Since the purpose of the model is to understand long term labour market dynamics in the Netherlands, realistic and better adapted policies in the model may reveal other behaviour that might have been overlooked.

References

ABU. (2022). Arbeidsmigranten in nederland: De cijfers [Accessed: 2023-06-06]. https://www.abu.nl/app/uploads/2021/06/HR_ABU_NBBU_Factsheet_Arbeidsmigranten_A3_juni-2021_DEF.pdf

Ackoff, R. L., et al. (1974). Redesigning the future. *New York*, *29*.

Agrey, L., & Lampadan, N. (2014). Determinant factors contributing to student choice in selecting a university. *Journal of Education and Human Development*, *3*(2), 391–404.

Aiyetan, O. A., & Das, D. (2018). System dynamics approach to mitigating skilled labour shortages in the construction industry: A south african context. *Construction Economics and Building*, *18*(4), 45–63.

Altun, Y., Das, K., Mielikäinen, T., Malerba, D., Stefanowski, J., Read, J., Žitnik, M., Ceci, M., & Džeroski, S. (2017). *Machine learning and knowledge discovery in databases: European conference, ecml pkdd 2017, skopje, macedonia, september 18–22, 2017, proceedings, part iii* (Vol. 10536). Springer.

Auping, W. L. (2018). Modelling uncertainty: Developing and using simulation models for exploring the consequences of deep uncertainty in complex problems.

Auping, W. L., Pruyt, E., & Kwakkel, J. H. (2015). Societal ageing in the netherlands: A robust system dynamics approach. *Systems Research and Behavioral Science*, *32*(4), 485–501.

Bankes, S. (1993). Exploratory modeling for policy analysis. *Operations research*, *41*(3), 435–449.

Binmore, K., Rubinstein, A., & Wolinsky, A. (1986). The nash bargaining solution in economic modelling. *The RAND Journal of Economics*, 176–188.

Blázquez Cuesta, M. (2005). Youth labour market integration in spain: Search time, job duration and skill mismatch. *Spanish Economic Review*, *7*(3), 191–208.

Blotenburg, Suzanne. (2022). Bonus voor voltijd werken: Dit zijn de plannen [Accessed: 2023-07-01]. https://www.rtlnieuws.nl/economie/artikel/5336214/voltijdsbonus-d66-jan-paternotte-personeelstekort

Boeri, T., & van Ours, J. (2021). *The economics of imperfect labor markets*. Princeton University Press.

Brücker, H., & Jahn, E. J. (2011). Migration and wage-setting: Reassessing the labor market effects of migration. *Scandinavian Journal of Economics*, *113*(2), 286–317.

Bryant, B. P., & Lempert, R. J. (2010). Thinking inside the box: A participatory, computer-assisted approach to scenario discovery. *Technological Forecasting and Social Change*, *77*(1), 34–49.

CBS. (2018). 1 op 3 werkt 10 jaar of langer bij dezelfde werkgever [Accessed: 2023-05-07]. https://www.cbs.nl/nl-nl/nieuws/2018/17/1-op-3-werkt-10-jaar-of-langer-bij-dezelfde-werkgever

CBS. (2020a). Ho-cohorten; behaalde hbo- en wo-diploma's [Accessed: 2023-04-24]. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83282NED/table?ts=1682338786070

CBS. (2020b). Ho-cohorten; behaalde hbo- en wo-diploma's [Accessed: 2023-04-12]. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83282NED/table?ts=1682338786070

CBS. (2020c). Immi- en emigratie; geslacht, leeftijd, nationaliteit, regio [Accessed: 2023-04-16]. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85451NED/table?ts=1684824631282

CBS. (2020d). Mbo; doorstroom en uitstroom, achtergrondkenmerken 2008/'09 - 2021/'22 [Accessed: 2023-04-12]. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84553NED/table?ts=1684917239384

CBS. (2020e). Vo en vso; doorstroom en uitstroom, achtergrondkenmerken [Accessed: 2023-04-12]. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84754NED/table?ts=1683017326037

CBS. (2020f). Voortijdig schoolverlaters; geslacht, onderwijssoort en herkomst [Accessed: 2023-04-24]. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85368NED/table?ts=1682337792911

CBS. (2022a). Arbeidsmarkt iets minder gespannen in derde kwartaal [Accessed: 2022-17-11]. https://www.cbs.nl/nl-nl/nieuws/2022/46/arbeidsmarkt-iets-minder-gespannen-in-derde-kwartaal

CBS. (2022b). Bbp, productie en bestedingen; kwartalen, mutaties, nationale rekeningen [Accessed: 2023-04-16]. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84106NED/table?ts=1687249910860

CBS. (2022c). Bevolking; kerncijfers [Accessed: 2023-03-15]. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85496NED/table?ts=1688544026199

CBS. (2022d). Meer asiel- en kennismigranten van buiten de eu in 2021 [Accessed: 2023-06-06]. https://www.cbs.nl/nl-nl/nieuws/2022/29/meer-asiel-en-kennismigranten-van-buiten-de-eu-in-2021

CBS. (2022e). Werkzame beroepsbevolking [Accessed: 2023-05-22]. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85272NED/table?ts=1681378883067

CBS. (2023a). Beroepsbevolking [Accessed: 2023-06-06]. https://www.cbs.nl/nl-nl/visualisaties/dashboard-beroepsbevolking/beroepsbevolking

CBS. (2023b). Bevolking in de toekomst [Accessed: 2023-07-03]. https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/bevolkingsgroei/toekomst#:~:text=De%20bevolking%20van%20Nederland%20blijft,migratie%20en%20een%20stijgende%20levensduur.

Dawson, N., Rizoiu, M.-A., Johnston, B., & Williams, M.-A. (2020). Predicting skill shortages in labor markets: A machine learning approach. *2020 IEEE International Conference on Big Data (Big Data)*, 3052–3061.

Devadoss, S., & Luckstead, J. (2018). Us immigration policies and dynamics of cross-border workforce in agriculture. *The World Economy*, *41*(9), 2389–2413.

Drewes, T., & Michael, C. (2006). How do students choose a university?: An analysis of applications to universities in ontario, canada. *Research in Higher Education*, *47*, 781–800.

Elfferich, A. (2022). *Stayrate en arbeidsmarktpositie van internationale afgestudeerden in nederland* (tech. rep.). Nuffic.

Eurostat. (2022a). 41% of young adults hold a tertiary degree [Accessed: 2023-03-08]. https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20220524-2

Eurostat. (2022b). Full-time and part-time employment by sex, age and educational attainment level [Accessed: 2022-17-11]. https://ec.europa.eu/eurostat/databrowser/view/lfsa_epgaed/default/map?lang=en

EurWORK. (2022). Labour shortage [Accessed: 2023-05-13]. https://www.eurofound.europa.eu/observatories/eurwork/industrial-relations-dictionary/labour-shortage

Fagiolo, G., Dosi, G., & Gabriele, R. (2004). Matching, bargaining, and wage setting in an evolutionary model of labor market and output dynamics. *Advances in Complex Systems*, *7*(02), 157–186.

Filardo, A. J., & Gordon, S. F. (1998). Business cycle durations. *Journal of econometrics*, *85*(1), 99–123.

Fishwick, P. A. (1995). *Simulation model design and execution: Building digital worlds*. Prentice Hall PTR.

Forrester, J. W. (1961). *Industrial dynamics*. MIT Press.

Forrester, J. W., Mass, N. J., & Ryan, C. J. (1976). The system dynamics national model: Understanding socio-economic behavior and policy alternatives. *Technological Forecasting and Social Change*, *9*(1-2), 51–68.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E., et al. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences*, *116*(14), 6531–6539.

Friedman, J. H., & Fisher, N. I. (1999). Bump hunting in high-dimensional data. *Statistics and computing*, *9*(2), 123–143.

Gavrilov, L. A., & Gavrilova, N. S. (2011). Mortality measurement at advanced ages: A study of the social security administration death master file. *North American actuarial journal*, *15*(3), 432–447.

Giabelli, A., Malandri, L., Mercorio, F., & Mezzanzanica, M. (2020). Graphlmi: A data driven system for exploring labor market information through graph databases. *Multimedia Tools and Applications*, 1–30.

Gompertz, B. (1825). Xxiv. on the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. in a letter to francis baily, esq. frs &c. *Philosophical transactions of the Royal Society of London*, (115), 513–583.

Hall, J. W., Lempert, R. J., Keller, K., Hackbarth, A., Mijere, C., & McInerney, D. J. (2012). Robust climate policies under uncertainty: A comparison of robust decision making and info-gap methods. *Risk Analysis: An International Journal*, *32*(10), 1657–1672.

Jennings, N. R. (1999). Agent-based computing: Promise and perils.

Korkmaz, S., & Korkmaz, O. (2017). The relationship between labor productivity and economic growth in oecd countries. *International Journal of Economics and Finance*, *9*(5), 71–76.

Kruiskamp, P. (2022). *Sbi 2008 versie 2018 update 2022* (tech. rep.). CBS.

Kwakkel, J. H. (2017). The exploratory modeling workbench: An open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. *Environmental Modelling Software*, *96*, 239–250. https://doi.org/https://doi.org/10.1016/j.envsoft.2017.06.054

Leitner, S. M. (2022). A skill-specific dynamic labour supply and labour demand framework: A scenario analysis for the western balkan countries to 2030. *LABOUR*, *36*(4), 471–504.

Lempert, R. J. (2003). Shaping the next one hundred years: New methods for quantitative, long-term policy analysis.

Lempert, R. J., Groves, D. G., Popper, S. W., & Bankes, S. C. (2006). A general, analytic method for generating robust strategies and narrative scenarios. *Management science*, *52*(4), 514–528.

