DIPLOMARBEIT

Age structured optimal control

An application to optimal human capital investment

Ausgeführt am Institut für
Stochastik und Wirtschaftsmathematik
der Technischen Universität Wien

unter der Anleitung von
Univ.Prof. Dipl.-Ing. Dr.techn. Alexia Fürnkranz-Prskawetz

durch

Michael Freiberger BSc
Rabengasse 13/11, 1030 Wien

14. Februar 2017

Abstract

In this thesis we analyse macro level optimal control models concerning education policies. The individuals in the economy thereby are allowed to differ in age and skill, with age as a continuous variable and the skill level being influenced by the education, which the individuals receive. Using this framework we derive the social optimal age specific education rates and their development over time. We further examine the effects of combinations of different education types, the influence of demographic changes and the impact of varying production technologies represented by different elasticities of substitution and productivities of the workers.

Our analysis shows, that a growing or decreasing population has a significant impact on the optimal education policies if workers are not perfectly substitutable. We also derive that the optimal age specific education rates strongly depend not only on the costs of different education types, but also on the productivities of workers of different skill levels. Further we introduce a more detailed educational structure and additional possibilities to upgrade skills. Our results indicate that under specific parameter constellations it is optimal to also utilize these augmented policy options.

Preface

I would like to begin this thesis by thanking all the people, who supported me during the creation of this work. First of all I want to thank Prof. Alexia Fürnkranz-Prskawetz, who introduced me to this interesting and innovative topic, and also always gave constructive-minded input and suggestions for improvement. Next I want to express my gratitude to the research group of Michael Kuhn and Ivan Frankovic at the Vienna Institute for Demography, who provided me their optimization algorithm. Without this code as a basis and the time Ivan took to explain it for me, all the wonderful simulations and colourful pictures in this thesis would not have been possible. I also want to thank my family, who always supported me and offered their assistance. Finally I want to express my special gratitude to my girlfriend Eva, who inexplicably somehow was able to cope with my highly stressed attitude and strange moods in the last several months.

Contents

1 | Introduction

Thinking about classic dynamic macroeconomic models, the Solow-growth model or the Ramsey-Cass-Koopmans model are likely to come in mind. A main assumption in these and many similar models is the homogeneity of all individuals populating the economy. There are several reasons behind this premise. First of all, especially in long-run economic growth models the objective is to maximize the utility of a infinitely-lived individual (or household). This is the simplest way to model altruistic motives in an economy since the individuals consider the entire future, when making their decisions. Additionally, it is in general assumed, that all individuals are born under the same conditions and face the same problem of utility maximization. In this case it is only logical, that all individuals therefore make the same decisions and act identically. This not only makes it possible to focus on one representative individual, but also simplifies the analytical treatment of the model.

After all we know that people do not start their life under the same conditions and hence also act different. Hence, in reality people are heterogeneous. The most basic differences people exhibit might be

- their gender,
- their endowment,
- their birth date respectively their age,
- their skills or education.

The most obvious distinction by gender is of interest in models concerning fertility decision and decisions within households, but in more general economic models the sex often plays a subordinate role.[1] The term endowment meanwhile covers several different aspects. It could either describe the economic endowment an individual receives from his parents containing for example financial wealth or the education a person gets provided, or the genetic endowments regarding for example physical or cognitive abilities.

In the following thesis we will nevertheless focus on the latter two aspects. The birth date especially plays a role if one drops the assumptions of infinitely-lived individuals. Some people are born at different dates and this leads to a population, that can be distinguished by age. That the share respectively the size of each age-group in reality is a dynamic variable and changes over time, can be seen in figure 1.1. The diagrams illustrate the population pyramids

[1]From the standpoint of total gender equality the sex of a person also must not have an effect on the alternatives he or she can choose from or the decision he/she makes.

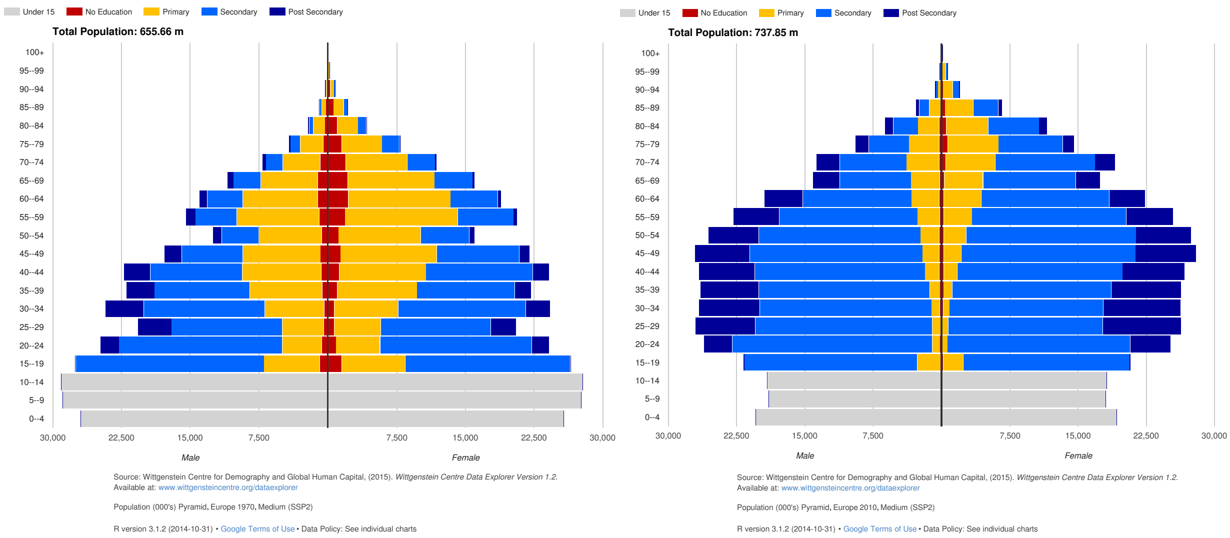


Figure 1.1: Population pyramids for Europe distinguished by education level for 1970 (left plot) and 2010 (right plot) (source: Wittgenstein centre (2015))

for Europe for the years 1970 (left panel) and 2010 (right panel). The length of each bar represents the size of the corresponding age-group (left bars for men and right bars for female) and we see, that they have significantly changed between 1970 and 2010. Roughly said in 2010 there are less people in the age-groups between 0 and 19 then in 1970, but meanwhile the other age-groups have increased in number. One can deduce several other facts from these graphs (increasing life expectancy, decreasing fertility,...), but we will now turn to the differences in skills.

The skill set of a person is often measured in the education he or she has attained. Figure 1.1 provides further information about the highest education level obtained[2]. The colours of each bar illustrate the number of people of each age-group having completed no education (red), primary education (yellow), secondary education (light blue) or post-secondary education (dark blue). Comparing the two plots we see that while in 1970 people with no education formed significant shares at all ages, in 2010 nearly everybody above the age of 15 has completed at least primary education. Also the number of people with secondary or post-secondary education has increased by a significant amount in those 40 years.

Overall we have seen, that age and education are very important factors describing a population and that individuals are highly heterogeneous with respect to these characteristics. Since the composition of the population regarding age as well as education changes over time, they should be explicitly included in economic models. The first theoretical models to include age heterogeneity were the Overlapping-Generations models (OLG-models). In the early versions it was assumed, that people would live two time periods, in one they are young in the other they are old. At each time point the economy is then populated by two generations born at two different time points. While the younger generation becomes old, when heading to the next time period, the older existing generation dies at the end of the current period. As this modelling wasn't able to reflect reality good enough (especially the economic dependency in young

[2]See chapter 3 for further information about the ISCED education levels.

and old age, while generating working income in the years in between), OLG-models were soon extended to three or more general n age-groups. Then again the solutions of these models often depended on the number of age-groups, thereby reducing the validity of the model. The next step in the generalisation of age heterogeneity in these models was the introduction of age as a continuous variable and finally lead to age-structured control models. These age-structured control models exhibit the benefits of continuous models (compared to discrete models) and higher generality, but on the other hands enhance the degree of mathematical complexity. As the standard theory about optimal control models is not applicable anymore, they also require the derivation of a new theory characterising new optimality conditions.

All these heterogeneities have an impact on the labour market and the education policies. As far as production is concerned people with higher skills are likely to exhibit higher productivities, which can also depend on the age of a person. We also have to bear in mind, that the rates, at which people of different ages or skills can be substituted, can differ. Regarding education policies one should respect that different kinds of educations have varying rates of success and also cause diverse costs. Again also age plays an important rule since younger people might have higher learning abilities and also profit more from their education, because they have a longer remaining time in the labour market. In all upcoming models we analyse the social optimal education policies through setting up the problem of a social planer. Therefore we also assume full employment and the population development is exogenously given. The main questions we want to answer within this thesis are:

- At which ages should education or training be applied in the social optimum.
- How do different elasticities of substitution effect the social optimal education policies? Does the case of perfect substitutability lead to special results?
- What are the consequences of demographic changes for an optimal education policy and how does the social planer react to them?
- How do different cost structures and skill productivities affect the decisions of the social planer? Which education paths get into focus and which get neglected.
- To which extend are newly introduced education possibilities used, or are they even favourable at all?

The thesis is structured as follows. Chapter 2 revisits the model of Prskawetz et al. (2012) with two skill groups, which will be used as a base model for further extensions. In Chapter 3 an additional skill level is introduced and the results are compared to chapter 2. In chapter 4 the available education tools for the social planer are augmented and we examine the impacts on the solutions. In Appendix B the derivation of the optimality conditions for each model are presented and chapter 5 shortly summarises the theory behind these conditions from Feichtinger et al. (2003). Appendix A gives further insight in CES production functions and lastly in appendix C the basic structure of the numerical optimization algorithm for the numerical solutions of the models is explained.

2 | Optimal human capital accumulation

In this chapter I will review the paper of Prskawetz et al. (2012) and support their results with my own simulations.

2.1 The model

In the following we will analyse an economy changing over time t within the time horizon $[0, T]$. T is assumed to be relatively large and indicates the end of the planning horizon. The individuals populating the economy differ with respect to two characteristics: Their age and their skills.

- The age of a person is denoted by s and ranges between 0 and ω. However in this model age should not directly be interpreted as biological age, but rather as time spent in the labour market. Individuals might start their working career at their biological age of 20 (but with $s = 0$) and leave it at age 65 with $s = 45$ years of working experience. Nevertheless we will refer to s as age, to keep the terminology relatively simple.
- Regarding the skills, we distinguish between two groups: the low-skilled workers $L(t, s)$ (at time t with age s) and the high-skilled workers $H(t, s)$.

Interchange between the two skill groups is based on skill-depreciation and learning-by-doing, which both cannot be controlled, and education which can be changed through governmental actions. Figure 2.1 summarises the different flows between the two groups.

The government can increase the transition of low-skilled to high-skilled workers by raising the educational efforts, i.e. increasing the education rate $u(t, s)$. The effectiveness of these efforts may depend on time and especially on age. The learning abilities of younger people are likely to be higher than those of older individuals regardless of their higher life/working experience. As a result in the model the education rate $u(t, s)$ is going to be multiplied with the effectiveness parameter $l(t, s)$. In addition to these education rates, costless "learning by doing" skill improvement takes place at the rate $e(s)$, which is assumed to be independent of time t. The only flow from high to low-skilled workers is described by skill depreciation at rate $\delta(t, s)$, that leads to people losing their high qualification and becoming low-skilled again. This effect might be due to forgetting or the obsolescence of the acquired skills resulting from technological progress.

Formally the flow dynamics between high and low-skilled workers can be described by the

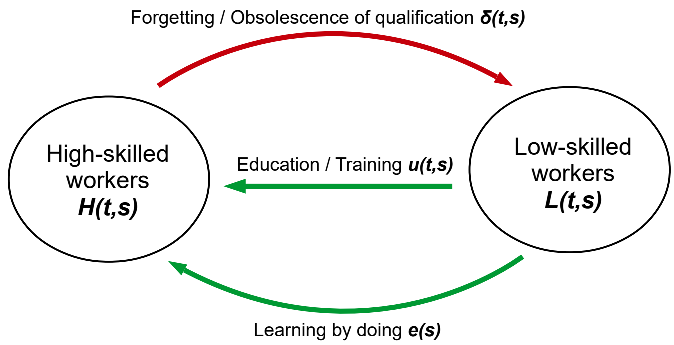


Figure 2.1: Transition flows between the two skill groups

following partial differential equations.

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) L(t,s) = \delta(t,s)H(t,s) - e(s)L(t,s) - l(t,s)u(t,s)L(t,s) \tag{2.1}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) H(t,s) = -\delta(t,s)H(t,s) + e(s)L(t,s) + l(t,s)u(t,s)L(t,s) \tag{2.2}$$

The operator $\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) L(t,s) = \lim_{\varepsilon \to 0} \frac{L(t+\varepsilon,s+\varepsilon)-L(t,s)}{\varepsilon}$ describes the change of the number of low-skilled workers of a cohort born at time-point $(t-s)$. For the solution of a PDE (partial differential equation) one generally needs boundary conditions. In our model we are given the initial composition of the population at the beginning of the time horizon $t=0$ and the number of individuals entering the economy with age 0 at each time point t, i.e.

$$L(0,s) = L_b(s), \qquad H(0,s) = H_b(s) \qquad \forall s \in [0,\omega]$$

$$L(t,0) = L_0(t), \qquad H(t,0) = H_0(t) \qquad \forall t \in (0,T]$$

with those functions being exogenously given. Since the system is closed, the inflows of high-skilled workers of one age-group s at time t corresponds to the outflows of low-skilled workers of the same age-group and vice versa. Hence also the population size of each cohort stays constant over time

$$L(t,s) + H(t,s) = L_0(t-s) + H_0(t-s) =: N_0(t-s) \qquad \forall t \geq s$$

As a result we would be able to eliminate $L(t,s)$ or $H(t,s)$ from the model, but we keep both variables to make the model and results better understandable.

A visualisation of the mathematical model formulation can be seen in figure 2.2, which illustrates the model in a Lexis diagram. Each cohort is represented by one of the lines with slope one and the PDEs (2.1) and (2.2) describe the evolution of the low or resp. high-skilled workers along these characteristic lines with the main driving forces $e(s)$, $u(t,s)$ and $\delta(t,s)$ (green). On the left border of the time-age-space we see the boundary condition for the structure of the population at $t=0$ (red) and on the bottom the condition for the individuals entering with age $s=0$ (blue).

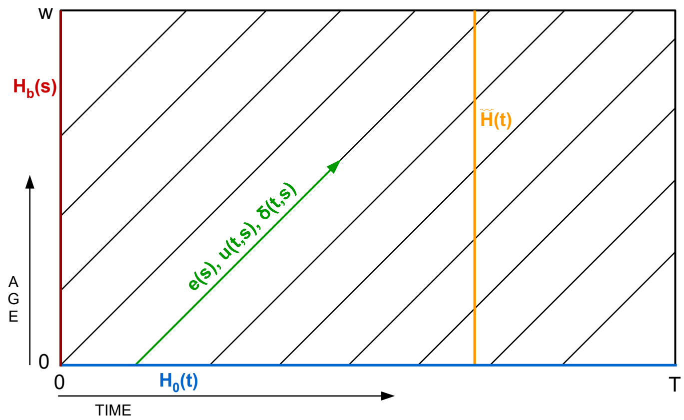


Figure 2.2: Lexis diagram with boundary constraints

We further assume full employment in our model, implying that all high- and low-skilled workers contribute to final production. The final production function is of the CES-type[1] using sub-aggregates of high and low-skilled workers $\tilde{H}(t)$ and $\tilde{L}(t)$ as inputs.

$$Y(t) = \left(\theta_L(t)\tilde{L}(t)^\rho + \theta_H(t)\tilde{H}(t)^\rho\right)^{1/\rho}$$

We allow for imperfect substitution between low and high skilled workers. For $\rho \in (-\infty, 1]$ it holds that $\frac{1}{1-\rho}$ is the partial elasticity of substitution between low- and high-skilled workers. In the case of $\rho = 1$ we obtain the case of perfect substitution, which means that low- and high-skilled workers can be exchanged at a fixed rate for any combination of the two.[2] $\theta_L(t)$ and $\theta_H(t)$ are the productivity parameters of the different skilled workers, with the productivity of highly educated people being higher.

The sub-aggregates are themselves of the CES-type and take the following form

$$\tilde{L}(t) = \left(\int_0^\omega \pi_L(s)L(t,s)^{\lambda_L}ds\right)^{1/\lambda_L}$$

$$\tilde{H}(t) = \left(\int_0^\omega \pi_H(s)H(t,s)^{\lambda_H}ds\right)^{1/\lambda_H}$$

Again imperfect substitution between age-groups is possible with $\frac{1}{1-\lambda_i}$ being the partial elasticity of substitution between the different age groups of the same skill. We also have $\pi_i(s)$

[1]CES = Constant elasticity of substitution

[2]The case $\rho \geq 0$ seems to be the more relevant case, since it results in low and high-skilled individuals being substitutes. $\rho \leq 0$ would lead to complementary inputs, which are more difficult to justify economically.

as a productivity parameter for different age-groups.[3] These aggregations are represented by the orange line in figure 2.2. One should note that this production function also has a downside. Despite the possibility of imperfect substitution between ages, there is no impact of the age-difference between two workers. For example a relatively young worker according to this production function can be as good substituted with another young worker of a similar age-group as with a worker, who is e.g. 30 years older. This may limit the realism of this production function, especially for the more physical work intensive low skilled sector, but nevertheless the production function is kept highly general with respect to various other aspects. A more detailed discussion about these kinds of production functions can be found in appendix A.2. There I also present two possible adaptation of the basic CES-production function, which can be made to adapt for the flaw mentioned above.

However educational efforts by the government are not costless. We assume that the per capita costs of education depend on age s and on the desired education rate, and are modelled using the function $p(s, u(t, s))$. This cost function is further discussed later in this section and the exact functional form for the numerical solutions is introduced in table 2.1. Finally the total educational costs $P(t)$ at every time-point can be calculated by the integral

$$P(t) = \int_0^\omega p(s, u(t, s))L(t, s)ds$$

Finally the social planer maximizes the aggregated production minus the aggregated education costs, i.e. $Y(t) - P(t)$. This difference is discounted by the term e^{-rt} with $r \geq 0$ being the exogenously given constant discount rate/ interest rate.[4]

Now we can summarize our model in the following problem formulation

Problem 1

The social planer solves the age-structured optimal control problem below:

$$\max_{u(t,s)} \int_0^T e^{-rt} \left[\left(\theta_L(t)\tilde{L}(t)^\rho + \theta_H(t)\tilde{H}(t)^\rho \right)^{1/\rho} - P(t) \right] dt$$

$$\text{s.t:} \quad \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s} \right) L(t, s) = \delta(t, s)H(t, s) - e(s)L(t, s) - l(t, s)u(t, s)L(t, s)$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s} \right) H(t, s) = -\delta(t, s)H(t, s) + e(s)L(t, s) + l(t, s)u(t, s)L(t, s)$$

$$L(0, s) = L_b(s), \qquad H(0, s) = H_b(s) \qquad \forall s \in [0, \omega]$$

$$L(t, 0) = L_0(t), \qquad H(t, 0) = L_0(t) \qquad \forall t \in (0, T]$$

[3]For the numerical results we are assuming $\pi_i(s)$ to be hump-shaped with respect to age. The parameters exhibit a relatively later peak in productivity for higher educated individuals.

[4]Since our model does not contain physical capital as a production input, we can take the interest rate r as exogenously given and identical to the time preference rate.

$$\tilde{L}(t) = \left(\int_0^\omega \pi_L(s) L(t,s)^{\lambda_L} ds \right)^{1/\lambda_L}$$

$$\tilde{H}(t) = \left(\int_0^\omega \pi_H(s) H(t,s)^{\lambda_H} ds \right)^{1/\lambda_H}$$

$$P(t) = \int_0^\omega p(s, u(t,s)) L(t,s) ds$$

$$u(t,s) \geq 0$$

Before we close the model description, some remarks should be made:

- The operator $\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)$ should be seen as one single operator and not as the sum of the partial derivatives. The latter one would lead to the necessity of the states $L(t,s)$ and $H(t,s)$ to be differentiable with respect to both time and age, which can happen not to be fulfilled, e.g. if there are discontinuities in the boundary conditions $L_0(t)$ and $H_0(t)$. Taking it as one operator only needs the differentiability along the characteristic lines, which results automatically from the model formulation.

- There is no separate educational sector in this model and individuals do not have to allocate their time between working in production and following educational activities. This means that the opportunity costs for education efforts by the social planer should be somehow included in the cost-function. In the case of perfect substitutability between skills ($\rho = 1$) and age in the low skill sector ($\lambda_L = 1$) one possibility to model the cost function would be

$$p(s, u(t,s)) = \underbrace{\tilde{p}(s, u(t,s))}_{\text{direct costs}} + \underbrace{\theta_L(t)\pi_L(s)u(t,s)}_{\text{opportunity costs}}$$

  In this case, individuals in education do not contribute to production and the losses in production $\theta_L(t)\pi_L(s)u(t,s)$ can be added to the direct cost of education resulting in the total cost. This means, that under the given parameter constellation ($\rho = 1$ and $\lambda_L = 1$) the general model formulation also covers the case of general secondary and tertiary education. On the other hand for other parameters this is not true and education should be strictly seen as on the job training.

  However in the further progress we will soon make the assumption, that the costs are independent of age, to keep the model analytically traceable and therefore only implement the direct costs.

- The cost function $p(s,u)$ is assumed to be strictly convex in u. This not only has mathematical advantages for the solution later on, but should also be the intuitive structure: Starting with a low desired education rate, training should be focused on the most talented (highest potential) individuals and should therefore be relatively cheap. Hence, additional individuals in training to raise the education rate should cost more than the average before. This leads exactly to a convex cost structure for the educational efforts.

2.2 Derivation of the optimality conditions

For the solution of this model we will use the theory presented in chapter 5. These optimality conditions are derived and proven in Feichtinger et al. (2003).

To be able to apply this theory, we first need to make assumptions about the exogenously given functions, that are sufficient for our model to fulfil the standing assumption in Feichtinger et al. (2003).

Assumption 1 (Standing assumptions)

The following assumptions shall hold for the rest of this chapter

- The initial populations $L_0(t)$ and $H_0(t)$ are piece-wise continuous.
- All other exogenously given functions are continuous.
- All derivatives appearing in the calculations are themselves continuous.
- The cost function $p(s, u)$ is monotonically increasing and strictly convex in u.
- All population parameters ($L_0(t)$, $L_b(s)$, $H_0(t)$ and $H_b(s)$) and all productivity coefficients ($\theta_L(t)$, $\theta_H(t)$, $\pi_L(s)$ and $\pi_H(s)$) are strictly positive.

The derivation of the optimality conditions takes some space-consuming calculations and can be found in appendix B.1. The following proposition 1 summarizes the final optimality conditions resulting from these calculations.

Proposition 1

Let $(L, H, \widetilde{L}, \widetilde{H}, P, u)$ be an optimal solution of problem 1. Then the partial differential equation (2.3) for $\Delta(t, s)$, the difference between the co-states/shadow-prices of high- and low-skilled workers, has a unique solution

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\Delta(t, s) = (r + e(s) + \delta(t, s) + l(t, s)u(t, s))\Delta(t, s) - \\ - p(s, u(t, s)) - (f_H(t, s) - f_L(t, s)) \tag{2.3}$$

$$\Delta(T, s) = 0, \qquad \Delta(t, \omega) = 0$$

with

$$f_L(t, s) = Y(t)^{1-\rho}\theta_L(t)\pi_L(s)\tilde{L}(t)^{\rho-\lambda_L}L(t, s)^{\lambda_L - 1} \qquad \text{and}$$

$$f_H(t, s) = Y(t)^{1-\rho}\theta_H(t)\pi_H(s)\tilde{H}(t)^{\rho-\lambda_H}H(t, s)^{\lambda_H - 1}$$

And for almost every $(t,s) \in [0,T] \times [0,\omega]$ *it holds that*

$$u(t,s) = \arg\max_{u \geq 0} \Big(\Delta(t,s) l(t,s) u - p(s,u) \Big) \tag{2.4}$$

We have been able to eliminate the co-state variables $\mu_L(t,s)$ and $\mu_H(t,s)$ corresponding to $L(t,s)$ and $H(t,s)$ since the optimality condition for $u(t,s)$ only depends on the difference of the two and subtracting the PDEs for the co-states results in the equation (2.3) for $\Delta(t,s) = \mu_H(t,s) - \mu_L(t,s)$. Moreover $f_H(t,s)$ equates the marginal productivity of a high skilled worker as it holds that $f_H(t,s) = \frac{\partial Y(t)}{\partial L(t,s)}$ (analogue for $f_L(t,s)$)[5]. The difference $(f_H(t,s) - f_L(t,s))$ therefore represents the marginal gain in productivity by a worker becoming high skilled and it is not surprising, that this terms effects the dynamic of $\Delta(t,s)$

As we assumed that the cost function $p(s,u)$ is monotonically increasing and strictly convex, the optimality condition (2.4) can be written as a first order condition (for the case of an inner solution)

$$p'(s,u(t,s)) := \frac{\partial p(s,u(t,s))}{\partial u} = \Delta(t,s) l(t,s) \tag{2.5}$$

The cost function being strictly convex leads to the derivative $p'(s,u)$ being strictly monotonically increasing and therefore being invertible with respect to u. This means the optimal control can explicitly be expressed through (taking boundary solutions into account)

$$u(t,s) = p'^{-1}_{+}(s, \Delta(t,s) l(t,s)) := \begin{cases} p'^{-1}(s, \Delta(t,s) l(t,s)) & \text{if } \Delta(t,s) l(t,s) \geq p'(s,0) \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

For a clearer interpretation of the condition $\Delta(t,s) l(t,s) - p'(s,0) > 0$, we split up $\Delta(t,s)$ and obtain

$$\mu_H(t,s) l(t,s) > \mu_L(t,s) l(t,s) + p'(s,0)$$

On the left side we have the benefit of applying training to a marginal low-skilled worker. $\mu_H(t,s)$ is the shadow-price of a high-skilled worker, but the education is only efficient with "probability" $l(t,s)$. On the right hand side we see the shadow-value of a low-skilled worker (which is also only lost with probability $l(t,s)$) and the costs for the education of the first marginal worker. Intuitively this means if benefits of the first marginal worker becoming high-skilled are bigger than the costs, there are incentives to set a positive education rate $u(t,s)$.

Considering that the inverse function of a monotonically increasing function is also monotonically increasing, equation (2.6) shows that the bigger the differences in the shadow-values are, the higher the optimal education rate will be.

[5]Strictly seen the term $\frac{\partial Y(t)}{\partial L(t,s)}$ makes mathematically no sense, because age is continuous. The reason, why we still use the term and further information can be found in appendix A.2

2.3 Perfect substitutability

In this section we assume, that workers of all ages and skills are perfect substitutes ($\rho = \lambda_L = \lambda_H = 1$). For this case already relatively simple calculations lead to interesting results.

Proposition 2

In the case of perfect substitutability the following statements hold

(a) The optimal education rate does not depend on the initial and boundary data ($L_b(s)$, $H_b(s)$, $L_0(t)$ and $H_0(t)$).

(b) If all other data is time invariant, the optimal education rate is also time invariant on the interval $[0, T-\omega]$.

Proof. For the original proof see Prskawetz et al. (2012, p. 168).

(a) First we note that under the given conditions $f_i(t,s) = \theta_i(t)\pi_i(s)$ and therefore the PDE (2.3) is independent of $L(t,s)$ and $H(t,s)$ and hence also of the initial and boundary conditions. Bearing in mind that $u(t,s)$ depends only Δ, we have proven part (a).

(b) In this case all functions on the right hand side of equation (2.3) are independent of t and we can rewrite the PDE by using $z(\tau;a) = \Delta(\tau-a, \omega-a)$ with $\tau \in [0,T]$ being the location parameter and $a \in [0, \min\{\tau, \omega\}]$ as

$$-\frac{dz}{da} = \Big(r + e(\omega-a) + \delta(\omega-a) + l(\omega-a)p_+'^{-1}(s, zl(\omega-a))\Big)z -$$

$$-p(\omega-a, p_+'^{-1}(s, zl(\omega-a))) - (f_H(\omega-a) - f_L(\omega-a)$$

The right side is clearly independent of τ and in combination with $z(\tau,0) = \Delta(\tau,\omega) = 0$ it holds that $z(\tau;a) = z(a) \quad \forall \tau \in [0,T]$. For $t \in [0, T-\omega]$ we obtain

$$\Delta(t,s) = z(t+\omega-s, \omega-s) = z(\omega-s)$$

i.e. $\Delta(t,s)$ is independent of t. Keeping in mind that $u(t,s)$ only depends on $\Delta(t,s)$ we have proven the second part of the proposition.

■

As a result of this proposition we directly see, that demographic changes over time (increasing/decreasing number of workers entering the labour market or a change in their composition) do not effect the decisions of the social planer. The second part shows us, that our model is also feasible for analysing a potential steady state, if the parameters are all time invariant. In this case we obtain, that the solution before $T-\omega$ is the stationary solution.

2.3.1 Numerical benchmark results

I will now present the results of my own simulation. For the exact modus operandi in the numerical optimisation see appendix C. Table 2.1 shows the parameter values chosen for the simulation of the case of perfect substitutability, which I will further refer to as the "Benchmark case".[6]

Parameter	Description	Value	Range
T	time horizon	140	
ω	maximal time in the labour market	40	
r	time preference rate / interest rate	0.03	
ρ	elasticity of substitution across skill levels= $\frac{1}{1-\rho}$	1	
λ_L	elasticity of substitution across age= $\frac{1}{1-\lambda_L}$ for low-skilled	1	
λ_H	elasticity of substitution across age= $\frac{1}{1-\lambda_H}$ for high-skilled	1	
$\delta(t,s)$	skill depreciation rate	0	$\forall(t,s)$
$l(t,s)$	effectiveness of education rate	0.1386	$\forall(t,s)$
$e(s)$	learning by doing rate	0	$\forall s$
$\theta_L(t)$	low-skilled productivity	0.3	$\forall t$
$\theta_H(t)$	high-skilled productivity	0.7	$\forall t$
$\pi_L(s)$	low-skilled age productivity $\propto c_L exp(\frac{q_1^2}{(s-m_L)^2-q_2^2})$	see figure 2.3	
$\pi_H(s)$	high-skilled age productivity $\propto c_H exp(\frac{q_1^2}{(s-m_H)^2-q_2^2})$	see figure 2.3	
$L_0(t)$	number of low-skilled workers entering labour market	1000	$\forall t$
$H_0(t)$	number of high-skilled workers entering labour market	10^{-6}	$\forall t$
$p(u)$	cost-function for educational efforts	$p(u) = \frac{9}{520}u + \frac{3}{260}u^2$	

Table 2.1: Benchmark parameter set

In this benchmark parametrisation we eliminate skill depreciation and learning by doing and focus on the educational decisions. As already mentioned before we model the age-structured productivity inverse hump-shaped as illustrated in figure 2.3. The working ages of maximal productivity are defined by $m_L = 13$ and $m_H = 20$. The parameters are normalised so that

$$\int_0^\omega \pi_L(s)ds = \int_0^\omega \pi_H(s)ds = 1$$

and therefore differences in productivity of the two skill levels are solely identified through $\theta_L(t)$ and $\theta_H(t)$. Additionally we demand that the two skill parameters should also add up to one, because we then obtain a normalised production function and assure constant returns to scale for the Cobb-Douglas production function, which emerges for $\rho \to 0$. Lastly one should also mention, that $H_0 = 10^{-6}$ and is therefore not set equal to zero to avoid numerical difficulties, which would arise in the case of imperfect substitutability later on.

In figure 2.4 the optimal education rate for a fixed time-point $t = 40$ is presented. The numerical results support the theoretical results from proposition 2, i.e. the numerical solution is independent of time before $t = T - \omega$. In figure 2.4 we chose $t = 40 < 140 - 40$ arbitrary to represent the stationary solution.

[6] I have chosen the same parameters as in Prskawetz et al. (2012). Also the results of my own simulation match the ones of the authors.

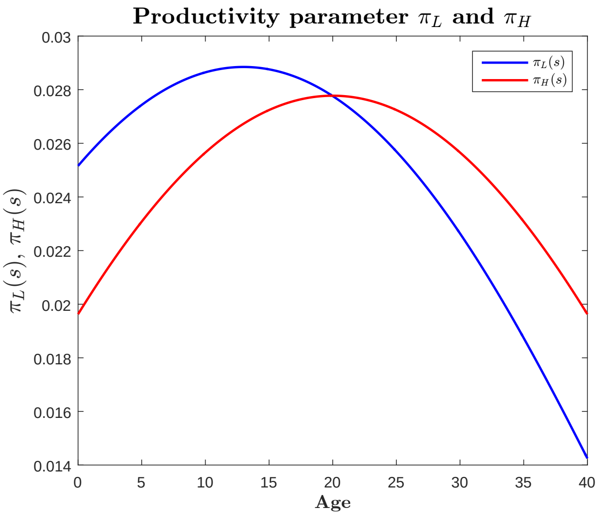


Figure 2.3: Age-structured productivity $\pi_L(s)$ and $\pi_H(s)$

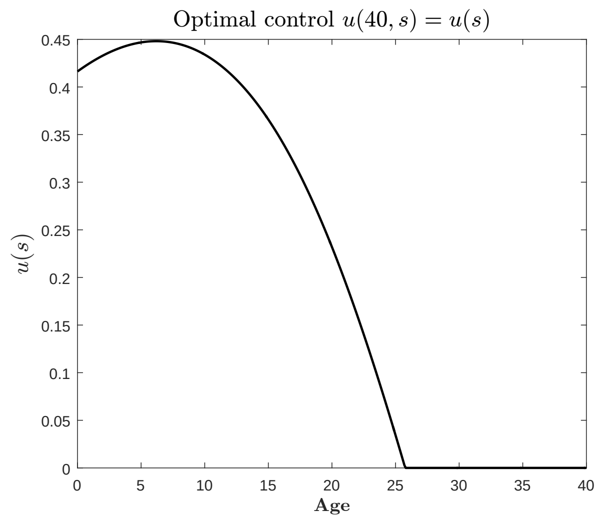


Figure 2.4: Optimal education rate $u(40, s) = u(s)$ for the benchmark parameters

Figure 2.4 shows that the optimal education rate increases in the first seven years, reaches its peak and drops again afterwards. After the (working) age of 26, no educational efforts are made by the social planer. Since the low- and high-skilled workers are perfectly substitutable, this cannot be explained by substitutional effects, but through the fact, that older workers have less time left in the labour market and therefore the education cost surpass the benefits of the worker becoming highly educated.[7] The relatively low benefits are not only driven by the shorter time left in the labour market but also by the worker's declining productivity in older ages (see figure 2.3).

One should note, that an increasing education rate in younger ages is not a standard result. In many papers about optimal education, the highest investments take place at the youngest ages and they decline as age increases. In the following section we will analyse under which

[7]Equation (2.5) shows this relationship. On the left side we see the marginal education cost, on the right side the benefits of a worker receiving education

conditions an increasing education rate for some working ages can occur.