Levanon, G., Cheng, B., & Paterra, M. (2014). The risk of future labor shortages in different occupations and industries in the united states. *Business Economics*, *49*(4), 227–243.

Logtens, T., Pruyt, E., & Gijsbers, G. (2012). Societal aging in the netherlands: Exploratory system dynamics modeling and analysis. *Proceedings of the 30th International Conference of the System Dynamics Society*, 22–26.

Makeham, W. M. (1860). On the law of mortality and the construction of annuity tables. *Journal of the Institute of Actuaries*, *8*(6), 301–310. https://doi.org/10.1017/S204616580000126X

Ministerie van Onderwijs, Cultuur en Wetenschap. (2022a). De staat van het onderwijs 2022 [Accessed: 2023-04-24]. https://www.onderwijsinspectie.nl/documenten/rapporten/2022/04/13/de-staat-van-het-onderwijs-2022

Ministerie van Onderwijs, Cultuur en Wetenschap. (2022b). Prognose aantal studenten mbo [Accessed: 2023-03-07]. https://www.ocwincijfers.nl/sectoren/middelbaar-beroepsonderwijs/studenten/prognose-aantal-studenten-mbo/

Ministerie van Onderwijs, Cultuur en Wetenschap. (2022c). Prognose aantal studenten wo [Accessed: 2023-03-07]. https://www.ocwincijfers.nl/sectoren/hoger-onderwijs/kengetallen-wetenschappelijk-onderwijs/studenten/prognose-aantal-studenten-wo/

Ministry of Education, Culture and Science. (2020). Onderwijsverslag: De staat van het onderwijs 2020 [Accessed: 2023-03-16]. https://www.onderwijsinspectie.nl/onderwerpen/staat-van-het-onderwijs/archief

Ministry of Health, Welfare and Sport. (2022). Levensverwachting (waargenomen) en prognose 1950-2070 [Accessed: 2023-04-24]. https://www.vzinfo.nl/levensverwachting/leeftijd-en-geslacht

Ministry of Social Affairs and Employment. (2022). Kamerbrief over aanpak krapte op arbeidsmarkt [Accessed: 2022-17-11]. https://www.rijksoverheid.nl/documenten/kamerstukken/2022/06/24/kamerbrief-over-aanpak-krapte-op-arbeidsmarkt

Neugart, M., & Richiardi, M. G. (2018). Agent based models of the labor market.

Niazi, M., & Hussain, A. (2011). Agent-based computing from multi-agent systems to agent-based models: A visual survey. *Scientometrics*, *89*(2), 479–499.

NIDI. (2021). Beleidskeuzes in een vergrijzende maatschappij [Accessed: 2022-17-11]. https://nidi.nl/demos/beleidskeuzes-in-een-vergrijzende-maatschappij/

NOS. (2023). Nog steeds zorgen om basisvaardigheden leerlingen, 'scholen kunnen meer doen' [Accessed: 2023-05-07]. https://nos.nl/artikel/2474543-nog-steeds-zorgen-om-basisvaardigheden-leerlingen-scholen-kunnen-meer-doen

OECD. (2019). *Negotiating our way up*. https://doi.org/https://doi.org/https://doi.org/10.1787/1fd2da34-en

OECD. (2022a). *Education at a glance 2022*. https://doi.org/https://doi.org/https://doi.org/10.1787/3197152b-en

OECD. (2022b). Fertility rates [Accessed: 2023-03-16]. https://data.oecd.org/pop/fertility-rates.htm

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, *263*(5147), 641–646.

Osborne, M. J., & Rubinstein, A. (1990). *Bargaining and markets*. Academic press.

Pruyt, E., & Logtens, T. (2015). Model-based exploration of societal aging in the netherlands. *International Journal of System Dynamics Applications (IJSDA)*, *4*(1), 57–95.

PWC. (2017). *The long view how will the global economic order change by 2050?* (Tech. rep.). PWC.

Rapoport, B., & Thibout, C. (2018). Why do boys and girls make different educational choices? the influence of expected earnings and test scores. *Economics of Education Review*, *62*, 205–229.

Rijksoverheid. (2020). Wet temporisering verhoging aow-leeftijd [Accessed: 2023-04-13]. https://www.rijksoverheid.nl/onderwerpen/pensioen/toekomst-pensioenstelsel/aow-leeftijd-stijgt-minder-snel

Rijksoverheid. (2021). Bedragen minimumloon 2021 [Accessed: 2023-04-16]. https://www.rijksoverheid.nl/onderwerpen/minimumloon/bedragen-minimumloon/bedragen-minimumloon-2021

Rijksoverheid. (2023a). After vmbo [Accessed: 2023-05-07]. https://www.government.nl/topics/secondary-education/different-types-of-secondary-education/pre-vocational-secondary-education-vmbo/after-vmbo

Rijksoverheid. (2023b). Schooladvies toelating voortgezet onderwijs [Accessed: 2023-05-07]. https://www.rijksoverheid.nl/onderwerpen/schooladvies-en-eindtoets-basisschool/toelating-voortgezet-onderwijs-gebaseerd-op-definitief-schooladvies

Rijksoverheid. (2023c). Senior general secondary education (havo) and pre university education (vwo) [Accessed: 2023-05-07]. https://www.government.nl/topics/secondary-education/different-types-of-secondary-education/senior-general-secondary-education-havo-and-pre-university-education-vwo

Rijksoverheid. (2023d). Stap-budget voor scholing en ontwikkeling [Accessed: 2023-06-14]. https://www.rijksoverheid.nl/onderwerpen/leven-lang-ontwikkelen/leven-lang-ontwikkelen-financiele-regelingen/stap-budget

Rijksoverheid. (2023e). Topics education [Accessed: 2023-05-07]. https://www.government.nl/topics/themes/education
Rijksoverheid. (2023f). Verbeteren van welzijn studenten door aanpassen bsa [Accessed: 2023-05-09]. https://www.rijksoverheid.nl/actueel/nieuws/2023/05/08/verbeteren-van-welzijn-studenten-door-aanpassen-bsa
Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy sciences*, *4*(2), 155–169.
ROA. (2020). Actuele arbeidsmarktsituatie per bedrijfssector, gemiddelde 2019-2020 [Accessed: 2023-03-11]. https://roastatistics.maastrichtuniversity.nl/AIS/Login.aspx?
Rutten, A. T. G. J., Knoef, M. G., & van Vuuren, D. J. (2023). *The effect of the statutory retirement age on spousal labor force participation* (tech. rep.). Netspar.
Senge, P. M., & Forrester, J. W. (1980). Tests for building confidence in system dynamics models. *System dynamics, TIMS studies in management sciences*, *14*, 209–228.
Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly journal of economics*, *70*(1), 65–94.
Solow, R. M. (1957). Technical change and the aggregate production function. *The review of Economics and Statistics*, *39*(3), 312–320.
Sterman, J. D. (1994). Learning in and about complex systems. *System dynamics review*, *10*(2-3), 291–330.
Sterman, J. D. (2001). System dynamics modeling: Tools for learning in a complex world. *California management review*, *43*(4), 8–25.
Sterman, J. D. (2002). All models are wrong: Reflections on becoming a systems scientist. *System Dynamics Review: The Journal of the System Dynamics Society*, *18*(4), 501–531.
Stiroh, K. J. (2001). What drives productivity growth? *Economic Policy Review*, *7*(1).
Stock, J. H., & Watson, M. (2011). Dynamic factor models. *Oxford Handbooks Online*.
Sutrisno, A., & Handel, O. (2011). Dynamic aging population in germany: A case study about demographic change. *University of Bergen*.
Tweede Kamer der Staten Generaal. (2023). Buitenlandse studenten in het nederlandse hoger onderwijs [Accessed: 2023-06-07]. https://www.tweedekamer.nl/kamerstukken/plenaire_verslagen/kamer_in_het_kort/buitenlandse-studenten-het-nederlandse-hoger
Ventana Systems. (2010). *Vensim reference manual* (tech. rep.). Harvard, MA.
Verdoorn, P. J. (2002). Factors that determine the growth of labour productivity. *Productivity growth and economic performance: Essays on Verdoorn's Law*, 28–36.
Webb, M. (2019). The impact of artificial intelligence on the labor market. *Available at SSRN 3482150*.
Witlox, Malini. (2022). Mama, ik wil leren programmeren [Accessed: 2023-07-01]. https://www.ad.nl/tech/mama-ik-wil-leren-programmeren~ab5c64d1/
Worldbank. (2020). Fertility rate, total (births per woman) - netherlands [Accessed: 2023-03-16]. https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?locations=NL

Appendix

Appendix A — Model background

In table 4 the dominant education level in each industry is calculated. Based on data of ROA (2020) the distribution of educational attainment per industry was obtained. This ranged from primary all the way until a university masters degree. In the model three new categories were created and named: low, middle and high. Where low contained the subrange: primary, vmbo, havo, vwo, mbo; middle contained hbo; and high contained wo. The first three collumns in the table show what percentage of an industry is made up by these three different categories. Then a weighted value is taken as follows:

$$\sum_{e=1}^{3} e \cdot \frac{X_e}{100}, \tag{26}$$

Where $e = 1$ corresponds to low education, $e = 2$ to middle education and $e = 3$ to high education. Hence for the agriculture industry the value $X_1 = 82$.