2.3.2 Increasing education rate

To analyse, why and when an education rate increasing in age can appear, we will further assume the cost function to be independent of age and quadratic[8], i.e. $p(u) = bu + \frac{c}{2}u^2$, and the learning abilities $l(s)$ to be constant. It should be noted that the assumption about $l(s)$ has no effect on the results. Since we are trying to identify the reason for an increasing education rate, which can obviously not be encouraged by decreasing learning abilities, choosing $l(s)$ constant is justified for this analysis. Declining learning abilities in age might extinguish an otherwise appearing increasing education rate, but will never be the reason for it. The following proposition summarises under which parameter constellations an increase in the education rate can occur.

Proposition 3

Assume that

- *all data are time-invariant (except the initial data* $L_0(t)$ *and* $H_0(t)$*) and continuous in age* s
- *learning abilities are constant* $l(s) = l$
- *the cost-function is quadratic and independent of age,* $p(u) = bu + \frac{c}{2}u^2$

If there exist a value $s_0 \in [0, \omega] : u(s_0) > 0$ *(this implies that the optimal control is not identical to* 0 *on* $[0, \omega]$*) and*

$$\theta_H \pi_H(s_0) - \theta_L \pi_L(s_0) \;\leq\; \frac{b}{l} d(s_0)$$

with $d(s) := \Big(r + \delta(s) + e(s)\Big)$*, then* u *is differentiable at* s_0 *and* $u'(s_0) > 0$*.*

Proof. The original proof can be found in Prskawetz et al. (2012, p. 169).

The quadratic cost-function leads to $p'(u) = b + cu$ and inserting it in equation (2.6) results in

$$u(t,s) = \begin{Bmatrix} \frac{1}{c}\left(\Delta(s)l - b\right) & \text{if } \Delta(s)l \geq b \\ 0 & \text{otherwise} \end{Bmatrix} = \max\left\{0, \frac{1}{c}\left(\Delta(s)l - b\right)\right\}$$

And since the data is time-invariant the PDE (2.3) becomes an ODE

$$\Delta'(s) = \Big(r + e(s) + \delta(s) + l \cdot u(s)\Big)\Delta(s) - \Big(bu(s) + \frac{c}{2}u(s)^2\Big) - \Big(f_H(s) - f_L(s)\Big)$$

Because all data is continuous, $\Delta(s)$ is continuous and therefore also $u(s)$. Now let (a_0, a) be

[8] This corresponds to the form of educational costs used for the simulations (see table 2.1).

a subinterval of the support of u (which is $\neq \emptyset$, due to the assumption $\exists s_0 : u(s_0) > 0$), with $a = \max\{s : s \in \text{supp}(u)\}$. So we have $u(s) = \frac{1}{c}\left(\Delta(s)l - b\right) > 0$ on (a_0, a) and $u(s)$ is also continuous differentiable on this interval. The condition $u(s) > 0$ at the same time implies $\Delta(s) > \frac{b}{l}$.

We now assume, that $u'(s) \leq 0$ on (a_0, a). We easily see, that $a_0 = 0$ must hold in this case. Since $u'(s) = \frac{l}{c}\Delta'(s)$ it must also hold that $\Delta'(s) \leq 0$ on this interval. Inserting $u(s) = \frac{1}{c}\left(\Delta(s)l - b\right)$ in the ODE for $\Delta(s)$ results in

$$\Delta'(s) = \Big(r + e(s) + \delta(s) + \frac{l}{c}\left(\Delta(s)l - b\right)\Big)\Delta(s) - \Big(\frac{b}{c}\left(\Delta(s)l - b\right) + \frac{1}{2c}\left(\Delta(s)l - b\right)^2\Big) - \Big(f_H(s) - f_L(s)\Big)$$

And rearranging leads to the equation

$$\Delta'(s) = q_2\Delta(s)^2 + q_1(s)\Delta(s) + q_0(s)$$

$$\text{with} \quad q_2 = \frac{l^2}{2c}, \qquad q_1(s) = \left(r + e(s) + \delta(s) - \frac{bl}{c}\right), \qquad q_0(s) = \frac{b^2}{2c} - \left(f_H(s) - f_L(s)\right)$$

The condition now cannot hold if there exists one s', so that the function $Q(x) = q_2x^2 + q_1(s)x + q_0(s) > 0$ for all $x > \frac{b}{l}$. Since $q_2 > 0$ it is assured that the function $Q(x)$ is convex, there are two possible scenarios for the condition to be fulfilled.

(i) $Q(x)$ has no roots, what holds if $q_1(s)^2 - 4q_2q_0(s) < 0$. Substituting the terms leads to the condition

$$\theta_H\pi_H(s) - \theta_L\pi_L(s) < \frac{b^2}{2c}\Big(1 - (1 - \frac{c}{bl}d(s))^2\Big)$$

(ii) The bigger root is still smaller than or equal $\frac{b}{l}$, i.e.

$$\frac{-q_1(s) + \sqrt{q_1(s)^2 - 4q_2q_0(s)}}{2q_2} \leq \frac{b}{l}$$

and $q_1(s)^2 - 4q_2q_0(s) \geq 0$. This case corresponds to the conditions

$$\theta_H\pi_H(s) - \theta_L\pi_L(s) \;\geq\; \frac{b^2}{2c}\Big(1 - (1 - \frac{c}{bl}d(s))^2\Big) \qquad \text{and}$$

$$\theta_H\pi_H(s) - \theta_L\pi_L(s) \;\leq\; \frac{b}{l}d(s)$$

Combining those to cases, we obtain, that $Q(x) > 0$ for all $x > \frac{b}{l}$, if

$$\theta_H\pi_H(s) - \theta_L\pi_L(s) \;\leq\; \frac{b}{l}d(s)$$

what concludes the proof. ■

This proposition basically states that it is sufficient for the derivative of the education rate to be positive for a certain age, if the difference in total productivity $\theta_H\pi_H(s) - \theta_L\pi_L(s)$ is relatively small compared to $d(s)$ the sum of discounting, depreciation and learning by doing rate at this age.

It is reasonable that a small discrepancy in productivity between the skill levels leads to a postponement of education, since the benefits of the slightly higher productivity might be outweighed by the cost resulting from the educational efforts. On the other hand a high value of $d(s)$ can be economically also interpreted as follows. Applying educational efforts raises costs, which lead to expenditures by the social planer he therefore cannot save, what reasons the term r in $d(s)$. So if the interest rate is higher, the social planer is more likely to save instead of investing in education. A person becoming high-skilled at a certain age is also exposed to the risk of losing his costly education at the rate $\delta(s)$. Therefore a high depreciation rate at a certain age might lead to postponement of education. At last applying costly educational efforts eliminates the possibility of cost-free learning by doing, so high values of $e(s)$ might again lead to shifting education to older ages.

The assumption in the proposition above are weaker than the assumptions in proposition 3 in Prskawetz et al. (2012, p.168). This improvement is due to the fact that the authors demanded the quadratic function $Q(x)$ to be positive for all non-negative values of x, while it is sufficient if $Q(x) > 0$ for all $x \geq b/l$. This leads to a simplification in the calculations and makes the case distinction in proposition 3 in Prskawetz et al. (2012) unnecessary.

2.4 Imperfect substitutability and demographic change

In proposition 2 we have seen, that in case of perfect substitutability changes in the demographic structure do not have any effect on the optimal educational rates. Apparently this is not the case if workers are not perfectly substitutable. In the following we will still assume that high and low-skilled workers can be perfectly substituted, i.e. $\rho = 1$ and focus on the imperfect substitutabilities between different age-groups, i.e. $\lambda_L \neq 1$ and $\lambda_H \neq 1$.[9] We are going to analyse two different aspects of demographic change in the following sections: The long-run and short-run effects.

2.4.1 Long-run effects of demographic change

In this section we are going to analyse population developments of the form

$$L_0(t) = L_0 e^{\gamma t} \qquad \text{and} \qquad H_0(t) = H_0 e^{\gamma t}$$

so we are able to analyse an exponentially increasing, decreasing or constant population $N(t)$.

$$N(t) = \int_0^\omega L(t,s) + H(t,s)ds = \int_0^\omega L(t-s,0) + H(t-s,0)ds = \int_0^\omega e^{\gamma(t-s)}(L_0 + H_0) =$$

[9]Since we introduce age as a characteristic of individuals, changing the age-specific substitution rates is more likely to reveal new insight and should be the main focus.

$$= e^{\gamma t} \int_0^{\omega} e^{-\gamma s} N_0 ds \quad = \begin{cases} e^{\gamma t} N_0 \left(\frac{1}{\gamma}(1 - e^{-\gamma\omega}) \right) & \gamma \neq 0 \\ \omega N_0 & \gamma = 0 \end{cases} \tag{2.7}$$

Numerical results

In the following numerical simulations we compare the optimal education rates for the scenarios with

- a low substitutability of low-skilled workers ($\lambda_L = 0.1$) and high substitutability of high-skilled workers ($\lambda_H = 0.9$)
- high substitutability of low-skilled workers ($\lambda_L = 0.9$) and low substitutability of high-skilled workers ($\lambda_H = 0.1$)

both for an increasing population ($\gamma = 0.0072$), a constant population ($\gamma = 0$) and a decreasing population ($\gamma = -0.0072$). All other parameters are chosen as in the benchmark case (see table 2.1). To make the results better comparable, we will also consider the normalised education rate $v(t, s)$ for a given time-point t, which is defined by

$$v(t, s) = \frac{u(t, s)}{\int_0^{\omega} u(t, s) ds}$$

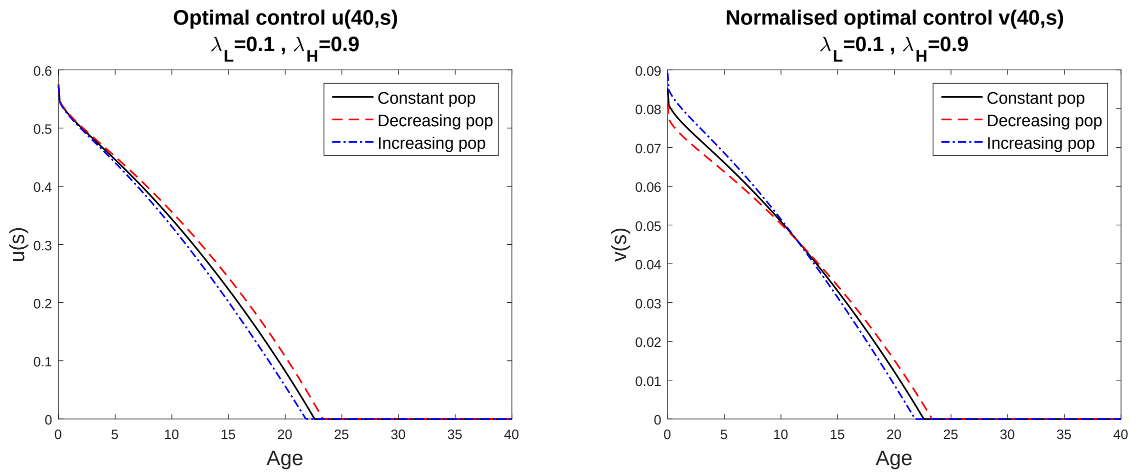
to identify shifts in the importance of certain age-groups.

Figure 2.5 shows the results for the case $\lambda_L = 0.1$ and $\lambda_H = 0.9$. In the left frame we see that the education rate is monotonically decreasing for all ages in contrast to the benchmark case. It is apparent that an increasing population size leads to a lower education rate compared to the constant population case, while a decreasing population results in higher education rates, both holding for all ages. As already described above it might be more insightful to compare the normalised educational efforts in the right frame of figure 2.5. There we can identify, that changing population sizes lead to a shift of the focus of education between ages. For a decreasing population we obtain a shift from younger to older ages, while for an increasing population size it holds vice versa (always compared to the constant population case).

For the other combination of elasticities of substitution (i.e. $\lambda_L = 0.9$ and $\lambda_H = 0.1$) in figure 2.6, one should note, that the education rate sharply increases when approaching the age of zero. Intuitively this is due to the fact that high-skilled labour can only be badly substituted between the age-groups, leading to a relative high optimal number of young high-skilled individuals, regarding the fact that only an insignificant number ($H_0 = 10^{-6}$) enters the labour market already high-skilled at age zero. As a result the optimal education rates at very young ages are very pronounced compared to the rates at middle or higher ages.[10]

Analysing the effects of population changes we see, that the results are now just the opposite. For a decreasing population we find an adjustment towards younger ages, while for an increasing

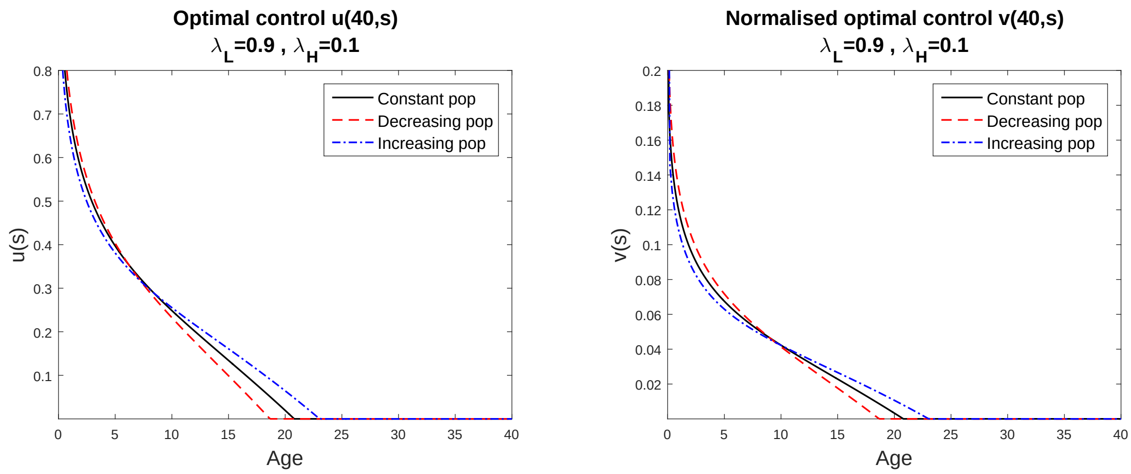
[10]This pattern also causes some troubles, while calculating the optimal control numerically. These problems are discussed in appendix C.



(a) Optimal education rate $u(40, s)$ (b) Optimal normalised education rate $v(40, s)$

Figure 2.5: Optimal controls for three population scenarios for $\lambda_L = 0.1$ and $\lambda_H = 0.9$

population the shift is towards the older ages. This time the pattern is not only visible in the normalised rates in figure 2.6b, but also in the non-normalised results (figure 2.6a).



(a) Optimal education rate $u(40, s)$ (b) Optimal normalised education rate $v(40, s)$

Figure 2.6: Optimal controls for three population scenarios for $\lambda_L = 0.9$ and $\lambda_H = 0.1$

The intuitive explanation for these various shifts is that for certain elasticities of substitution the social planer aims for a special input-distribution in both low and high-skilled workers. An increasing number $L_0(t)$ leads to an excess of young low-skilled, but also of high-skilled individuals if the education rate doesn't change compared to the stationary case. Now one has to distinguish, within which skill group workers of different ages can be substituted worse. If $\lambda_L << \lambda_H$ the social planer will focus on re-establishing the age-distribution of the low-skilled workers. To reduce the number of young low-skilled workers the education rate at younger ages will increase, while lower education efforts at higher ages lead to an increase in older low-skilled workers (see figure 2.5b). On the other hand if $\lambda_L >> \lambda_H$ the age-composition of the high-skilled will be reinstalled by diminishing the education rate at lower ages and increasing it for higher ages. For a decreasing population size the argumentations are just the other way round.

After all we see, that its not the change in the population size itself, that leads to differences in the age-specific optimal education rates, but the change in the age-composition of the population resulting out of it. To clarify, why the age structure changes, we consider the relative size of an age-group at a time point t compared to the total population size, i.e.

$$\frac{L(t,s)+H(t,s)}{\int_0^\omega [L(t,s)+H(t,s)]\,ds}$$

It holds that the size of a cohort doesn't change over time, since we have abstracted from mortality:

$$L(t,s)+H(t,s) =: N(t,s) = N(t-s,0) = N_0(t-s) = N_0 e^{\gamma(t-s)} = (L_0+H_0)e^{\gamma(t-s)}$$

So together with equation (2.7) we can reduce the relative population size to the term

$$\frac{L(t,s)+H(t,s)}{\int_0^\omega [L(t,s)+H(t,s)]\,ds} = \frac{\gamma}{1-e^{-\omega\gamma}}e^{-s\gamma} \tag{2.8}$$

In the stationary case all age-groups have the same share in the total society, while we have now shown, that for an increasing population the percentage of one age-group is lower, the older the age-group is (see the term $e^{-s\gamma}$ in equation (2.8)). For a decreasing population the statements are vice-versa.

Analytical foundation for the numerical results

Under the assumptions, that $\rho = 1$ and that all data except the boundary and initial data is time-invariant, we can lead the case of an exponentially changing population back to the stationary case. To show this we introduce the variables $L^\gamma(t,s) = e^{-\gamma(t-s)}L(t,s)$ and $H^\gamma(t,s) = e^{-\gamma(t-s)}H(t,s)$. Using the fact, that it holds that

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial s}\right)\left[e^{-\gamma(t-s)}L(t,s)\right] = e^{-\gamma(t-s)}\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial s}\right)L(t,s)$$

we see, that $L^\gamma(t,s)$ and $H^\gamma(t,s)$ also fulfil the PDEs in problem 1, but the boundary conditions become constant, i.e. $L^\gamma(t,0) = L_0$. Substituting $L^\gamma(t,s)$ (resp. $H^\gamma(t,s)$) into the other equations of problem 1 results in the problem

$$\max_{u(t,s)} \int_0^T e^{-(r-\gamma)t}\left[\theta_L \tilde{L}^\gamma(t) + \theta_H \tilde{H}^\gamma(t) - P(t)\right]dt \tag{2.9}$$

$$L^\gamma(t,0) = L_0, \qquad H^\gamma(t,0) = L_0 \qquad \forall t \in (0,T] \tag{2.10}$$

$$\tilde{L}^\gamma(t) = \left(\int_0^\omega e^{-\gamma\lambda_L s}\pi_L(s)L^\gamma(t,s)^{\lambda_L}ds\right)^{1/\lambda_L} \tag{2.11}$$

$$\tilde{H}^\gamma(t) = \left(\int_0^\omega e^{-\gamma\lambda_H s}\pi_H(s)H^\gamma(t,s)^{\lambda_H}ds\right)^{1/\lambda_H} \tag{2.12}$$

$$P(t) = \int_0^\omega e^{-\gamma s} p(s, u(t,s)) L^\gamma(t,s) ds \tag{2.13}$$

Summarising we see, that the optimal solution $u(t,s)$ for the non stationary problem is equivalent to the optimal solution of a stationary problem with the given data

$$\pi_L^\gamma(s) = e^{-\gamma\lambda_L s}\pi_L(s) \quad , \quad \pi_H^\gamma(s) = e^{-\gamma\lambda_H s}\pi_H(s) \quad , \quad p^\gamma(s,u) = e^{-\gamma s}p(s,u(t,s)) \quad , \quad r^\gamma = r - \gamma$$

Hence the social planer in the long run reacts to an exponentially changing population the same way he would react to a change in the cost and productivity structure described above. The case of an increasing population size ($\gamma > 0$) is equivalent to the costs of education becoming relatively cheaper in older ages. However this doesn't lead to higher education rates in older ages generally, since the productivity of high skilled workers declines at the rate $\gamma \cdot \lambda_H$ in comparison to the constant population case. On the other hand the productivity of low-skilled declines too, but at the rate $\gamma \cdot \lambda_L$, which is in general different to $\gamma \cdot \lambda_H$. Intuitively the cases $\lambda_L < \lambda_H$ and $\lambda_L > \lambda_H$ can be analysed for an increasing population as follows:

- If high-skilled workers are better substitutable with respect to age ($\lambda_L < \lambda_H$) a positive value for γ results in a relatively faster declining productivity of high-skilled workers. Thus there are less incentives to educate people of higher age, i.e. $u(t,s)$ is smaller for higher ages s compared to the stationary case. This interpretation fits the results in figure 2.5, where the productivity decline overcompensates the relatively lower cost for education at older ages.

- For better substitutable low-skilled workers ($\lambda_L < \lambda_H$), we obtain a relatively stronger decreasing productivity of low-skilled workers, which motivates higher education rate at older ages. At the same time the relatively declining costs for education leads to a postponement in education efforts resulting in lower education rates at younger ages (see figure 2.6).

2.4.2 Short-run effects of the demographic change

In this section we want to study the short-run effects of demographic changes. To do so, we will analyse how the social planer reacts if the population size changes exponentially after a fixed time-point $\bar{t}$. Furthermore we will investigate the optimal education rates shortly after, but also before $\bar{t}$. For the simulations I have chosen $\bar{t} = 20$. Preliminary I have to say, that the results of my own simulations do not fit the results of Prskawetz et al. (2012) as good as in the last sections. This might be due to the fact, that I was not able to exactly reconstruct how the authors treated the influence of the initial population structure. At $t = 20$ the initial population plays a decisive role on the optimal education rates, since every individual with 20 years or less in the labour market in the beginning is still working at $t = 20$. Additionally we do not have perfect substitutability with respect to age, so formally we can not apply proposition 2 about the stationary solution. Nevertheless the simulations have shown, that on a time interval with enough distance to the start and endpoint of the time horizon the solution is approximately stationary, because it only changes insignificantly with respect to time within this period of time. Taking this knowledge we can think of two possible ways to account for the influence of the initial population:

- Start the system already before $t = 0$, so that the initial population doesn't affect the system anymore at $t = 20$.
- Set the initial population equal to the corresponding approximated stationary population, which we already implicitly calculated in the last section.

I chose the second approach and although my results are not identical to the ones in Prskawetz et al. (2012), my simulations still lead to the same qualitative properties in the optimal solutions.

In figure 2.7 we see the optimal education rates for the two parameter constellations at time-point $t = 18$ two years before the change in the population size occurs. We see that the effects are minor but still significant. For the case $\lambda_L << \lambda_H$ we obtain the same results as in the last section, while for the $\lambda_L >> \lambda_H$ combination we now obtain that an increasing population leads to a higher education rate for all ages (and vice versa). After all this means that in both cases the social planer reacts to population changes even before they happen.[11]

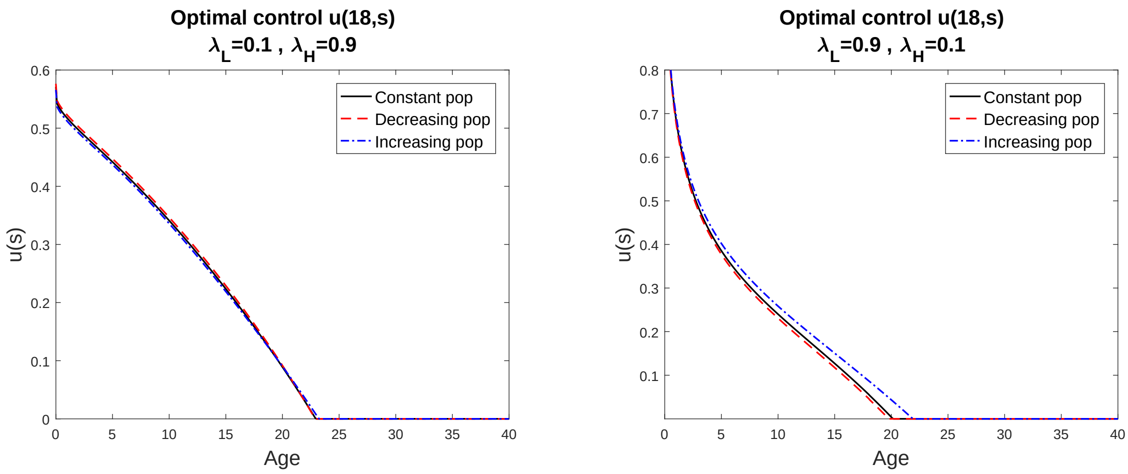


Figure 2.7: Optimal education rates at $t = 18$ (two years before the change in population size)

This anticipation phenomenon becomes even more clear, when considering the per capita education effort (PCE)

$$PCE(t) = \frac{\int_0^\omega L(t,s)u(t,s)ds}{\int_0^\omega N(t,s)ds}$$

This value describes the relative share of individuals being in education at a certain time t. Especially for the time period at the beginning with constant population this is a relevant indicator. Figure 2.8 shows the time path of the $PCE(t)$ from $t = 5$ to $t = 30$ for the case

[11] It should be noted, that compared to other models in our model perfect foresight for the social planer is relatively realistic. A change in the number of individuals entering the labour market can be partially expected around 15 years before it happens by analysing the birth numbers.

$\lambda_L >> \lambda_H$. The figure shows, that the PCE is constant over time for the constant population case, while a changing population size has an impact on the PCE even before the change takes place. Corresponding to figure 2.7 a decreasing population leads to decreasing per capita education effort, while increasing per capita results in increasing per capita education efforts in the time period before and shortly after the changing point.

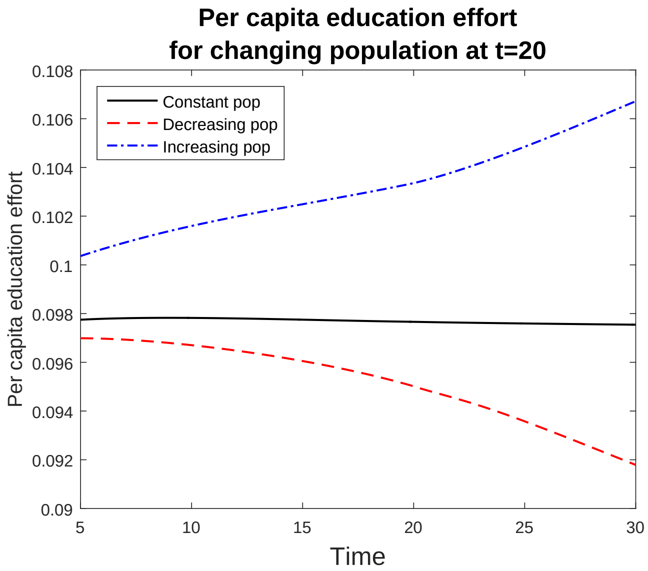


Figure 2.8: Per capita education effort for changing population after $t = 20$

2.5 Conclusio

The model of Prskawetz et al. (2012) introduces age as a characteristic of the individuals in an economy and enables the social planer apply education efforts or training specific to certain age-groups. In the numerical solution for the benchmark scenario with perfect substitutability we have obtained non-standard results, namely an increasing education rate in the first years in the working life. To support this numerical result, we have derived sufficient conditions for an increasing optimal education rate at some age in proposition 3. In the case of perfect substitutability and time invariant parameters we also obtained a time invariant solution, which is not affected by changes in the initial population or in the population entering with age 0.

Dropping the assumption of perfect substitutability with respect to age shows that these elasticities have a pronounced impact on the optimal education rates. In both numerically observed parameter constellations (with low-skilled or high-skilled workers being relatively better substitutable) we obtained monotonically decreasing optimal education rates. Additionally changes in the population structure become influential in these scenarios. In the long-run

- the education focus shifts to younger ages for an increasing population size in the case of high-skilled labour being relatively better substitutable with respect to age. For a decreasing population the shift happens towards older ages.

- if low-skilled labour is relatively better substitutable, the results are vice-versa. A decreasing population size leads to higher education rates at younger ages and lower rates at higher ages, while a growing population implies lower educational efforts at younger ages and higher at older ages.

Furthermore we have shown, that demographic changes also have short-run effects and the social optimal solution anticipates the change and reacts already to it, before it eventually happens. We have also analysed the case of relatively better substitutable low-skilled labour and obtained higher (lower) per capita education rate for an increasing (decreasing) population shortly after and even several years before the population size begins to rise (decline).

After all we have seen that the elasticities of substitution are of major interest and that the optimal solutions are relatively sensitive to changes in these parameters. As a result they have to be carefully estimated, if one wants to obtain realistic and policy relevant results. Also a lot of other features of the model like skill depreciation, learning-by-doing, different learning abilities, which kept the model very general in the first place, have been more or less eliminated in the numerical analysis to focus on the effects of population change for different age-specific elasticities of substitution. However these features should enable the possibility of very realistic simulations, if aimed for, and make the model also a matter of interest for different kinds of analysis.

Lastly the model also exhibits some shortcomings (e.g. the production function, only two skill levels), which I already partly addressed within the last chapter and I will try to resolve some of them in the upcoming chapters.

3 | Extension with three skill levels

3.1 Motivation

Despite providing interesting insights in the structure of education policies, assuming only two different skill groups is a major simplification for analysing realistic scenarios. The ISCED[1] classification used for international statistics concerning education subjects distinguishes between nine different levels of education listed in table 3.1.

ISCED level	Description
0	Early childhood education
1	Primary education
2	Lower secondary education
3	Upper secondary education
4	Post-secondary non-tertiary education
5	Short-cycle tertiary education
6	Bachelor's or equivalent
7	Master's or equivalent
8	Doctoral or equivalent

Table 3.1: ISCED coding of educational levels

The levels 0 and 1 contain the lowest forms of education (up to the age of 10-12) and are called *primary education*. For level 2 "usually, the aim is to lay the foundation for lifelong learning and human development upon which education systems may then expand further educational opportunities."[2] Upper secondary education (level 3) contains both vocational training and general education and is "typically designed to complete secondary education in preparation for tertiary education or provide skills relevant to employment, or both."[3] Level 4 education aims at offering a possibility for specialisation or making progression after having completed level 3 education, but cannot hold up to the complexity of tertiary education.[4]. This also leads to its name "Post-secondary but non-tertiary" education.

Level 5 programs have generally more complex content than level 3 or 4, but are shorter

[1]International Standard Classification of Education, for the manual see UNESCO (2011)
[2]UNESCO (2011), p. 33
[3]UNESCO (2011), p. 38
[4]UNESCO (2011), p. 43

than the other tertiary education levels. On the one hand they are typically preparing the individuals for the labour market, but on the other hand they "may also provide a pathway to other tertiary education programmes"[5]. The remaining levels 6,7 and 8 should be relatively clear due to their short description and represent the classic higher academic education.

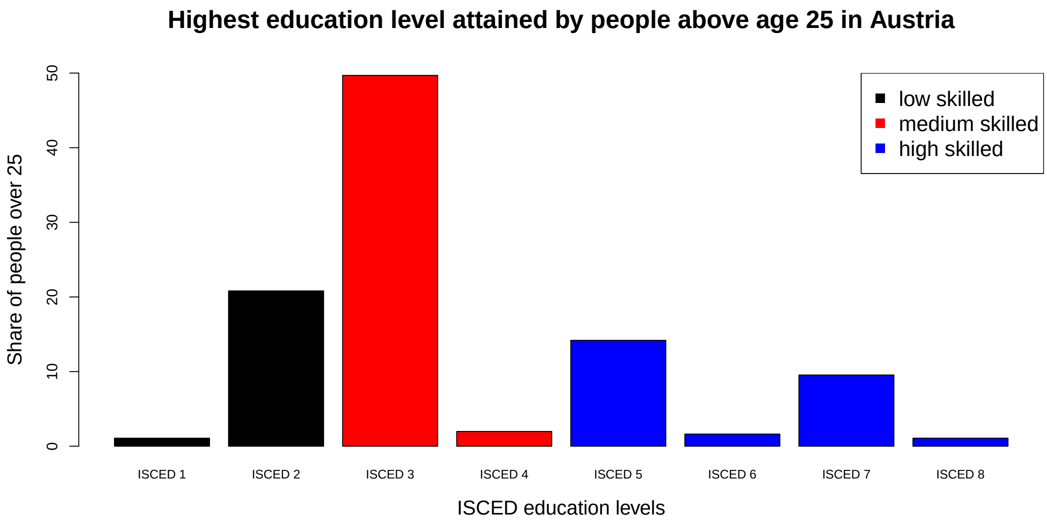


Figure 3.1: Shares of the highest attained educational level of people above age 25 in Austria in 2014 (source: UNESCO (2016))

Figure 3.1 shows the distribution of the highest level of education attained by individuals older than 25 in Austria. The data is taken from UNESCO (2016), the online database of the UNESCO Institute for statistics. We can see, that the levels 2 (lower secondary), 3 (upper secondary), 5 (short-circle tertiary) and 7 (Master's or equivalent) have the highest shares and together cover over 94% of the population above age 25. Taking into account the importance of the different levels, their descriptions and their designations it seems convenient to collapse the 9 different levels into the following three skill levels:

- low skill → levels 0, 1, 2

  (The level 0 is non existent in figure 3.1, because children in Austria mandatory have to spend nine years in education and therefore at least attain ISCED level 1.)

- medium skill → levels 3, 4

- high skill → levels 5, 6, 7, 8

This subdivision is illustrated in figure 3.1 by the colours of the bars. In Austria the low-skilled workers together make up around 22% of the population, nearly 52% of the population is medium skilled, while the remaining 26% are high skilled. Summing up we see, that a three level educational system reflects the reality rather well and seems to be an appropriate classification. In general the separation of the working population into three different skill groups or education levels is widely spread in the literature. Especially in empirical papers

[5]UNESCO (2011), p. 48

this setup is used often, e.g. in Lutz and Samir (2011), Philipov et al. (2014) or Loichinger and Weber (2016), and also statistical databases commonly present their data in a three-level-model. Lutz and Samir (2011) additionally included "no education obtained" as an education level, which on the other hand has become nearly irrelevant in modern societies of developed countries. Philipov et al. (2014) and Loichinger and Weber (2016) in their respective papers also collapsed the ISCED educational levels in three groups in the same way as described above.

3.2 The model

Introducing a new additional skill group in the model, requires an adaptation of the dynamics and functions from the previous model from Prskawetz et al. (2012) shown in chapter 2. Figure 3.2 shows the flows between the skill groups. We postulate two important assumptions:

- At first there is no transition from medium to high skill. This assumption is reasoned by the fact, that for example in Austria there is seldom a change from medium to high skilled due to the different emphases of the two education forms. Medium-skilled education (represented by apprenticeships) for a big share consists of vocational training and has a higher practical focus, while high-skilled education (academic education) is characterised by a more general form of education, which goes beyond the focus of direct practical applications. Later on (chapter 4) we will introduce a control for the social planer to upgrade workers from medium to high skill and analyse the changes in the optimal solutions. This extension is motivated by the relatively new "Lehre mit Matura" in Austria. This program enables young individuals running through their apprenticeship to additionally obtain higher academic education later in their career.

- There is no chance of becoming high skilled by learning by doing. Only the medium skill level can be attained by such an upgrading. Again this is motivated by the different main aims of the education forms. Medium-skill might be attained through long years of practical work, while high-skill education cannot be obtained by practical experience.