Table 4 Dominant educational level per industry

Industry	Low	Middle	High	Weighted Values	Vensim Values
Agriculture	82	14	4	1.22	1
Mining	66	23	12	1.48	2
Industry	61	25	13	1.5	2
Energy	61	24	18	1.63	2
Water and Waste	78	15	5	1.23	1
Construction	80	14	4	1.2	1
Reatail	85	11	4	1.19	1
Transport	77	15	6	1.25	1
Horeca	85	10	2	1.11	1
ICT	35	39	24	1.85	3
Finance	39	34	27	1.88	3
Real Estate	39	34	27	1.88	3
Business Services	31	32	36	2.03	3
Rental	71	20	8	1.35	1
Government	42	30	26	1.8	3
Education	24	46	21	1.79	3
Healthcare	35	35	30	1.95	3
Culture and Sports	51	34	15	1.64	2
Other Services	67	18	14	1.45	2
Households	67	18	14	1.45	2

Note: Threshold range Low: [1.11, 1.41). Mid: [1.41, 1.7). High [1.7, 2.03]. The last column depicts the values given to the model. A '1' means that in that industry mbo or lower educated people are dominating the workforce. A '2' means that hbo people dominate that industry. And a '3' means wo people dominate that industry. The goal of this table is to create a vector of 20 values that can be used by the Vensim software for the allocate available function.

After all the weighted values have been calculated we divide the range of the weighted values in three equally big sections. These can be seen under the table, named 'thresholds'. Finally each weighted value gets labeled either 1, 2 or 3 and this determines how the preferences of each industry will be modelled in Vensim. For instance, industries active in business services will prefer highly educated people, while industries in construction would prefer lower educated people.

Vensim allocate available function

Next, the allocate available function can be explained. The model makes use of a function called 'allocate available' within the Vensim modelling software. This section is largely based on the information available in the user guide of Vensim here or copy this link https://www.vensim.com/documentation/allocation_overview.html (Ventana Systems, 2010). By clicking on the 'next' arrow in the top right the different pages of the user guide can be viewed. The allocation function builds upon 6 properties,

- Conservation of matter: The amount received by demanders must be equal to the amount provided by the suppliers, if summed across all demanders and suppliers)
- Nonnegative: Quantities must be non-negative
- Conservation of intent: No supplier will provide more than it desired to supply and demanders shall receive no more than it has chosen to demand
- No loopholes: If there is adequate supply or demand, each demander or supplier should receive or supply appropriately.
- Clear differentiation: When supply is insufficient, extremely low priority demanders should receive little. High priority should receive more. For suppliers this works the same way
- Continuity: Small changes in priorities, supply and demand should cause small changes in allocations.

The function makes use of the simple principle in economics that price will adjust such that supply and demand match. In this case there is no price for which market clears but this way of thinking still proved to be useful in the implementation of it. In the model, the unemployed group of persons can be seen as the scarce resource. Industries demand workers and have a preference for certain suppliers. The model makes a distinction between four different suppliers, 'unqualified', 'low edu', 'mid edu', and 'high edu'. Where the latter three are referencing to mbo, hbo and wo respectively, and the first one are all persons that have obtained at most an high school degree. Consider figure 18 which depicts a supply curve with constant demand. If demand is low (d1), the highest priority suppliers are used. When demand is higher (d2), more suppliers are used. If the demand is very high (d3) then all suppliers are used to their capacities. For example, if demand for workers in the agriculture industry is low, it will prioritise hiring from the 'unqualified' and 'low edu' group since it prefers to have workers with these educational backgrounds (see table 4). If demand is high in that specific industry, it will also take on workers that are maybe over-qualified or not necessarily preferred in that industry.

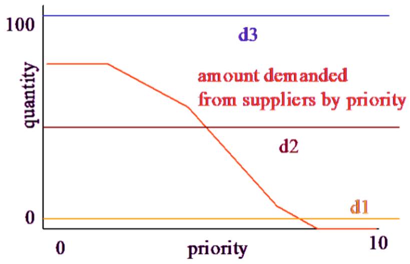


Figure 18 Simple supply demand curve

The preferences or priorities are modelled as a curve which may change over time. Vensim provides the following choices seen in figure 19. For most behaviour a normal distribution is customary and also chosen as a basis for the priority curves of this model.

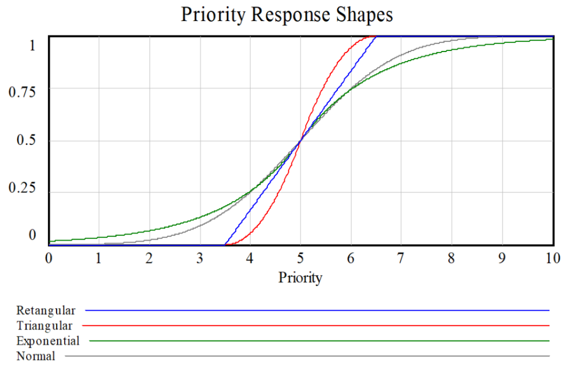


Figure 19 Priority response curves

The priority curves in the model are chosen to be normally distributed with a width (i.e. standard deviation) of 3. The priority curves for the four suppliers are given in figure 20.

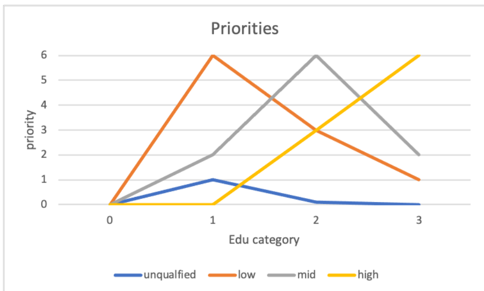


Figure 20 Priority curves

From the figure it can be seen that that industries with a preference for persons falling in edu category 1 (x-axis) have the highest priority for hiring from the pool of unemployed workers that are qualified as 'low' educated (see table 4). Industries that have a preference for hiring persons that are higher educated, like finance, will have the highest priority for people in the unemployed pool of workers that are highly educated. Persons that are 'unqualified' will almost never be hired by industries that have prefer only highly educated people, as indicated by the blue line.

The model has 4 stocks that portray the preference for hiring a certain educational level. This is subscripted for the 20 industries. The priority curves in figure 20 play an important role in determining the preferences. The four stocks work in the same way except that their initial value differ, since this depends on the initial priority curves. The Vensim model structure is adapted from the model from Logtens et al. (2012). This gives,

$$Plh_i(t) = Plh_i(t_0) + \int_{t_0}^{t} [sPlh_i(t) - dPlh_i(t)]\, dt, \tag{27}$$

where $Plh_i(t)$ is the preference for hiring low educated people for industry i, $Plh_i(t_0)$ is the value from the initial priority curves from figure 20, and $sPlh_i(t)$ and $dPlh_i(t)$ are the increase and decrease of the preference over time respectively. So for this stock the orange curve is considered, and the initial value $Plh_i(t_0)$ is the y-value corresponding to the dominant edu level for industry i taken from table 4. For example $Plh_{agriculture}(t_0) = 6$, since from table 4 we see that the agriculutre industry has a prefernce for hiring 'low' educated people. With that in mind we look at the orange curve of figure 20 at point 1 on the x-axis. The corresponding y-value there is 6.

As can be seen from equation 27 the preference for hiring workers with a specific educational attainment may change over time, this is dependent on the two flow variables written in the integral. The inflow is given as,

$$sPlh_i(t) = (6 - Plh_i(t)) \cdot \alpha_1, \tag{28}$$

where 6 comes from the maximum preference possible, which is by chosen to be 6 by design, $Plh_i(t)$) is the value of the preference stock at time t, and α_1 is the annual change in preference of workers. The subscript is used to make a distinction between the different α variables in the equations. This last variable can be seen as the speed of convergence of the preference stock to the midpoint with a value of 3. Similarly for the decrease flow we have,

$$dPlh_i(t) = (Plh_i(t) - 0) \cdot \alpha_1, \tag{29}$$

where the value 0 comes from the minimum preference possible, which is chosen by design.
The remaining three stocks which model the preference of an industry for hiring unqualified/mid/high educational levels work in the exact same way. The only difference is the initial value of the stock which is based on the blue, grey, and yellow curves respectively from figure 20.

Table 5 Industry descriptions

Industry	Description
Agriculture	Agriculture, Forestry, and Fishing. The exploitation of natural plant and animal resources, including crop farming, livestock breeding, and the production of other plants and animals on agricultural farms or in natural habitats
Mining	Mining and Quarrying. Extraction of Minerals. The extraction of naturally occurring minerals in solid form (such as coal, peat, and ores), in liquid form (such as petroleum), or in the form of a gas (such as natural gas).
Industry	The mechanical, physical, or chemical processing of materials, substances, or components into new products. The processed materials, substances, or components are raw materials from agriculture, forestry, fishing, and mining, as well as (semi)finished products from the industry.
Energy	Production, distribution, and trade of electricity, natural gas, steam, and cooled air.
Water and Waste	Water collection and distribution; waste and wastewater management and remediation.
Construction	General and specialized construction and civil engineering works, construction installation, and building finishing.
Retail	Workers active in retail stores. Includes supermarkets.
Transport	the transportation of passengers or goods, whether or not scheduled, by rail, pipeline, road, water, or air
Horeca	Hotels, restaurants and cafes. Accommodation, meal, and beverage provision.
ICT	The production and distribution of information, the provision of infrastructure for transmitting that information, as well as activities in the field of data and communication information technology and the processing of data and other information.
Finance	Insurance and pension funds. Financial markets, securities brokers.
Real Estate	Rental of real estate. Intermediation and management of real estate.
Business Services	Specialized professional, scientific, and technical activities. These activities require a high level of education and provide specific expertise.
Rental	Rental of movable goods and other business services.
Government	Public administration and government services.
Education	Includes all forms of public and private education, at every level and for every profession, whether oral, written, or through radio and television
Healthcare	Treatment in general and specialized hospitals
Culture and Sports	All activities related to cultural and sports activities and events.
Other services	Residual category. The activities of interest groups, computer and consumer goods repair, and many other activities in the field of personal services that are not classified elsewhere.
Households	Activities include, cleaning staff, private chefs and other jobs related to household activities that are paid.