The parameter and variable names are kept the same, only the index M or H is added to distinguish between the different levels. I.e. we have different depreciation rates $\delta_M(t,s)$ and $\delta_H(t,s)$ and the social planer can decide in which education he wants to invest in, resulting in different education rates $u_M(t,s)$ and $u_H(t,s)$.

In compliance to the basic model we obtain partial differential equations describing the changes in size of a skill-group of a certain age. It should be noted, that introducing the third skill level eliminates the beneficial structure in the basic model of the outflows of one group being the inflows of the other.

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) L(t,s) = \delta_M(t,s)M(t,s) + \delta_H(t,s)H(t,s) - e(s)L(t,s) - l_M(t,s)u_M(t,s)L(t,s) - \\ - l_H(t,s)u_H(t,s)L(t,s)$$

$$L(t,0) = L_0(t) \qquad \forall t \in (0,T]$$

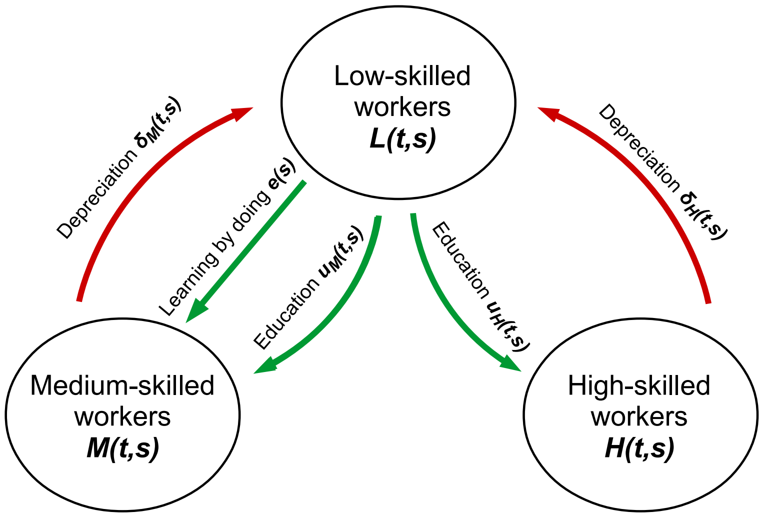


Figure 3.2: Transition flows between different skill groups in the extended model

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial s}\right)M(t,s)=-\delta_M(t,s)M(t,s)+e(s)L(t,s)+l_M(t,s)u_M(t,s)L(t,s)$$

$$M(t,0)=M_0(t)\qquad \forall t\in(0,T]$$

$$\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial s}\right)H(t,s)=-\delta_H(t,s)H(t,s)+l_H(t,s)u_H(t,s)L(t,s)$$

$$H(t,0)=H_0(t)\qquad \forall t\in(0,T]$$

For the costs of education we assume the same qualitative structure for both kinds of education resulting in the following term for the total cost of education.

$$P(t)=\int_0^{\omega}\Big(p_M(s,u_M(t,s))L(t,s)+p_H(s,u_H(t,s))L(t,s)\Big)ds$$

Regarding the production function we again first build the sub-aggregates over all ages for all skills. In addition to the low and high skilled sub-aggregates we now also include the sub-aggregate for the medium skill level as an input for the final good production function $Y(t)$. All functions are again of the constant elasticity of substitution (CES-)type.

$$\widetilde{L}(t)=\left(\int_0^{\omega}\pi_L(s)L(t,s)^{\lambda_L}ds\right)^{1/\lambda_L}$$

$$\widetilde{M}(t)=\left(\int_0^{\omega}\pi_M(s)M(t,s)^{\lambda_M}ds\right)^{1/\lambda_M}$$

$$\widetilde{H}(t)=\left(\int_0^{\omega}\pi_H(s)H(t,s)^{\lambda_H}ds\right)^{1/\lambda_H}$$

$$Y(t) = \Big(\theta_L(t)\widetilde{L}(t)^\rho + \theta_M(t)\widetilde{M}(t)^\rho + \theta_H(t)\widetilde{H}(t)^\rho\Big)^{1/\rho}$$

All together we can again summarize the problem faced by the social planer as follows:

Problem 2

In the extended model the social planer solves the following optimisation problem

$$\max_{\{u_M(t,s),u_H(t,s))\}} \int_0^T e^{-rt}\Big(Y(t) - P(t)\Big)dt$$

subject to the PDEs

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) L(t,s) = \delta_M(t,s)M(t,s) + \delta_H(t,s)H(t,s) - e(s)L(t,s) - \\ - l_M(t,s)u_M(t,s)L(t,s) - l_H(t,s)u_H(t,s)L(t,s) \qquad (3.1)$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) M(t,s) = -\delta_M(t,s)M(t,s) + e(s)L(t,s) + l_M(t,s)u_M(t,s)L(t,s) \qquad (3.2)$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) H(t,s) = -\delta_H(t,s)H(t,s) + l_H(t,s)u_H(t,s)L(t,s) \qquad (3.3)$$

the boundary conditions

$$L(t,0) = L_0(t), \qquad M(t,0) = M_0(t), \qquad H(t,0) = H_0(t) \qquad \forall t \in (0,T]$$

$$L(0,s) = L_b(s), \qquad M(0,s) = M_b(s), \qquad H(0,s) = H_b(s) \qquad \forall s \in [0,\omega]$$

the aggregate equations

$$\widetilde{L}(t) = \left(\int_0^\omega \pi_L(s)L(t,s)^{\lambda_L} ds\right)^{1/\lambda_L} \qquad (3.4)$$

$$\widetilde{M}(t) = \left(\int_0^\omega \pi_M(s)M(t,s)^{\lambda_M} ds\right)^{1/\lambda_M} \qquad (3.5)$$

$$\widetilde{H}(t) = \left(\int_0^\omega \pi_H(s)H(t,s)^{\lambda_H} ds\right)^{1/\lambda_H} \qquad (3.6)$$

$$Y(t) = \Big(\theta_L(t)\widetilde{L}(t)^\rho + \theta_M(t)\widetilde{M}(t)^\rho + \theta_H(t)\widetilde{H}(t)^\rho\Big)^{1/\rho} \qquad (3.7)$$

$$P(t) = \int_0^\omega \Big(p_M(s,u_M(t,s))L(t,s) + p_H(s,u_H(t,s))L(t,s)\Big)ds \qquad (3.8)$$

and the non-negativity constraints

$$u_M(t,s) \geq 0 \qquad u_H(t,s) \geq 0$$

3.3 Optimality conditions

For the application of the optimality conditions in Feichtinger et al. (2003) we need to make several assumptions first. We will use the standing assumptions of the basic model (Assumption 1) and extend them to all newly added functions and parameters. With these assumptions made applying the optimality conditions leads to intensive calculations similar to the basic model. The calculations can be found in appendix B.2 and the results are summarised in the following proposition 4.

Proposition 4

Let the standing assumptions still hold, also for the newly introduced parameters, initial and boundary functions for medium skill worker. If $(L, M, H, \widetilde{L}, \widetilde{M}, \widetilde{H}, P, u)$ is an optimal solution for the problem 2 than the system of partial differential equations

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \mu_L(t,s) = \Big(e(s) + l_M(t,s)u_M(t,s) + l_H(t,s)u_H(t,s) + r\Big)\mu_L(t,s) - \\ - \Big(e(s) + l_M(t,s)u_M(t,s)\Big)\mu_M(t,s) - l_H(t,s)u_H(t,s)\mu_H(t,s) + \\ + \Big(p_M(s, u_M(t,s)) + p_H(s, u_H(t,s))\Big) - f_L(t,s) \tag{3.9}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \mu_M(t,s) = \delta_M(t,s)\Big(\mu_M(t,s) - \mu_L(t,s)\Big) + r\mu_M(t,s) - f_M(t,s) \tag{3.10}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \mu_H(t,s) = \delta_H(t,s)\Big(\mu_H(t,s) - \mu_L(t,s)\Big) + r\mu_H(t,s) - f_H(t,s) \tag{3.11}$$

$$\mu_i(T,s) = 0 \quad \mu_i(t,\omega) = 0 \qquad \forall i \in \{L, M, H\}\ \forall t \in [0,T]\ \forall s \in [0,\omega] \tag{3.12}$$

with

$$f_L(t,s) = Y(t)^{1-\rho} \cdot \theta_L(t)\pi_L(s) \cdot \widetilde{L}(t)^{\rho-\lambda_L} L(t,s)^{\lambda_L - 1}$$

$$f_M(t,s) = Y(t)^{1-\rho} \cdot \theta_M(t)\pi_M(s) \cdot \widetilde{M}(t)^{\rho-\lambda_M} M(t,s)^{\lambda_M - 1}$$

$$f_H(t,s) = Y(t)^{1-\rho} \cdot \theta_H(t)\pi_H(s) \cdot \widetilde{H}(t)^{\rho-\lambda_H} H(t,s)^{\lambda_H - 1}$$

has a unique solution $(\mu_L(t,s), \mu_M(t,s), \mu_H(t,s))$. For almost every $(t,s) \in [0,T] \times [0,\omega]$ $u(t,s) = (u_M(t,s), u_H(t,s))$ maximizes the Hamiltonian in equation B.19 respectively solves the FOC

$$-\frac{\partial p_M(s, u_M(t,s))}{\partial u_M} + \Big(\mu_M(t,s) - \mu_L(t,s)\Big) l_M(t,s) \leq 0 \tag{3.13}$$

$$\left(-\frac{\partial p_M(s, u_M(t,s))}{\partial u_M} + \Big(\mu_M(t,s) - \mu_L(t,s)\Big) l_M(t,s)\right) \cdot u_M(t,s) = 0 \tag{3.14}$$

$$-\frac{\partial p_H(s, u_H(t,s))}{\partial u_H} + \Big(\mu_H(t,s) - \mu_L(t,s)\Big) l_H(t,s) \leq 0 \tag{3.15}$$

$$\Big(-\frac{\partial p_H(s,u_H(t,s))}{\partial u_H}+\Big(\mu_H(t,s)-\mu_L(t,s)\Big)l_H(t,s)\Big)\cdot u_H(t,s)=0 \tag{3.16}$$

These optimality conditions are similar to the conditions in the basic model (see proposition 1), but we are not able to reduce the system to the differences of the shadow prices due to the more complicated structure. On the other hand we are still able to characterise the optimal solutions by using the inverse functions of the derivatives of the cost-functions (based on their strictly convex structure)

$$u_M(t,s)=p'^{-1}_{M+}\Big(s,(\mu_M(t,s)-\mu_L(t,s))\,l_M(t,s)\Big)$$

$$=: \begin{cases} p'^{-1}_M\Big(s,(\mu_M(t,s)-\mu_L(t,s))\,l_M(t,s)\Big) & \text{if } \;(\mu_M(t,s)-\mu_L(t,s))\,l_M(t,s)\geq p'_M(s,0) \\ 0 & \text{otherwise} \end{cases}$$

and $u_H(t,s)$ analogue. The interpretations are similar to the basic model solution. There will be investment in medium (high) education if the shadow-price of one medium (high) skilled worker is higher than the shadow price of a low skilled worker taking into account the effectiveness of medium (high) skill education. Additionally the desired education rate will increase as the the difference between the shadow prices increases.

3.4 Perfect substitutability

In the next step we can directly extend the conclusion of proposition 2 to the extended model

Proposition 5

In the case of perfect substitutability the following statements hold

(a) The optimal education rates $u_M(t,s)$ and $u_H(t,s)$ do not depend on the initial and boundary data ($L_b(s)$,$M_b(s)$, $H_b(s)$, $L_0(t)$, $M_0(t)$ and $H_0(t)$).

(b) If all other data is time invariant, the optimal education rates are also time invariant on the interval $[0,T-\omega]$.

Proof. The proof of the proposition follows directly from applying the arguments of the proof of proposition 2 to the PDEs (3.9)-(3.11). ∎

Again this means that we can chose a fixed time point with enough time-lag to the end of the time horizon and interpret the given states and controls as the stationary solution. Trying to transfer the results from proposition 3 to the extended model turns out to be considerably more difficult. Nevertheless the numerical simulations show that increasing education rates are possible for both medium and high skill education.

Before we examine some examples, first we will summarise all the parameters chosen in table 3.2 similar to the basic model. We see that the parameters correspond to the ones in the benchmark case of the basic model with only some adaptations due to the additional parameters in the extended models

- High skill education is assumed to be 10% more effective than medium skill education, i.e. $l_H = l_M \cdot 1.1$
- The age-specific productivities have the same qualitative form as in the basic model and are illustrated in figure 3.3. Note that as in the basic model the integrals of these productivity functions are normalised to 1, so that all skill specific productivity differences are measured by $\theta_i(t)$ alone. Therefore $\pi_i(s)$ again measures only the age-specific differences within on skill group.
- We will generally assume that medium skill education causes α_{edu} times the costs of high skilled education. More precisely for any arbitrary but fixed education rate $\overline{u}$ the arising medium skill education cost equal the share α_{edu}% of the costs, that would arise for the same rate $\overline{u}$ of higher education.
- Regarding the skill specific productivities we will always suppose that low skill workers exhibit a very small productivity of 0.1. The productivities of medium and high skilled workers add up to 0.9 (to assure overall comparability between the different parameter constellations[6]) with a minimum of 0.1 each. This means $(\theta_M(t), \theta_H(t))$ is a convex combination of the points $(0.1, 0.8)$ and $(0.8, 0.1)$ with α_{prod} being the location parameter.

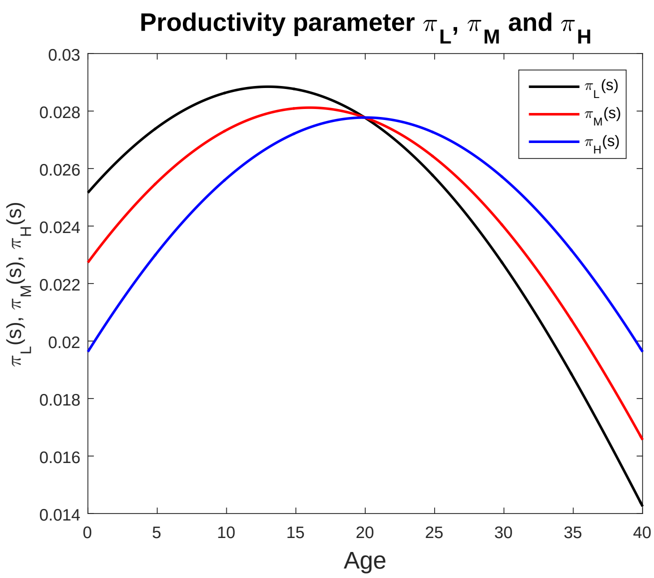


Figure 3.3: Age-structured productivity $\pi_L(s)$, $\pi_M(s)$ and $\pi_H(s)$.

Figure 3.4 illustrates the optimal education rates for two different parameter constellations. While it holds for both that $\theta_M(t) = 0.38$ and $\theta_H(t) = 0.52$, the costs of medium and higher

[6]Only increasing one of the productivity parameters without lowering the other, would lead to an overall productivity gain and result automatically in higher output even without newly allocating the labour. As the final good is the numeraire, we measure the education cost in units of the final good. With the higher production both types of education would become relatively cheaper, which would make it difficult to compare the results of a productivity change.

Parameter	Description	Value	Range
T	time horizon	140	
ω	maximal time in the labour market	40	
r	time preference rate / interest rate	0.03	
ρ	elasticity of substitution across skill levels $= \frac{1}{1-\rho}$	1	
λ_i	elasticity of substitution across age $= \frac{1}{1-\lambda_i}$	1	$i = \{L, M, H\}$
$\delta_M(t,s)$, $\delta_H(t,s)$	skill depreciation rates	0	$\forall(t,s)$
$l_M(t,s)$	effectiveness of medium education rate	0.1386	$\forall(t,s)$
$l_H(t,s)$	effectiveness of high education rate	$0.1386 \cdot 1.1$	$\forall(t,s)$
$e(s)$	learning by doing rate	0	$\forall s$
$\theta_L(t)$	low-skilled productivity	0.1	$\forall t$
$\theta_M(t)$	medium-skilled productivity	$0.1 + \alpha_{prod} \cdot 0.7$	
$\theta_H(t)$	high-skilled productivity	$0.1 + (1-\alpha_{prod}) \cdot 0.7$	
$\pi_i(s)$	low-skilled age productivity $\propto c_i exp(\frac{q_1^2}{(s-m_i)^2-q_2^2})$	see figure 3.3	
$L_0(t)$	low-skilled worker entering labour market	1000	$\forall t$
$M_0(t)$, $H_0(t)$	medium/high-skilled worker entering labour market	10^{-6}	$\forall t$
$p_H(u)$	cost-function for high-education efforts	$p_H(u) = \frac{9}{520}u + \frac{3}{260}u^2$	
$p_M(u)$	cost-function for medium-education efforts	$p_M(u) = \alpha_{edu} \cdot p_H(u)$	

Table 3.2: Basic parameter set for the variation analysis

education are different (all other parameters are chosen as in table 3.2). In the left panel of figure 3.4 we assumed $\alpha_{edu} = 0.4$, while the right panel we raised this value to 70%. The left graph now shows that if medium skill education is relatively cheap compared to high skill education, the optimal rates for medium education surpasses the one for high education at all ages by a significant level. This means that the low costs overcompensate the lower productivity of medium skilled worker. As the relative costs for medium education increase, higher education becomes more and more beneficial and the high skill education rates rise above the medium skill, what can by seen the right plot of figure 3.4.