Table 6 Distribution of educational attainment per industry, (ROA, 2020)

	Primary	Vmbo	Havo	Vwo	Mbo	Hbo	Wo
Agriculture	7.0%	20.0%	1.5%	1.5%	47.0%	17.0%	7.0%
Mining	4.0%	12.0%	1.5%	1.5%	43.0%	25.0%	14.0%
Inudstry	5.0%	10.0%	1.5%	1.5%	40.0%	27.0%	14.0%
Energy	4.0%	10.0%	1.5%	1.5%	36.0%	26.0%	19.0%
Water and Waste	13.0%	18.0%	1.5%	1.5%	40.0%	17.0%	7.0%
Construction	6.0%	20.0%	1.0%	1.0%	49.0%	16.0%	5.0%
Retail	8.0%	25.0%	5.5%	5.5%	37.0%	13.0%	6.0%
Transport	7.0%	18.0%	3.0%	3.0%	42.0%	17.0%	8.0%
Horeca	10.0%	24.0%	7.5%	7.5%	31.0%	12.0%	3.0%
ICT	1.0%	4.0%	3.0%	3.0%	21.0%	42.0%	26.0%
Finance	1.0%	4.0%	2.5%	2.5%	26.0%	36.0%	28.0%
Real estate	1.0%	4.0%	2.5%	2.5%	26.0%	36.0%	28.0%
Business services	1.0%	3.0%	2.0%	2.0%	21.0%	33.0%	37.0%
Rental	8.0%	15.0%	3.0%	3.0%	35.0%	22.0%	10.0%
Government	1.0%	5.0%	2.0%	2.0%	30.0%	31.0%	27.0%
Education	1.0%	3.0%	2.0%	2.0%	14.0%	47.0%	32.0%
Healthcare	1.0%	3.0%	1.5%	1.5%	26.0%	36.0%	31.0%
Culture and sports	3.0%	9.0%	5.0%	5.0%	26.0%	36.0%	16.0%
Other services	4.0%	10.0%	2.0%	2.0%	45.0%	20.0%	16.0%
Households	4.0%	10.0%	2.0%	2.0%	45.0%	20.0%	16.0%

Table 7 GAV per industry, (CBS, 2022b)

	Gross Added Value (GAV) in €/year
Agriculture	1.33e+10,
Mining	5.546e+09,
Inudstry	8.7602e+10,
Energy	9.227e+09,
Water and Waste	4.415e+09,
Construction	3.6001e+10,
Retail	1.01787e+11,
Transport	3.4744e+10,
Horeca	1.5291e+10,
ICT	3.6673e+10,
Finance	4.7514e+10,
Real estate	5.2799e+10,
Business services	5.8315e+10,
Rental	5.1883e+10,
Government	5.0768e+10,
Education	3.5117e+10,
Healthcare	6.6954e+10,
Culture and sports	8.009e+09,
Other services	8.206e+09,
Households	8.09e+08

Table 8 Employed Female FTE's, (CBS, 2022e)

Females employed FTE for 10 year age cohorts	**15-25**	**26-35**	**36-45**	**46-55**	**56-65**	**66-75**
Agriculture	9000	4000	8000	17000	11000	5000
Mining	0	0	0	0	1000	0
Inudstry	19000	41000	37000	55000	32000	4000
Energy	0	2000	2000	1000	2000	0
Water and Waste	1000	3000	1000	2000	2000	0
Construction	2000	7000	10000	16000	9000	0
Retail	246000	89000	85000	105000	69000	6000
Transport	8000	24000	18000	30000	18000	1000
Horeca	132000	38000	19000	33000	19000	6000
ICT	8000	28000	18000	19000	7000	1000
Finance	6000	20000	27000	26000	15000	1000
Real estate	2000	7000	7000	9000	5000	1000
Business services	24000	67000	58000	72000	39000	5000
Rental	44000	54000	51000	55000	31000	5000
Government	15000	41000	57000	75000	46000	3000
Education	25000	92000	93000	94000	78000	9000
Healthcare	124000	297000	229000	265000	245000	26000
Culture and sports	24000	25000	16000	19000	19000	3000
Other services	14000	27000	25000	29000	24000	7000
Households	1000	2000	1000	5000	5000	3000

Table 9 Employed Male FTE's, (CBS, 2022e)

Males employed FTE for 10 year age cohorts	**15-25**	**26-35**	**36-45**	**46-55**	**56-65**	**66-75**
Agriculture	22000	16000	15000	27000	26000	8000
Mining	1000	1000	4000	2000	1000	0
Inudstry	61000	135000	130000	179000	142000	15000
Energy	1000	6000	8000	8000	6000	0
Water and Waste	3000	4000	4000	8000	12000	1000
Construction	31000	67000	91000	97000	86000	6000
Retail	213000	141000	130000	135000	101000	17000
Transport	37000	55000	53000	69000	66000	18000
Horeca	118000	31000	19000	27000	15000	4000
ICT	20000	68000	56000	53000	27000	3000
Finance	7000	30000	37000	49000	32000	4000
Real estate	3000	10000	8000	10000	8000	2000
Business services	22000	86000	85000	98000	83000	21000
Rental	62000	82000	57000	46000	41000	13000
Government	15000	49000	59000	83000	81000	8000
Education	20000	51000	42000	47000	52000	13000
Healthcare	21000	64000	44000	53000	55000	9000
Culture and sports	18000	25000	17000	20000	14000	5000
Other services	6000	8000	7000	14000	13000	4000
Households	0	0	0	1000	1000	0

Appendix B — Model equations

This section provides most equations present in the model.
The policy related to the increase in FTE per worker is dependent on a binary switch variable. This switch can take on the value 0 and 1. To understand how the FTE per worker is increased if this swtich is turned on, it is necessary to first describe some auxillary variables. Define τ as the starting year in the simulation when this policy is implemented. Define γ as the minimum age for which this policy is implemented. For example, if $\gamma = 40$ the policy will increase the FTE per worker for people aged 40 and above, if they are part of the working-age population. The maximum and minimum value for FTE per worker is also necessary, which are 1 and 0 by definition. Define α_2 as the annual change in FTE per worker, which can be understood as the speed of convergence of the FTE per worker to the maximum level of 1. All these variables are needed to describe the flow variables for the following stock,

$$F_a(t) = F_a(t_0) + \int_{t_0}^{t} [sF_a(t) - dF_a(t)]\, dt, \tag{30}$$

where $F_a(t)$ is the value of the stock FTE per worker for a specific age cohort a at time t, with initial value $F_a(t_0)$ and the in- and outflow variables $sF_a(t)$ and $dF_a(t)$. Here the inflow, or increase in FTE per worker is defined as,

$$sF_a(t) = \mathbb{1}_{t\geq\tau}\mathbb{1}_{a\geq\gamma}\mathbb{1}_{a\in WP}(1 - F_a(t)) \cdot \alpha_2, \tag{31}$$

where WP refers to the all the age cohorts that are part of the working-age population. Equation 31 is nonzero in the case when the time t is above a certain value τ, and the age cohort is above a threshold age γ but still within the set WP which is the working age population. The 1 comes from the maximum value that the stock FTE per worker can take, and α_2 is the annual change in FTE per worker. The outflow is defined as,

$$dF_a(t) = \mathbb{1}_{a\in WP}(F_a(t) - 0) \cdot \alpha_2, \tag{32}$$

where 0 is the minimum value the FTE per worker can take by definition.

In equation 20 a specific multiplier is used to portray the retraining of workers which will become available to work in a different industry. This variable is only 'activated' once the unemployment rate is above a certain threshold for a certain number of years. This yields,

$$m(t) = \mathbb{1}_{UR\geq 0.1}\mathbb{1}_{\mu\geq c}SMOOTH(m), \tag{33}$$

where UR is the unemployment rate, 0.1 is a model assumption for an unsustainable high unemployment rate, μ is the amount of years that the unemployment rate is above this threshold of 0.1 and c is the variable called threshold number of years long unemployment. SMOOTH refers to an internal Vensim function that smooths the variable m over time to make sure that the multiplier does not suddenly jump to a value of 1.2 as an example. More information can be found in the Vensim user manual here or via this url https://www.vensim.com/documentation/

mgu09_material_and_information_delays.html?q=delays (Ventana Systems, 2010).

The firing of workers is a flow that is used to determine the outflow of the employed workers stock from equation 17. Firing will occur when there are too many workers in a specific industry compared to industry demand. Therefore the flow looks if the relative FTE scarcity from equation 23 is negative. This yields,

$$f(t)_{i,e,a} = \max\left\{0, \frac{\max\{-\textit{relative FTE scarcity}_i, 0\}}{\textit{firing time} \cdot E_{i,e,a}(t)}\right\}, \tag{34}$$

where *firing time* is the average time a person will remain at the same employer.

Appendix C — Model visualisation

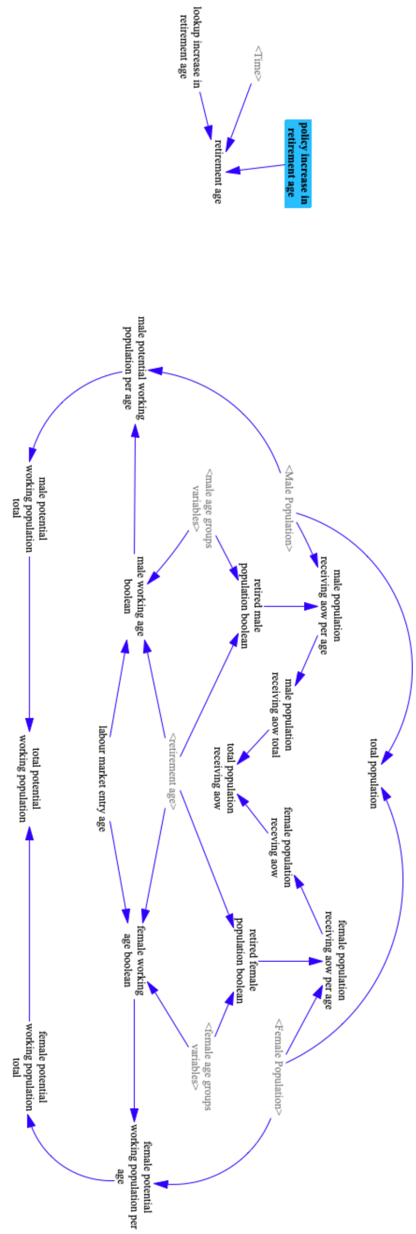


Figure 21 Population calculations

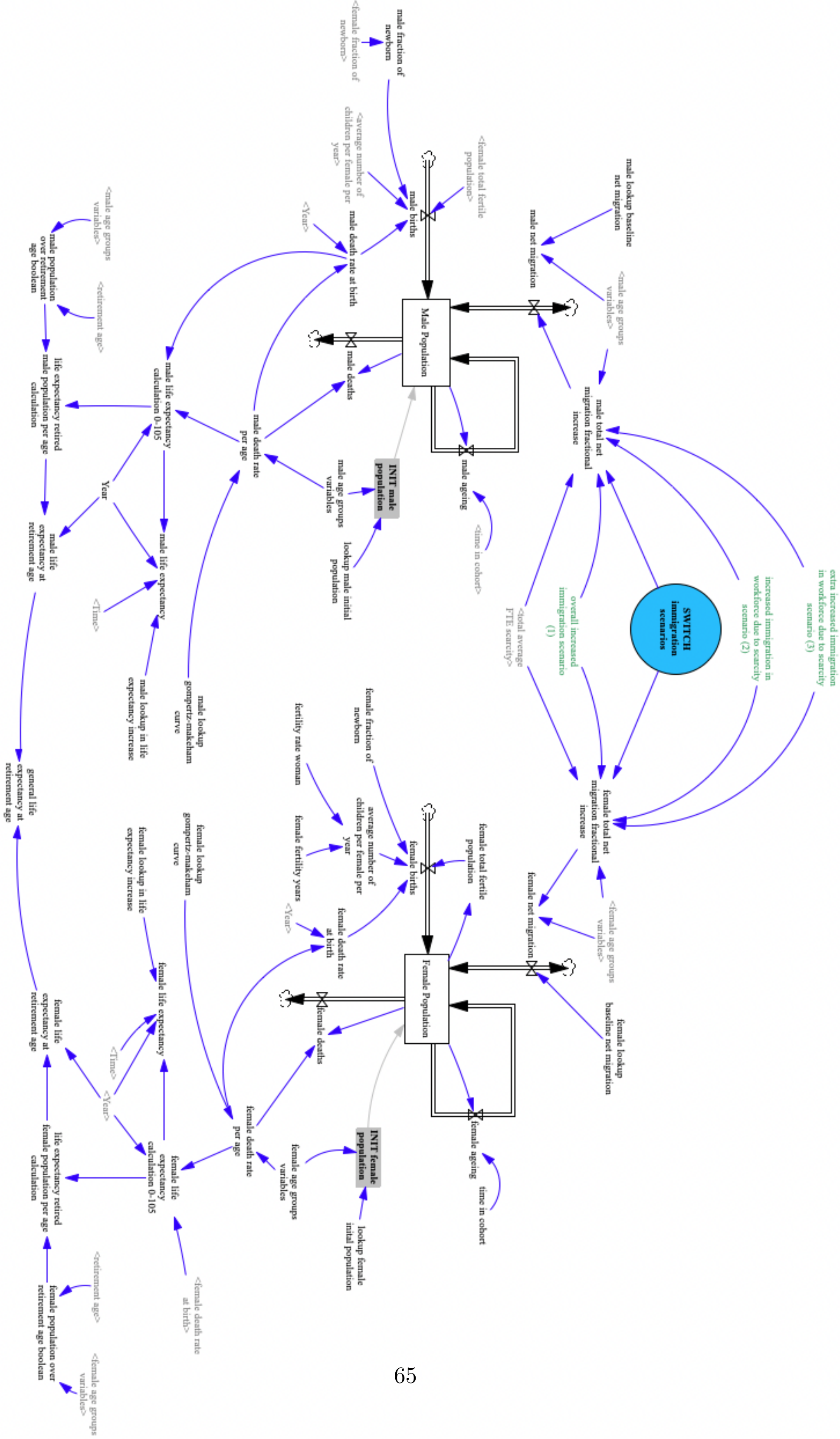


Figure 22 Population main part

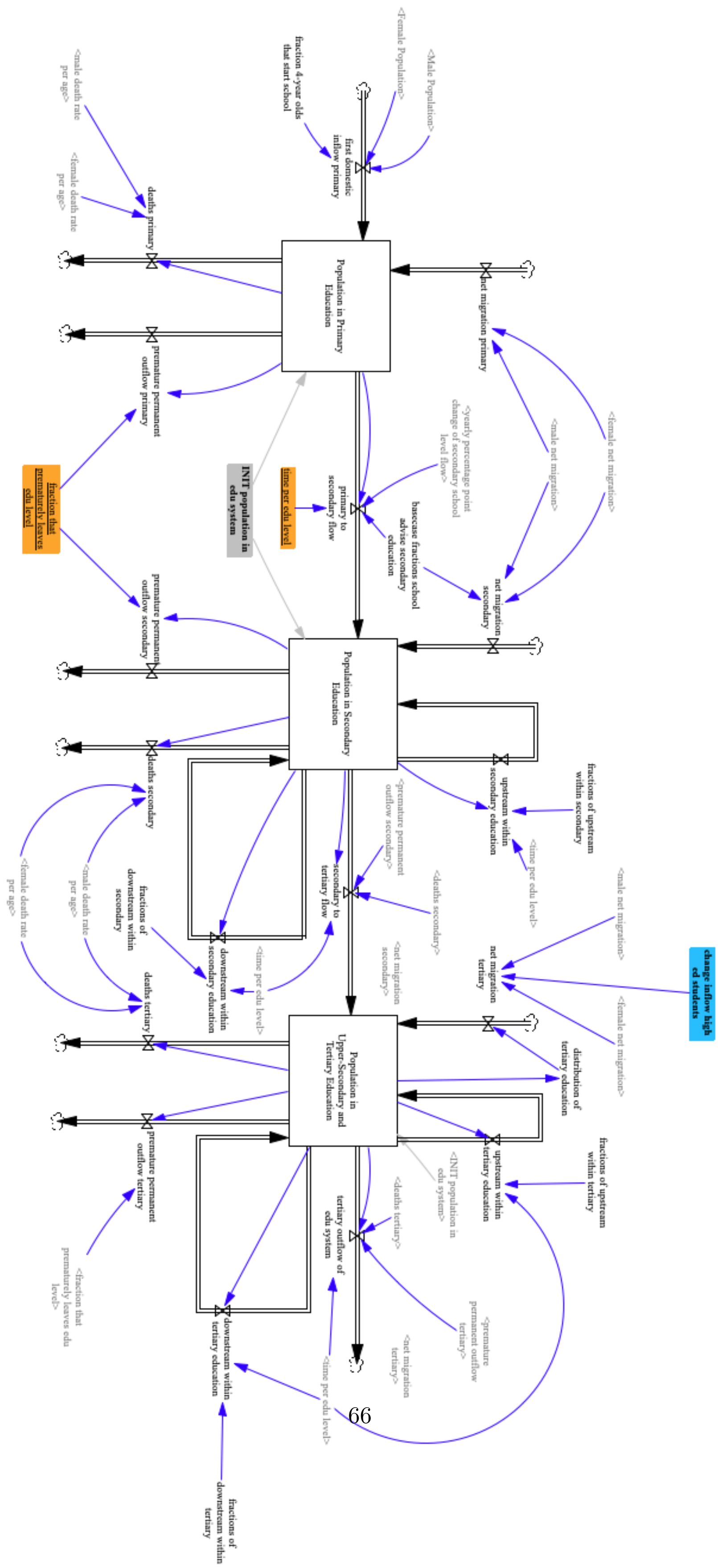


Figure 23 Education main part

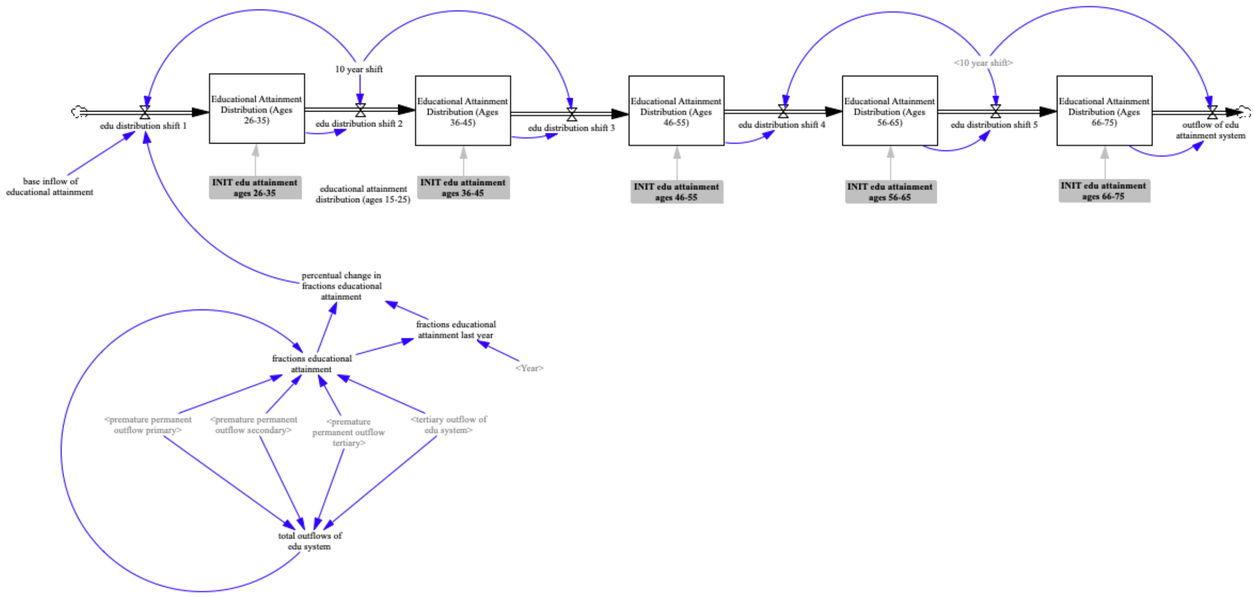


Figure 24 Educational attainment flow

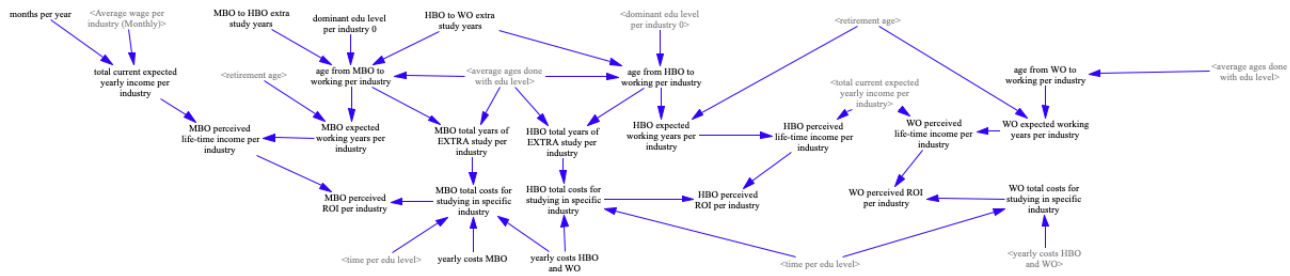


Figure 25 ROI of industries

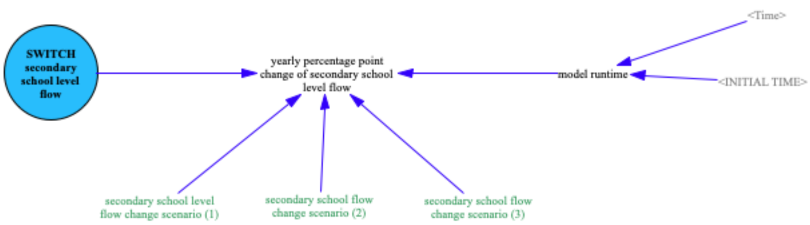


Figure 26 Edu policies

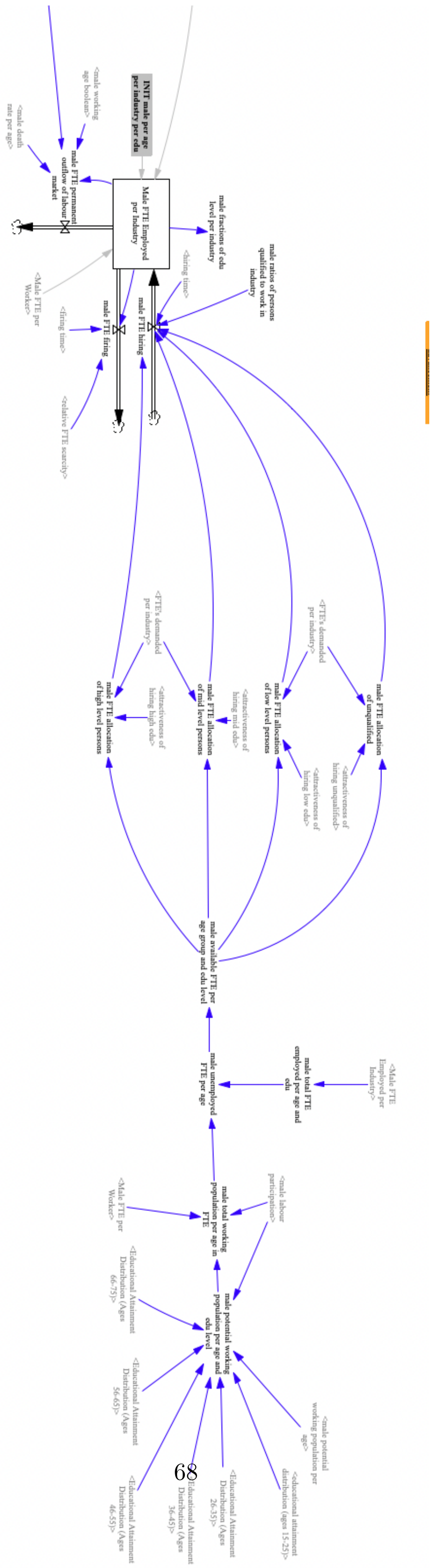


Figure 27 Males FTE employed

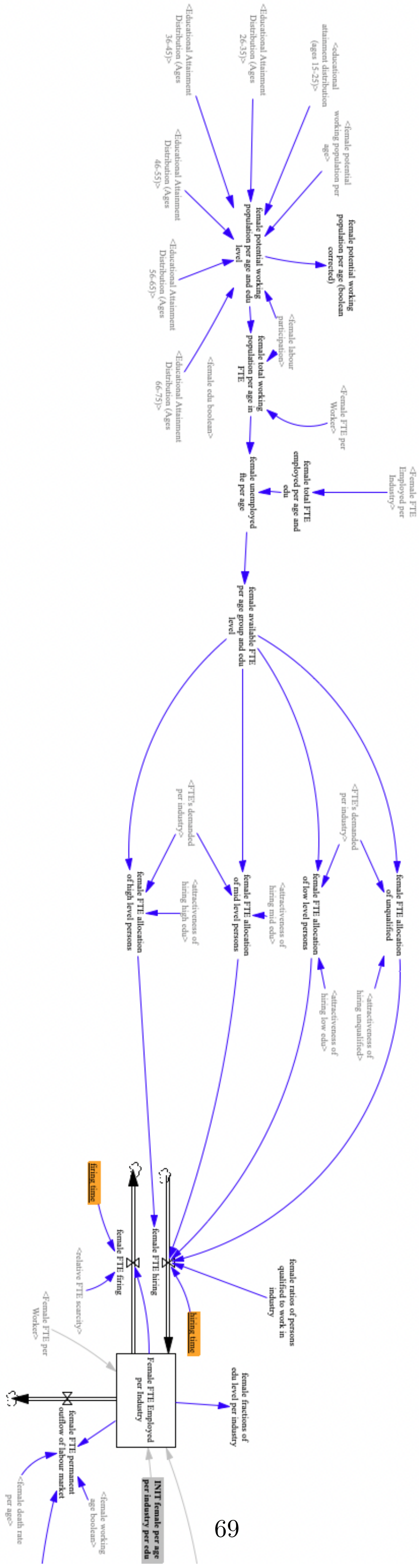


Figure 28 Females FTE employed

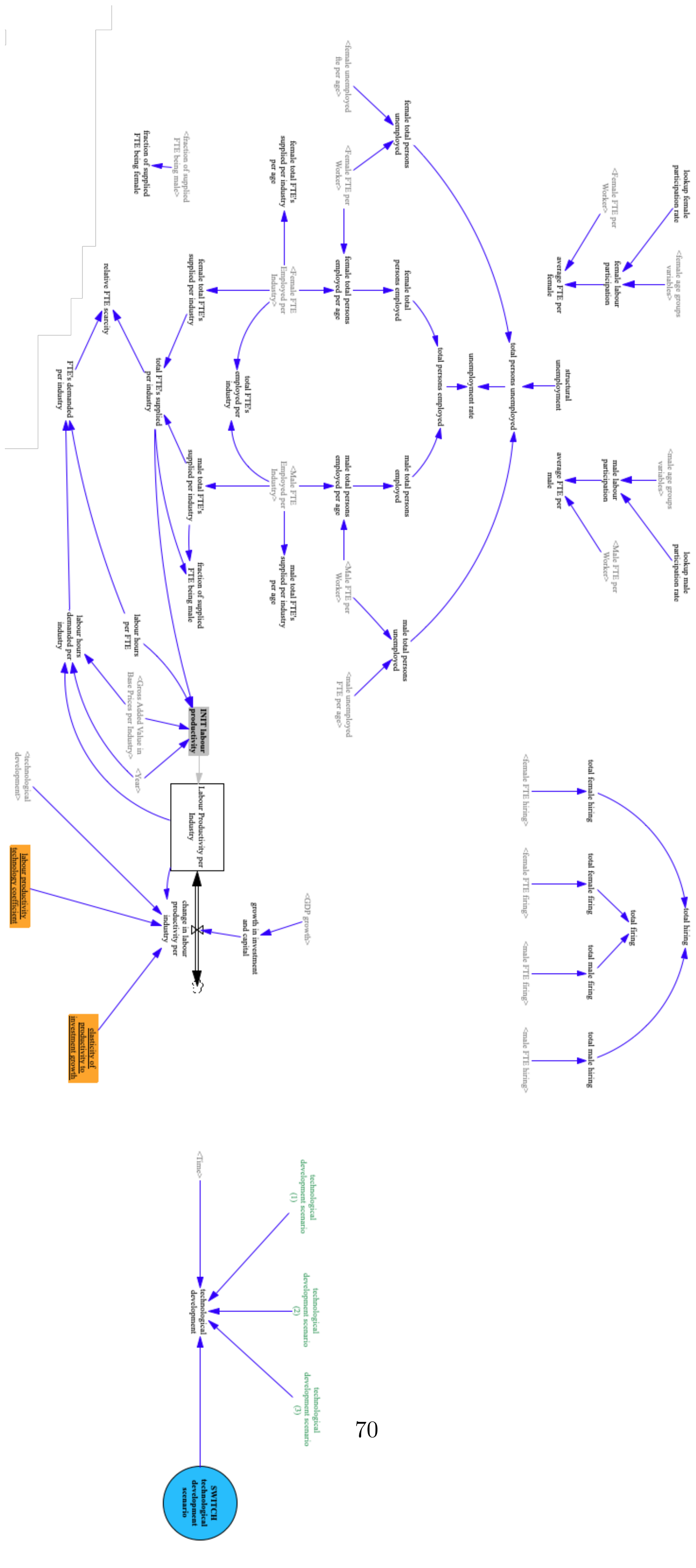


Figure 29 Labour market calculations

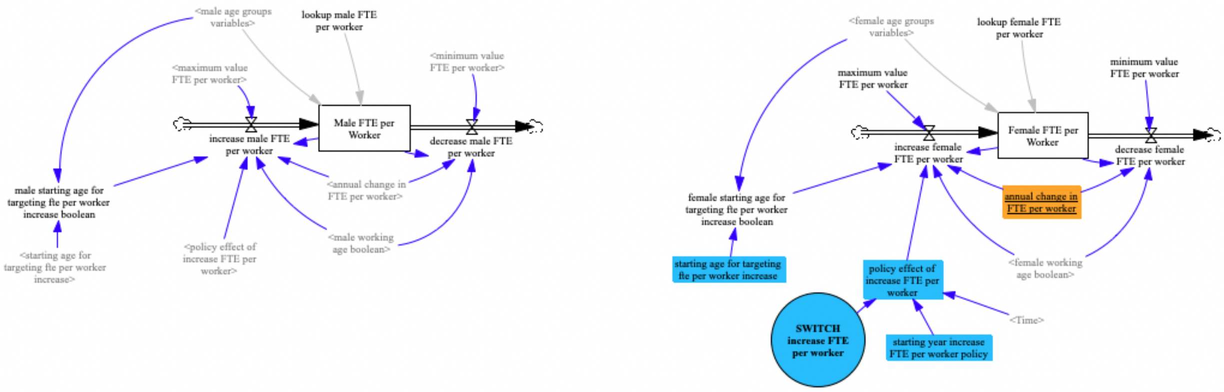


Figure 30 FTE policy

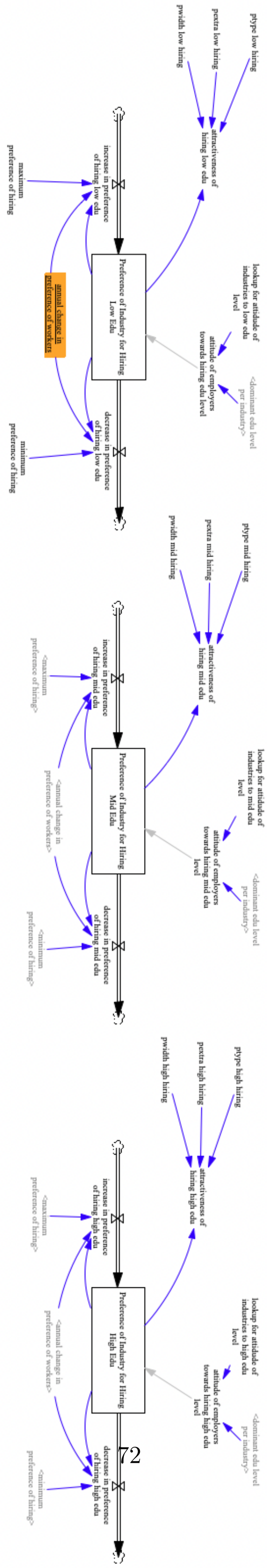


Figure 31 Preferences of workers

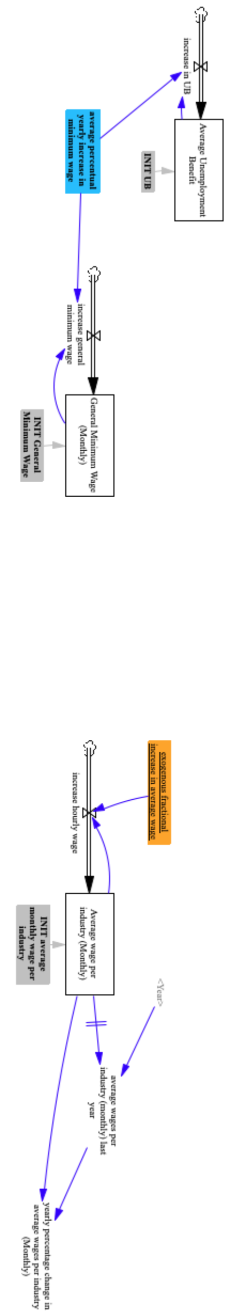


Figure 32 Wage setting in the labour market

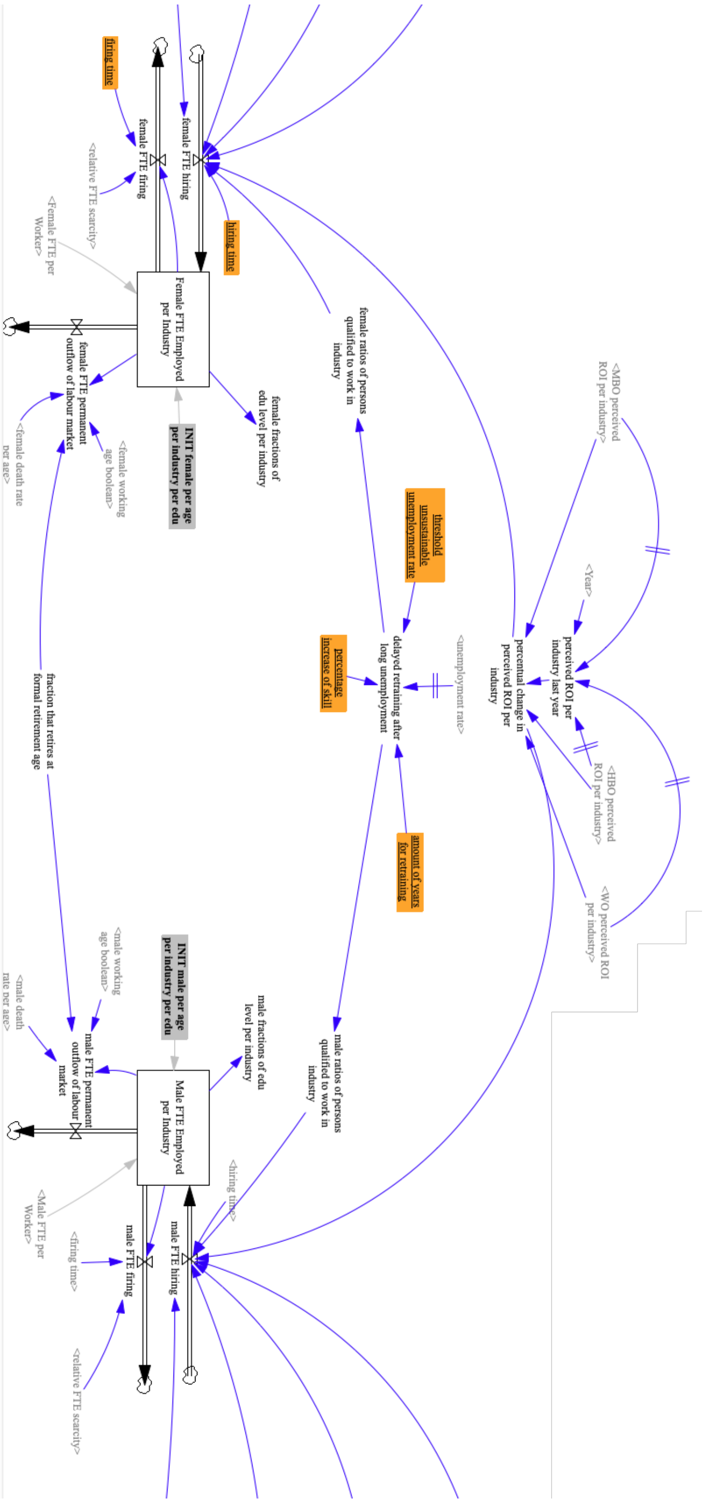