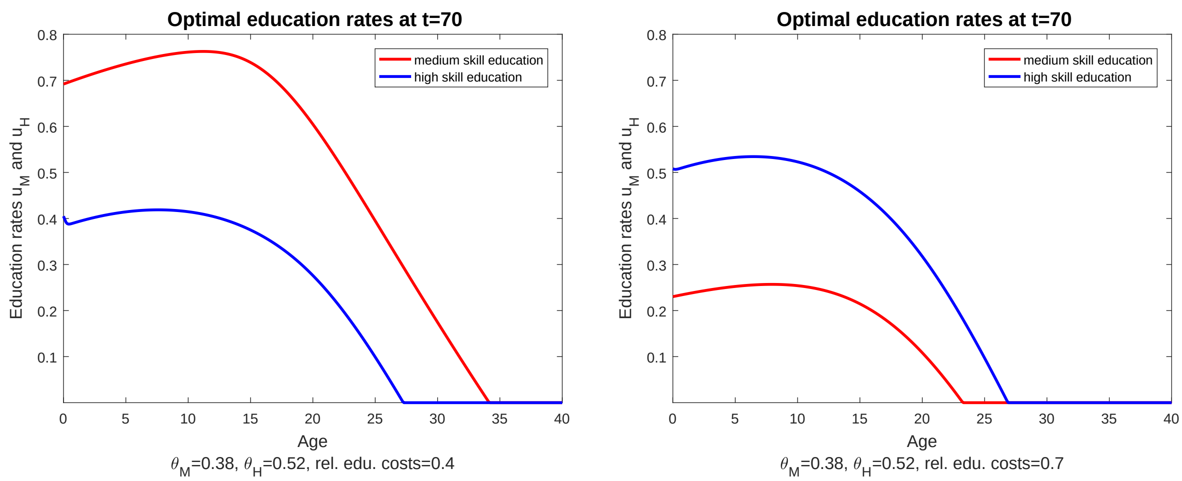


Figure 3.4: Optimal education rates in the extended model under perfect substitutability

For both parameter variations and both types of education, we can observes that the optimal rates are increasing in age for the first years and then decline after reaching their peaks until they drop to zero. All these turning points occur at different ages and it should be noted, that the peaks happen to be earlier in the right panel of figure 3.4 as compared to the left panel. This indicates that an increase of the costs for medium skill education not only leads to lower rates, but also to a shift to younger ages (for these particular parameter values). At the same time also the peak for higher education is reached earlier if medium skill education costs are higher. This might be surprising at first, since the costs for higher education didn't change. Nevertheless, due to the convex cost structure the additional high skilled worker needed, resulting from the lower number of medium skilled worker, are relatively expensive. To make up for these costs education is applied slightly earlier, to profit from the higher productivity for a longer time.

After all we can suspect, that a lot of different mechanisms are effecting the solutions, when certain parameters are changed. In the next section we are trying to filter out the effects of a change in the cost structure (as done in the example above), but also the impact of productivity shifts.

3.5 Relative costs and productivity - sensitivity analysis

We will now try to analyse the effects of different combinations of costs and productivity of medium and high-skilled workers. First we will assume perfect substitutability with respect to skill and age, but we will relax these assumptions later. As in the last section all other parameters as chosen as in table 3.2.

The parameter α_{edu} is varied from 0.1 to 1.5, while α_{prod} takes values between 0 to 0.7 to observe the changes of the optimal controls. This analysis therefore also contains cases, for which medium education is either more expensive than higher education or medium skill worker are more productive or even both. Though such scenarios are quite unrealistic we allow for these variations to illustrate a comprehensive sensitivity analysis. Indeed the upcoming analysis will show, that from the way the model is set up it follows, that under such unrealistic parameter constellations the social planer reacts accordingly and doesn't use the medium education if its unreasonably expensive or higher education, if its unrealistic unproductive.

Since proposition 5 holds, we can choose an (nearly) arbitrary t and analyse the approximated stationary solution. To be able to compare that relatively large number of optimal controls ($8 \cdot 15 = 120$), that follow from this sensitivity analysis, reasonably we need to define indicators which collapse the characteristics of the solutions. First we are going to calculate the following three indicators

- The aggregated education rate, where we integrate the rates over all ages to compare the total educational efforts for different constellations. $\Longrightarrow$ **Aggregated value** $AGG_i(t)$

$$AGG_i(t) = \int_0^\omega u_i(t,s)ds \qquad\qquad i = \{L, H\}$$

- We interpret the rates as density functions with respect to age, i.e. divide the rate at each time point by the aggregated value described above (This approach only makes sense, for optimal rates which are strictly positive at least for a small age-interval). With this interpretation we can calculate the respective mean values and standard deviation of the corresponding distributions. $\Longrightarrow$ **Mean value** $MEAN_i(t)$ and **Standard deviation** $SD_i(t)$ for $i = \{L, H\}$

$$MEAN_i(t) = \int_0^\omega s \cdot \frac{u_i(t,s)}{AGG_i(t)} ds = \frac{\int_0^\omega s \cdot u_i(t,s)ds}{\int_0^\omega u_i(t,s)ds}$$

$$SD_i(t) = \int_0^\omega (s - MEAN_i(t))^2 \frac{u_i(t,s)}{AGG_i(t)} ds = \frac{\int_0^\omega (s - MEAN_i(t))^2 u_i(t,s)ds}{\int_0^\omega u_i(t,s)ds}$$

$$= \frac{\int_0^\omega s^2 u_i(t,s)ds}{\int_0^\omega u_i(t,s)ds} - (MEAN_i(t))^2$$

Figure 3.5 tries to illustrate these indicators in a structured way. In the left column we can see aggregated education rates, the mean age at which education is received and the standard deviation of that age for the medium skill education, in the right column the same is pictured for the high skill education. We have chosen a representation, such that in both plots in the i-th column and j-th row the results from the same parameter constellation are shown. This means if one wants to analyse the indicators of the medium skill education decisions made by the social planer for the parameter constellation in the i-th column and j-th row, the indicators of high skill education for the same overall parameter constellation can be found in the i-th column and j-th row of the figures for higher education. Therefore it should be stressed that the skill-specific productivity on the y-axis is increasing for medium skill education (as we have labelled θ_M), but decreasing for high skill (θ_H is indicated), because the two productivities have to add up to 0.9 as described above. The colour of a square represents a numeric value indicated by a colour bar on the right side of each panel. Colours in a spectrum from dark blue to yellow are used. While dark blue indicates the lowest values of each panel, yellow marks the upper boundary of the range of each indicator. The parameter combinations for which the aggregated value is smaller than 0.1 (this correspond to a constant education rate of 0.0025 over all ages) are marked black, since the other indicators can not be calculated in a meaningful way. It is not surprising taking into account that different skilled workers are perfect substitutable, that for several parameter constellations there are no educational efforts for some skill group at all, since no skill group is essential for the final good production. This picture might change, if we vary the elasticity of substitution with respect to skill. Finally the dashed white lines separate the more realistic parameter set (medium skill education being cheaper than higher education, but high skilled workers being more or equally productive than medium skilled individuals), from the artificial ones.

Aggregated rates

As the first row of figure 3.5 shows, we see that for medium education the highest aggregated education rates are reached, if its productivity is relatively high and the costs are relatively cheap, which is intuitively the most advantageous parameter combination for medium education. A ceteris paribus decline in productivity or increase in costs leads to diminishing aggregated rates. Lastly we can also see that for a big range of parameter constellations there is no educational effort for medium skills at all. For these combinations medium skill is either not productive enough or too costly. Summarising there seems to be a trade-off between productivity and relative costs, which constitutes a fairly intuitive result.

On the other hand this kind of trade-off is much less pronounced for the high skill education (first row, second column). For high levels of productivity ($\theta_H \in [0.66, 0.8]$) the relative costs of high skill education have no significant impact on the aggregated education rates, while a decrease in productivity leads to a decrease in the aggregated education rate. As θ_H drops to around 0.5 or lower the relative education cost have an effect. One should also note, that the range of parameters, where the high-skill education rate drops to zero for all ages is smaller than for medium skill education. Again the productivity seems to be the most decisive factor if education is applied or not. As a last fact it should be mentioned that for high skill education the aggregated rates are between 0 and 22, for medium education the variation is higher and the aggregated rates range from 0 to 100.

Mean age of education

The second row in figure 3.5 shows the mean ages of education. For medium education the figure shows that for productivity values of $\theta_M = 0.31$ upwards the mean ages only varies slightly with increasing costs between 7 and 17 (except for a few combinations in the upper right corner of the plot). Increasing costs thereby lead to a shift of the education focus to younger ages. On the other hand there is also another interesting fact: If for low values of productivity medium education is still applied, because it is relatively cheap, then the education efforts are applied to the oldest age-groups close to retirement (between 25 and 35). This is surprising since education costs are age independent and individuals could profit from their education for a longer time, if investment occurred earlier. Taking a second look we can observe that the explanation lies within the age-specific productivities. For $\theta_M = \{0.1, 0.17, 0.24\}$ medium skill worker are only slightly more productive than low-skill worker, for which reason medium education is too costly in young ages. In older ages on the other hand figure 3.3 shows that the age-specific medium skill productivity surpasses the low skill productivity, what results in medium education being cost-effective and therefore in a positive education rate.

For the high-skill education the mean age again seems to be nearly independent of the relative cost structure and almost solely depends on the productivities. Only when high skilled workers become less productive we can see that relatively expensive high-skill education (relatively cheap medium skill education) leads to a shift of the mean age to older ages. However as for the aggregated values it must be mentioned, that the range is narrower for the high-skill education. The last described shift, is a change in the mean age of two years. As a whole the mean age of high-skill education only varies roughly between 8 and 13 years in the labour market. For medium education as already described the mean age ranges from 5 years into the working life up to 35 years.

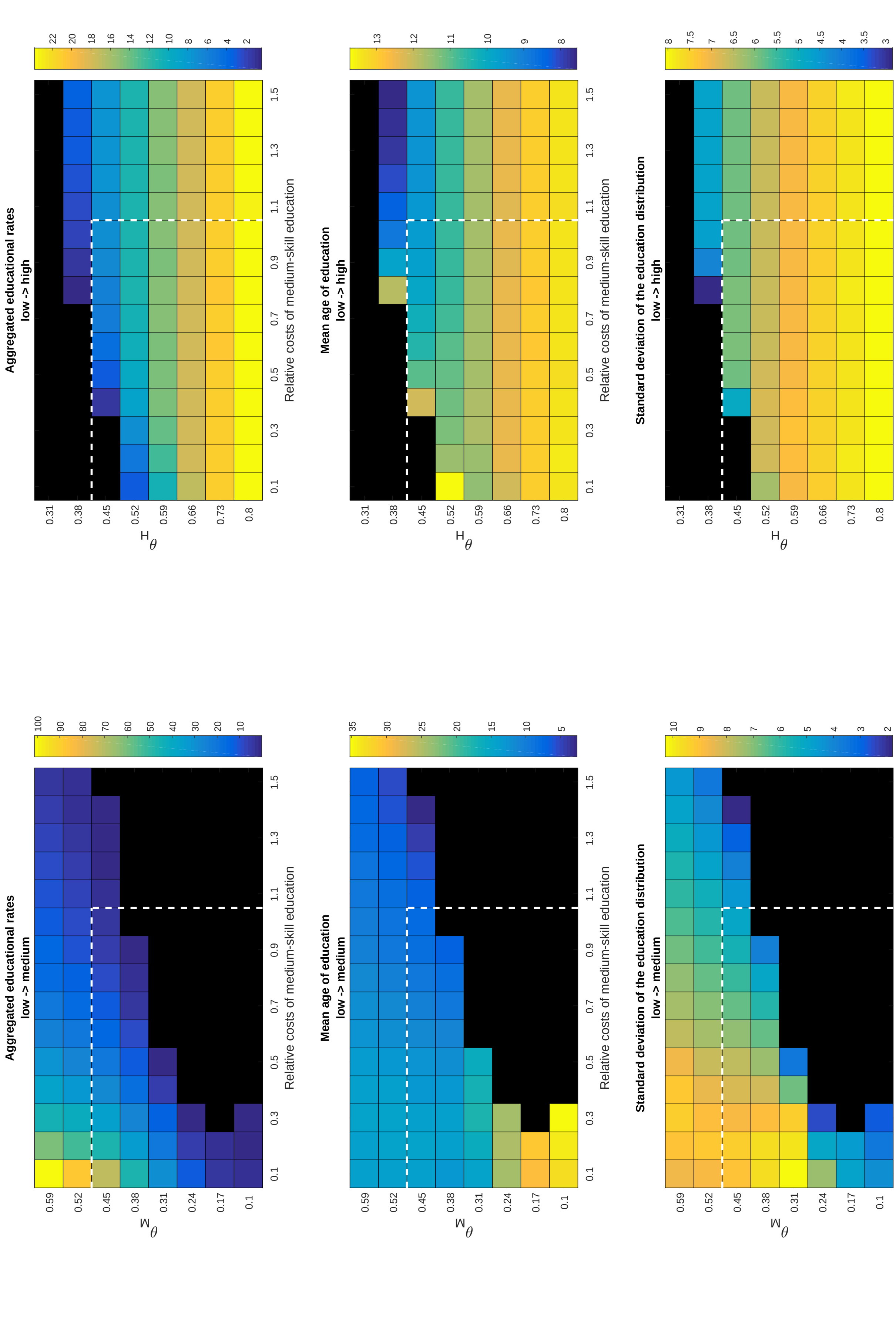


Figure 3.5: Analysis of the effects of changes in productivity and costs on the structure of the optimal education rate in the case of perfect substitutability

Standard deviation of education

Having analysed the mean age at which education is received it is also of interest to see, if the educational efforts are focused on a small time period around this age, or if they are more widely spread. Therefore we have calculated the corresponding standard deviations and they are illustrated in the last row of figure 3.5. For the medium education in the first column we first note that for the low productive, but cheap constellations with the high mean age, the standard deviation is relatively small and lies around 3-7 years. This is not very surprising, since the maximum age is set to $\omega = 40$ and with a mean age between 25 and 35 there is not a big age-range left. For productivity values of $\theta_M = 0.31$ upwards, we see that increasing costs for medium education lead to a smaller standard deviation from around 9 years down to about 4, what means more dense educational efforts around the mean age.

The standard deviation for high education is again nearly invariant to changes in the cost-structure (except for the relatively unproductive, but costly constellation), and decreases with declining productivity of high-skilled labour. This means a lower relative productivity leads to more focused educational efforts near the mean age. Combing the three figures for higher education we can argue that a decline in high skill productivity simultaneously leads to

- lower aggregated educational efforts,
- a shift of the efforts to younger ages,
- concentration of the education rate to a smaller age-group,

while the situation is more complicated and not so simple to summarise in the case of medium skill education.

3.5.1 Imperfect skill substitutability

To support the results of the previous section and show, that they are not critically dependent on the assumption of perfect substitutability, I conducted the same analysis with the elasticity of substitution with respect to skill changed to $\rho = 0.9$. In the case of perfect substitutability for a lot of parameter constellations one form of education was dominant, while the other wasn't used at all. This is relatively intuitive since perfect substitutability allows us to produce the output using only one input and completely neglecting the other inputs at the same time. So changing $\rho = 0.9$ lets us a priori suspect, that the range of parameter sets, for which both types of education are used, increases. Figure 3.6 shows, that these assumptions are being confirmed. The first row of figure 3.6 illustrates that high skill education is now applied to a significant extend for all analysed constellations, except for the case of highly productive and very cheap medium skill education (right panel). Also medium skill education is used for significantly more parameter combinations, however it is not significantly applied if it is relatively expensive compared to the productivity of medium-skill worker (left panel). Analysing the aggregated levels of the two education forms, we see, that there is not much change in the qualitative structure of the solutions compared to the case of perfect substitutability. For medium skill education there is still a trade-off between the relative costs and its productivity, while for higher education the main impact is its productivity, but taking a close look, we see that the effect of the relative costs for fixed levels of productivity is slightly more pronounced than in the case of $\rho = 1$. Paying attention to the dashed white lines, which again separate the more realistic from the relatively unrealistic parameter constellations, nevertheless shows us that both education forms are used to a significant extend for a wider range of unrealistic scenarios, because all skill levels have become necessary for production to at least a small degree.

Studying the mean ages of education in the second row of figure 3.6 we see, that the left plot for medium education exhibits the same patterns as in figure 3.5. If the productivity of medium skill workers is above 0.31, the mean ages mostly lie in a small range between 7 and 15 years of "working age", while for $\theta_M < 0.31$ the mean age increases significantly up to over 30 years. For high skill education it might look like there was a big impact of the change of ρ as the colours have changed, but one has to pay attention to the change of the colorbar/scale on the right side of the plot. While the scale only ranged from 8 to 14 years in the case of perfect substitutable skills, it now covers an interval from 3 to 17. In fact for all parameters where higher education was applied in the case of perfect substitutability too, the mean ages only slightly differ for the new scenario with $\rho = 0.9$. The scale of the plot had to change as for $\theta_H = 0.31$ we obtained a very high mean age for $\alpha_{edu} = 0.4$ and at the same time a very low mean age for $\alpha_{edu} = 0.1$ and $\theta_H = 0.38$.

Regarding the standard deviation, the values might even have changed the least compared to the aggregated and mean values, as the difference stays within 0.3 for almost all these values. For parameter combination, where medium resp. higher education hasn't been applied before, we obtain a rather smooth continuation of the overall pattern (except for the higher education indicators in the case of very cheap and highly productive medium skill education, but as already stated I will not rule out numerical issues in these scenarios). I didn't give any intuitive interpretations of figure 3.6 in this, since we have seen, that the results have not changed much also the intuition behind them stays the same. So the conclusions of the case of perfect substitutability are applicable for $\rho = 0.9$ too.