Figure 33 Extra influences on labour market

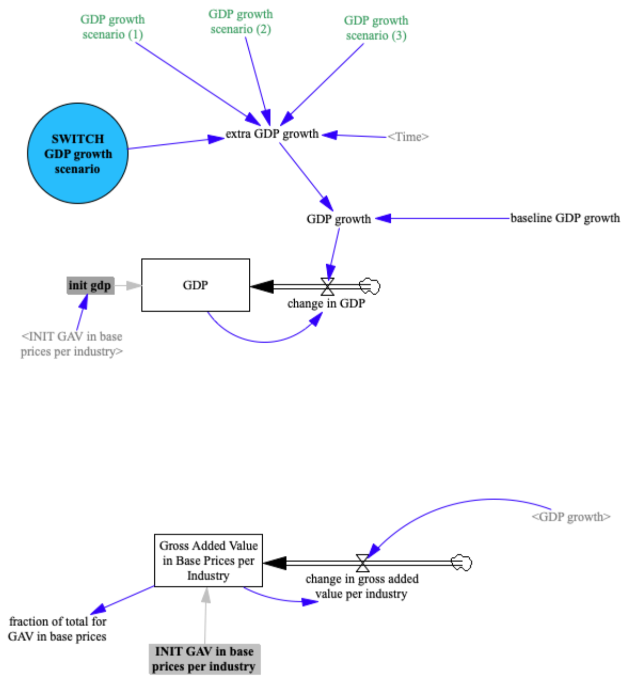


Figure 34 Economy

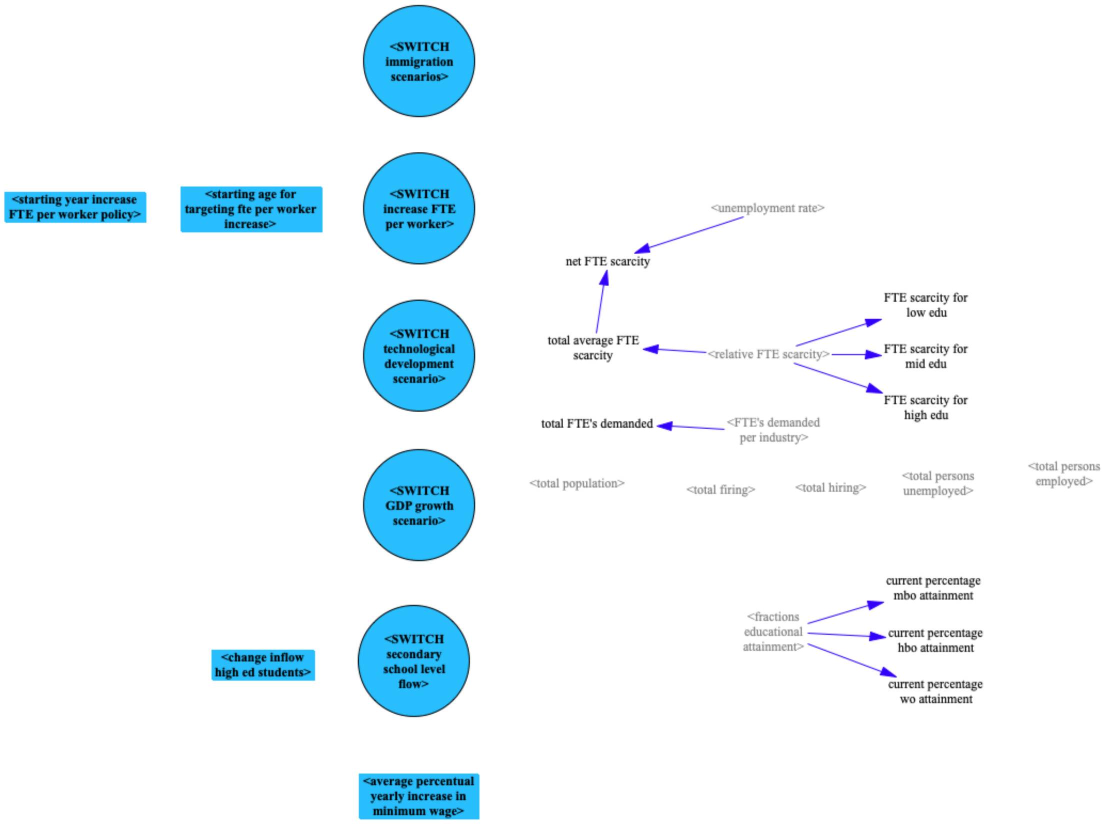


Figure 35 Switches and KPI's

Appendix D — Supplementary results

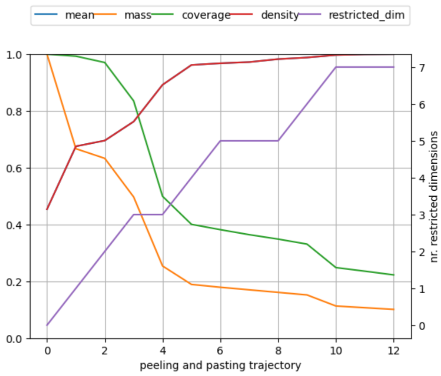


Figure 36 PRIM - peeling trajectory net fte scarcity

	coverage	density	id	mass	mean	res_dim
0	1.000000	0.453733	0	1.000	0.453733	0
1	0.992947	0.675462	1	0.667	0.675462	1
2	0.970614	0.695735	2	0.633	0.695735	2
3	0.835292	0.762575	3	0.497	0.762575	3
4	0.499412	0.892126	4	0.254	0.892126	3
5	0.400676	0.961905	5	0.189	0.961905	4
6	0.381869	0.967970	6	0.179	0.967970	5
7	0.364237	0.972157	7	0.170	0.972157	5
8	0.348663	0.982609	8	0.161	0.982609	5
9	0.331031	0.988158	9	0.152	0.988158	6
10	0.248310	0.997050	10	0.113	0.997050	7
11	0.235674	0.999377	11	0.107	0.999377	7
12	0.222598	1.000000	12	0.101	1.000000	7

Figure 37 PRIM - peeling trajectory net fte scarcity table

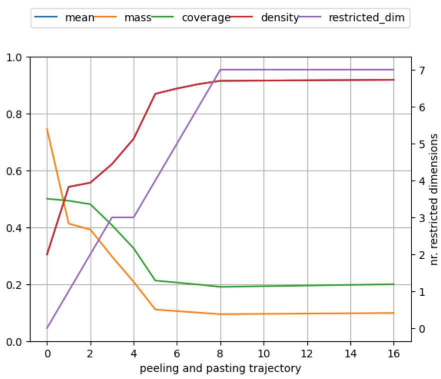


Figure 38 PRIM - second peeling trajectory net fte scarcity

	coverage	density	id	mass	mean	res_dim
0	0.500588	0.304468	0	0.746000	0.304468	0
1	0.493535	0.542211	1	0.413000	0.542211	1
2	0.481193	0.556973	2	0.392000	0.556973	2
3	0.407435	0.622447	3	0.297000	0.622447	3
4	0.326183	0.711538	4	0.208000	0.711538	3
5	0.212607	0.869591	5	0.110933	0.869591	4
6	0.205701	0.888325	6	0.105067	0.888325	5
7	0.198648	0.903743	7	0.099733	0.903743	6
8	0.190420	0.915254	8	0.094400	0.915254	7
9	0.191596	0.915730	9	0.094933	0.915730	7
10	0.192771	0.916201	10	0.095467	0.916201	7
11	0.193947	0.916667	11	0.096000	0.916667	7
12	0.195122	0.917127	12	0.096533	0.917127	7
13	0.196297	0.917582	13	0.097067	0.917582	7
14	0.197473	0.918033	14	0.097600	0.918033	7
15	0.198648	0.918478	15	0.098133	0.918478	7
16	0.199824	0.918919	16	0.098667	0.918919	7

Figure 39 PRIM - second peeling trajectory net fte scarcity table

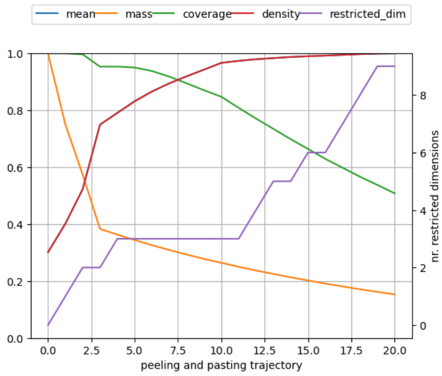