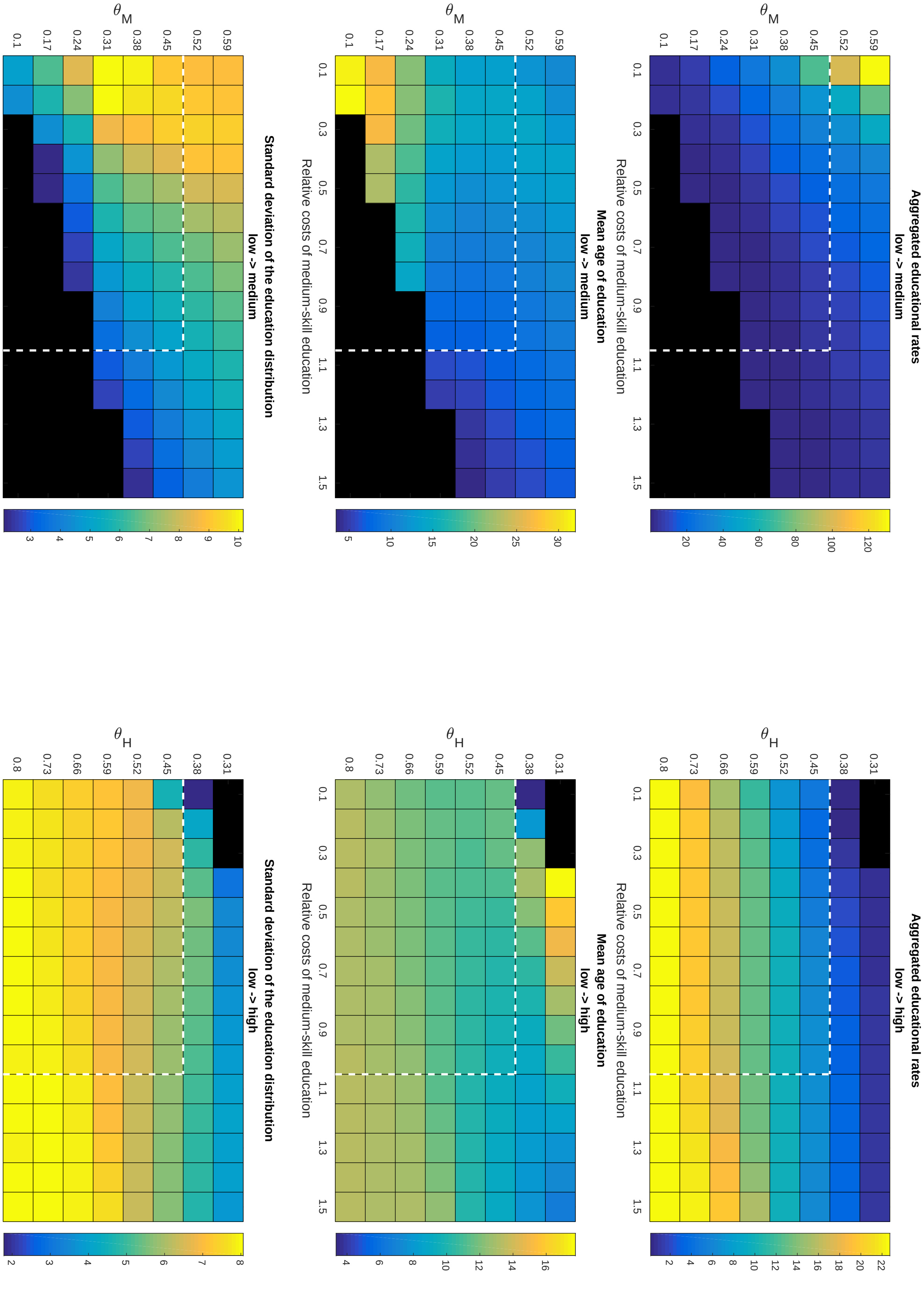


Figure 3.6: Analysis of the effects of changes in productivity and costs on the structure of the optimal education rate in the case of imperfect substitutability with respect to skill ($\rho = 0.9$)

4 | “Lehre mit Matura” Extension

In the model presented in chapter 3 a worker, who has already obtained medium education, only has the chance to become high skilled after he looses his medium skill qualification due to depreciation and becomes low-skilled again. This seems to be a flaw in the model, since there is no reason why a medium skilled worker shouldn’t also obtain further education, especially in a social-planer model, which tries to find the social optimum. The extended model in this section tries to adjust for this observation and enables medium-skilled workers to become high-skilled. I denote this extension the “Lehre mit Matura”- Extension, since Austria introduced a dual education system, where people can complete their medium skill education and at the same time obtain the permission to attain higher education later in their life.

4.1 The model

Figure 4.1 shows the flows in the new model and it basically resembles the model in chapter 3, we only added a second-level education flow $u_Z(t,s)$ between medium and high skill. We do not include a learning by doing transition between those two, because it seems plausible, that higher education, even if it consists of specialisation in the field in which the medium skill was obtained, still needs some kind of training/education. Additionally we set aside a depreciation flow, which would lead to individuals losing their high skill and becoming medium skilled again, since a technological change making the high qualification obsolete is likely to cause the obsolescence of the respective medium skill too and in this case the worker automatically becomes low skilled.

Regarding the costs of the second-level education, we assume them to be of the same structure and additive to the already existing costs[1], i.e.

$$P(t) = \int_0^\omega p_M(s,u_M(t,s))L(t,s) + p_H(s,u_H(t,s))L(t,s) + p_Z(s,u_Z(t,s))M(t,s)ds \qquad (4.1)$$

The mathematical formulation of the optimization problem resembles problem 2 and hasn’t changed at all beside the additional control, the new cost function and the new transition dynamic. Nevertheless all equations of the model are summarised below.

[1] One might argue about the functional form of the cost function, because medium and low skilled worker could possibly receive the same education and this function would not be consistent with the convex cost structure. For example let $p_H = p_Z$ being convex, then $p_H(u) + p_Z(u) = 2p_H(u) < p_H(2u)$. On the other hand, we mainly argued the convex cost structure through the “cream-skimming effect”, meaning that the most talented individuals receive education first leading to lower per capita costs. With this approach the functional form in equation (4.1) is again justified.

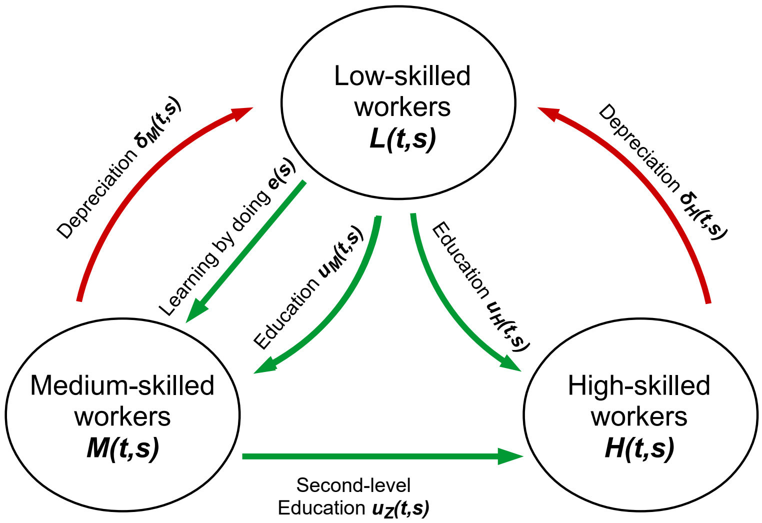


Figure 4.1: Transition flows between different skill groups in the "Lehre mit Matura"-extension

Problem 3

Facing the "Lehre mit Matura" extended model the social planer solves the following optimisation problem

$$\max_{\{u_M(t,s),u_H(t,s),u_Z(t,s))\}} \int_0^T e^{-rt}\Big(Y(t) - P(t)\Big)dt$$

subject to the PDEs

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) L(t,s) = \delta_M(t,s)M(t,s) + \delta_H(t,s)H(t,s) - e(s)L(t,s)- \\ - l_M(t,s)u_M(t,s)L(t,s) - l_H(t,s)u_H(t,s)L(t,s) \tag{4.2}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) M(t,s) = -\delta_M(t,s)M(t,s) + e(s)L(t,s) + l_M(t,s)u_M(t,s)L(t,s)- \\ - l_Z(t,s)u_Z(t,s)M(t,s) \tag{4.3}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) H(t,s) = -\delta_H(t,s)H(t,s) + l_H(t,s)u_H(t,s)L(t,s) + l_Z(t,s)u_Z(t,s)M(t,s) \tag{4.4}$$

the boundary conditions

$$L(t,0) = L_0(t), \qquad M(t,0) = M_0(t), \qquad H(t,0) = H_0(t) \qquad \forall t \in (0,T]$$

$$L(0,s) = L_b(s), \qquad M(0,s) = M_b(s), \qquad H(0,s) = H_b(s) \qquad \forall s \in [0,\omega]$$

the aggregate equations

$$\widetilde{L}(t) = \left(\int_0^\omega \pi_L(s)L(t,s)^{\lambda_L}ds\right)^{1/\lambda_L}, \qquad \widetilde{M}(t) = \left(\int_0^\omega \pi_M(s)M(t,s)^{\lambda_M}ds\right)^{1/\lambda_M} \tag{4.5}$$

$$\widetilde{H}(t) = \left(\int_0^{\omega} \pi_H(s) H(t,s)^{\lambda_H} ds \right)^{1/\lambda_H}, \qquad Y(t) = \left(\theta_L(t) \widetilde{L}(t)^{\rho} + \theta_M(t) \widetilde{M}(t)^{\rho} + \theta_H(t) \widetilde{H}(t)^{\rho} \right)^{1/\rho} \tag{4.6}$$

$$P(t) = \int_0^{\omega} \Big(p_M(s, u_M(t,s)) L(t,s) + p_H(s, u_H(t,s)) L(t,s) + p_Z(s, u_Z(t,s)) M(t,s) \Big) ds \tag{4.7}$$

and the non-negativity constraints

$$u_M(t,s) \geq 0 \qquad u_H(t,s) \geq 0 \qquad u_Z(t,s) \geq 0$$

After having stated the complete model we move on to the optimality conditions for this extended model.

4.2 Optimality conditions

The space-consuming derivation of the optimality conditions can again be found in appendix B.3. The calculations reveal that there are only minor changes compared to the conditions in proposition 4. We obtain similar first order conditions for the newly added control $u_Z(t,s)$ as for $u_M(t,s)$ and $u_H(t,s)$ in the last chapter. The PDE for $\mu_M(t,s)$ contains additional terms, while the dynamics of the other two co-state variables $\mu_L(t,s)$ and $\mu_H(t,s)$ stay the same.

Proposition 6

Let the standing assumptions hold, also for the newly introduced function $p_Z(s, u_Z)$. If $(L, M, H, \widetilde{L}, \widetilde{M}, \widetilde{H}, P, u)$ is an optimal solution for the problem 3 than the system of partial differential equations

$$\begin{aligned}\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \mu_L(t,s) = {} & \Big(e(s) + l_M(t,s) u_M(t,s) + l_H(t,s) u_H(t,s) + r \Big) \mu_L(t,s) - \\ & - \Big(e(s) + l_M(t,s) u_M(t,s) \Big) \mu_M(t,s) - l_H(t,s) u_H(t,s) \mu_H(t,s) + \\ & + \Big(p_M(s, u_M(t,s)) + p_H(s, u_H(t,s)) \Big) - f_L(t,s)\end{aligned} \tag{4.8}$$

$$\begin{aligned}\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \mu_M(t,s) = {} & \delta_M(t,s) \Big(\mu_M(t,s) - \mu_L(t,s) \Big) - l_Z(t,s) u_Z(t,s) \Big(\mu_H(t,s) - \mu_M(t,s) \Big) + \\ & + p_Z(s, u_Z(t,s)) + r \mu_M(t,s) - f_M(t,s)\end{aligned} \tag{4.9}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \mu_H(t,s) = \delta_H(t,s) \Big(\mu_H(t,s) - \mu_L(t,s) \Big) + r \mu_H(t,s) - f_H(t,s) \tag{4.10}$$

$$\mu_i(T,s) = 0 \quad \mu_i(t,\omega) = 0 \qquad \forall i \in \{L, M, H\} \ \forall t \in [0,T] \ \forall s \in [0,\omega] \tag{4.11}$$

with

$$f_L(t,s) = Y(t)^{1-\rho} \cdot \theta_L(t) \pi_L(s) \cdot \widetilde{L}(t)^{\rho - \lambda_L} L(t,s)^{\lambda_L - 1}$$

$$f_M(t,s) = Y(t)^{1-\rho} \cdot \theta_M(t) \pi_M(s) \cdot \widetilde{M}(t)^{\rho - \lambda_M} M(t,s)^{\lambda_M - 1}$$

$$f_H(t,s) = Y(t)^{1-\rho} \cdot \theta_H(t)\pi_H(s) \cdot \widetilde{H}(t)^{\rho-\lambda_H} H(t,s)^{\lambda_H - 1}$$

has a unique solution $(\mu_L(t,s), \mu_M(t,s), \mu_H(t,s))$. *For almost every* $(t,s) \in [0,T] \times [0,\omega]$ $u(t,s) = (u_M(t,s), u_H(t,s), u_Z(t,s))$ *maximizes the Hamiltonian in equation B.37 respectively solves the FOC*

$$-\frac{\partial p_M(s, u_M(t,s))}{\partial u_M} + \Big(\mu_M(t,s) - \mu_L(t,s)\Big) l_M(t,s) \leq 0 \tag{4.12}$$

$$\left(-\frac{\partial p_M(s, u_M(t,s))}{\partial u_M} + \Big(\mu_M(t,s) - \mu_L(t,s)\Big) l_M(t,s)\right) \cdot u_M(t,s) = 0 \tag{4.13}$$

$$-\frac{\partial p_H(s, u_H(t,s))}{\partial u_H} + \Big(\mu_H(t,s) - \mu_L(t,s)\Big) l_H(t,s) \leq 0 \tag{4.14}$$

$$\left(-\frac{\partial p_H(s, u_H(t,s))}{\partial u_H} + \Big(\mu_H(t,s) - \mu_L(t,s)\Big) l_H(t,s)\right) \cdot u_H(t,s) = 0 \tag{4.15}$$

$$-\frac{\partial p_Z(s, u_Z(t,s))}{\partial u_Z} + \Big(\mu_H(t,s) - \mu_M(t,s)\Big) l_Z(t,s) \leq 0 \tag{4.16}$$

$$\left(-\frac{\partial p_Z(s, u_Z(t,s))}{\partial u_Z} + \Big(\mu_H(t,s) - \mu_M(t,s)\Big) l_Z(t,s)\right) \cdot u_Z(t,s) = 0 \tag{4.17}$$

It is clear, that any feasible set of controls and states with $u_Z(t,s) = 0$ for all $(t,s) \in [0,T] \times [0,\omega]$ is also a feasible set for the extended model in chapter 3. This means that the objective value for optimal chosen controls can only increase, when second-level education is introduced. Hence we are going to apply and study the same sensitivity analysis as in the last chapter, to find out for which parameter combinations second-level education is applied and what its effects are, and for which parameter sets the new control isn't used at all.

4.3 Examples

Before performing the sensitivity analysis we will illustrate the impact of second-level education exemplarily for the parameter combinations in section 3.4 respectively in figure 3.4. However we need to define the new parameters $l_Z(t,s)$ and $p_Z(t,s)$ and find plausible values for them first. The rest of the parameters are chosen as in table 3.2 and we will stay within the case of perfect substitutability with respect to skill ($\rho = 1$).

- I decided to set the effectiveness of second-level education $l_Z(t,s)$ twice as high as for high skill education, i.e. $l_Z(t,s) = (0.1386) \cdot (2.2)$. This decision is based on the thought, that people, who already completed some form of education are more focused, when they decide to go through further training. They also have experience in finishing a degree or school and are less likely to be surprised by obstacles, that come with trying to complete their education.
- The per capita education costs of second-level education are set to 90% of the costs for high-skill education, i.e. $p_Z(u) = 0.9 \cdot p_H(u)$. People who are already medium skilled

might skip some parts of the higher education, which low skilled worker receive, but on the other hand it is still a different kind of education[2] and also the opportunity costs are higher for medium skilled worker in education than for a low skilled. Overall second-level education being cheaper, but still relatively close to the costs of high skill education, seemed most plausible.

Figure 4.2 now shows the optimal education rates for the productivity parameters $\theta_M = 0.38$ and $\theta_H = 0.52$ with the relative costs of medium skill education being 0.4 in the left panel and 0.7 in the right panel. First of all we observe, that the newly introduced control (green) is used to the least extend of the three education types, but is still significantly applied in both cases. Additionally we can see that the introduction of second-level education leads to an increase of medium skill education rates at younger ages in both parameter constellations and to a decrease at older ages in the case of relatively cheaper medium skill education. For high skill education we observe just the opposite, namely a slight decrease for both parameter combinations in the optimal education rates at the younger ages compared to the model without second-level education. It is also remarkable that the additional control u_Z stays basically unaffected by changes of the relative costs of medium skill education. For the economic interpretations of the observations I am referring to the sensitivity analysis. There we have again varied the parameters by smaller steps and also analysed different productivity scenarios, what made it easier to derive the intuitive reasons behind the effects.