Figure 40 PRIM - peeling trajectory unemployment

	coverage	density	id	mass	mean	res_dim
0	1.000000	0.301067	0	1.000	0.301067	0
1	1.000000	0.401422	1	0.750	0.401422	1
2	0.996457	0.522648	2	0.574	0.522648	2
3	0.953277	0.749347	3	0.383	0.749347	2
4	0.953277	0.790634	4	0.363	0.790634	3
5	0.950177	0.831589	5	0.344	0.831589	3
6	0.937777	0.866053	6	0.326	0.866053	3
7	0.918069	0.894498	7	0.309	0.894498	3
8	0.894597	0.919226	8	0.293	0.919226	3
9	0.870461	0.942686	9	0.278	0.942686	3
10	0.847431	0.966414	10	0.264	0.966414	3
11	0.808459	0.973600	11	0.250	0.973600	3
12	0.770815	0.979184	12	0.237	0.979184	4
13	0.734721	0.983111	13	0.225	0.983111	5
14	0.698406	0.987167	14	0.213	0.987167	5
15	0.664083	0.989769	15	0.202	0.989769	6
16	0.629318	0.991972	16	0.191	0.991972	6
17	0.597874	0.994475	17	0.181	0.994475	7
18	0.566430	0.997271	18	0.171	0.997271	8
19	0.537644	0.999177	19	0.162	0.999177	9
20	0.508193	1.000000	20	0.153	1.000000	9

Figure 41 PRIM - peeling trajectory unemployment table