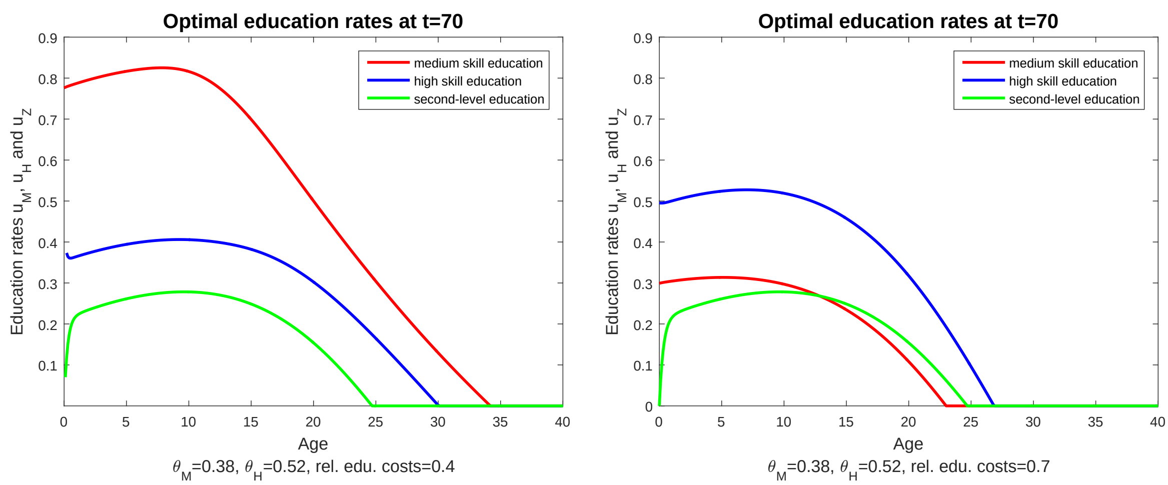


Figure 4.2: Optimal education rates in the “Lehre mit Matura”-Extension in the case of perfect substitutability

However in no case should the large medium-skill education rates be misinterpreted and directly taken as indicator for the importance of medium skilled workers in production. To make this more clearly we will take a look at the corresponding state variables, i.e. the composition of labour resulting from the shown education rates. Figure 4.3 shows the skill composition of the population of each age-group. The solid lines correspond to the “new” optimal controls

[2]See section 3.1, vocational training vs. general education.

presented in figure 4.2 above, while the dashed lines relate to the "old" optimal controls in figure 3.4 from the previous extended model in chapter 3. Especially the left plot shows that the possibility of second-level education can lead to a switch from an economy with a majority of medium skilled workers to one being heavily high-skill abundant. For all age-groups being six years into the labour market or longer high skilled workers mark the biggest share of the three skill levels. Also of all workers with 25 years or more working experience, more than 75% percent are high skilled. At the same time this number was below 40%, when second-level education was not an available option for the social planer. In the right plot we also see a shift towards high skilled workers, despite them being already the primary choice of labour before the introduction of second-level education. Meanwhile the number of medium skilled worker in the new model reaches only a slightly higher value than the low skilled workers for the ages above 20. After all we have seen that the impact of the new control on the labour composition can be considerably higher, than what the education rates might let suspect at first sight.

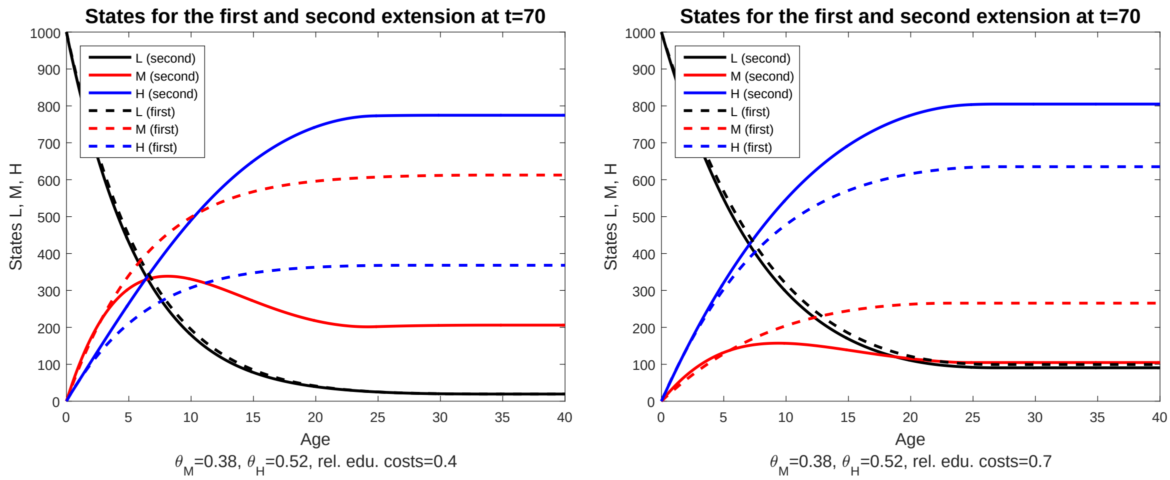


Figure 4.3: Optimal stationary states in the "Lehre mit Matura"-Extension compared to the extended model in chapter 3, both for the case of perfect substitutability

4.4 Sensitivity analysis

We will now perform the same sensitivity analysis for our new extension as in section 3.5. We chose the newly appearing parameters $l_Z(t,s)$ and $p_Z(s,u_Z)$ as described in the last section and focus on the case of perfect substitutability with respect to skill levels ($\rho = 1$). Before we conduct the sensitivity analysis itself, we will show for which parameter combinations the possibility of second-level education has an impact on the decisions of the social planer. Since we are again in the case of perfect substitutability and we therefore can focus on the stationary controls appearing for $\bar{t}$ "in the middle" of the observed time frame, we will compare the stationary output $Y(\bar{t})$ and costs $P(\bar{t})$. The difference between the two $SB(\bar{t}) = Y(\bar{t}) - P(\bar{t})$ can be interpreted as the long-run social benefit, which can be generated each time period. Figure 4.4 shows the relative gains G in social benefits through the introduction of second-level education with the values SB_{old} from the extended model in chapter 3 being the base values, i.e.

$$G = \frac{SB_{new} - SB_{old}}{SB_{old}}$$

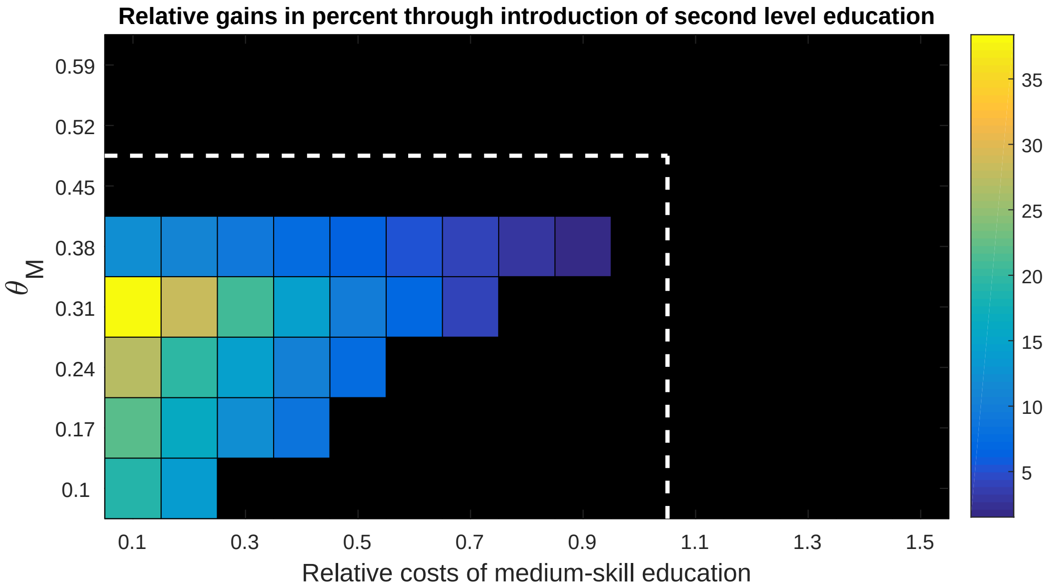


Figure 4.4: Relative gains in percent in the social benefits (output minus costs) through introduction of second-level education

This gain should theoretically always be positive as already mentioned above. Due to numerical errors this does not hold for our simulations[3], so we set all negative values equal to zero. Further examination showed that all negative values stayed below 0.3% in absolute terms. To be consistent we did the same for all positive gains below 0.3%, since we can not rule out, that they originate from numerical errors too. 0.3% now might seem like an arbitrary threshold, but the actual numbers reveal that the smallest positive relative gain bigger than 0.3% is even above 1.5%. As a result 0.3% can be taken as a reasonable threshold value to distinguish significant relative gains in social benefits, and the numerical inaccuracies should not be of bigger concern.

In Figure 4.4 we see, that for a big share of the examined parameters there is no gain in social benefits through the implementation of second-level education. Therefore also the optimal education rates do not change in these cases. We can spot three main reasons for this to happen:

- For values of $\theta_M \geq 0.45$ the skill-related productivity of medium skilled worker is either equal or higher then the one for high-skilled worker. Hence there is no incentive to enhance the skill level of medium skill workers in younger ages and the age-specific productivity difference in older ages (see figure 3.3) is not big enough to invest in additional training.

[3] One should not forget, that we solve the PDEs numerically and also use an optimisation algorithm to acquire an approximation for the optimal solution.

- If the costs of medium skill education are equal or even higher than the costs of high skill education, it is intuitive to directly apply high-skill education to low skilled people, if there is a need for high skilled workers. One might think that the convex cost structure would lead to a balanced application of high and second-level education, but due to the low effectiveness of medium-skill education this effect is cancelled out.

- For the remaining parameter set the combination of productivity and relative costs of medium skilled education is decisive. We should first mention the most intuitive reason for the application of second-level education is a temporary productivity gain. An individual becoming medium skilled at time a and afterwards high skilled at time b has a higher productivity on $[a, b]$ than a person directly receiving higher education at the later time point b. This productivity gain must outweigh the additional education costs to be profitable. If either the costs of medium education are too high, or the productivity of medium skill is too low, the two step education process does not pay off.

Keeping in mind the background of the last reasoning, it is only logical that a ceteris paribus decline in relative medium-skill education costs leads to a higher gain, if the new education possibility is already applied in the first place. An increasing productivity θ_M on the other hand first leads to higher gains (since the two step educational process becomes more profitable), but after the value of $\theta_M = 0.31$ it becomes more and more beneficial to keep the workers at the medium skill level, because the productivity differences become smaller and smaller. Finally the dashed white lines again represent the border of the more realistic parameter constellations as in chapter 3. Figure 4.4 now illustrates that there is no advantage of having a possible two step education process for all relatively unrealistic parameter sets, while for a large share of the others it is profitable to make use of the new education possibilities.

In the next step we are going to look again at the aggregated education rates, the mean age at which education was received and the standard deviation of this distribution . As we have seen, that for many parameter constellations the solutions did not change compared to the previous extended model in chapter 3, figure 4.5 focuses on the smaller range of parameter sets, where we have discovered some new implications. In the left column the results for medium-skill education are shown, in the middle column for second-level education and in the right column for higher education.

Aggregated rates

In the left panel in the first row we see that there are now additional combinations with significant positive medium skill education rates compared to figure 3.5. Additionally the aggregated rates are clearly higher than before in most cases, especially for the lowest values of θ_M. This is due to the new possible two-step-education. Examining the middle plot we can identify very high rates of second-level education, which confirm the distinctive impact of the new control in these cases. There is also an significant increase in the aggregated medium skill education rates for $\theta_M = 0.31$, but not so pronounced. The rates of second-level education decline with increasing θ_M as keeping the workers medium skilled becomes more favourably.

The rates stay unaffected by changes in the relative costs of u_M.

This additional source for high skilled workers leads to lower high-skill education rates, when second-level education is used to a high extend. The aggregated rates drop by around 20% for these parameter sets. The other values however remain relatively unchanged, as the new control is either not used at all or only to a small degree. Overall these results support the explanations about the investment into second-level education given in section 4.3.

Mean age

Regarding the average age of education obtained, we end up with a very interesting result for combinations of small θ_M and low costs. While the mean age of medium education drops drastically, the mean age for higher education increases significantly. The possible second-level education thus leads to higher rates of medium education for younger people, which then after a short time (the mean age for second-level education stays relatively constant at around 14 years for most constellations) obtain further education and become high skilled. As a result high-skill education for young individuals looses importance, consequently this leads to a higher mean age for this form of education. These interpretations also correspond to the analysis of figure 4.2.

Standard deviation

The rigorous drop in the mean age of medium-skill education for the parameter scenarios described above simultaneously leads to an increase of the standard deviation. For the other parameter constellations the changes are minor, but it should be noted, that for the cases with relatively cheap medium skill education and $\theta_M = 0.31$ we obtain even a small decrease in standard deviation compared to chapter 3. For higher education no significant differences in the standard deviation - compared to chapter 3 - are observed across the different parameter variations. The middle panel in the last row shows that the standard deviation for the second-level education stays constant for ceteris paribus changes in the relative costs, but decreases with increasing medium skill productivity. This corresponds to the behaviour of the aggregated rates for this education type and implies that as θ_M increases, the second-level educational efforts become also less widely spread around the (relatively constant) mean age.

Summary

Summarising we have seen that second-level education is applied for a big share of the most realistic parameter constellations (medium skill education being cheaper than high skill, but less productive at the same time). We also have observed that the new control together with higher medium education replaces the classic high skill education partially for favourable parameter constellations, but also supports and strengthens the overall relevance of high skilled workers. All this together is likely to lead to a higher share of high skilled individuals in the economy (as seen in figure 4.3) and to gains in the long-run social benefits, which can be distributed to the people.

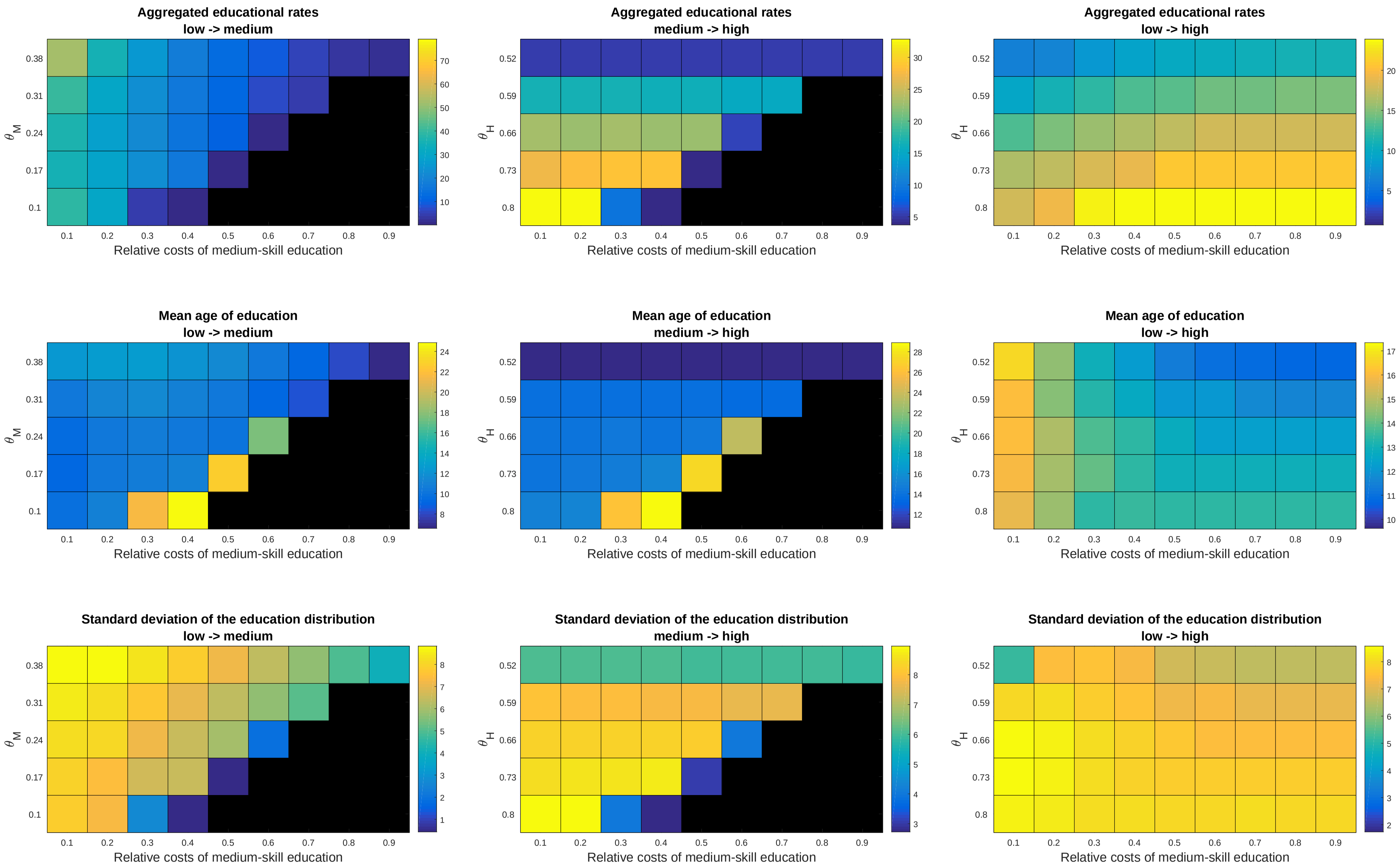


Figure 4.5: Analysis of the effects of changes in productivity and costs on the structure of the optimal education rate in the case of perfect substitutability

4.5 Demographic change

We will now come back to the analysis of the impact of demographic change for different elasticities of substitution across age. To keep the number of different combinations to a traceable level we not only assumed perfect substitutability with respect to skill ($\rho = 1$), but also with respect to age for the low skilled workers ($\lambda_L = 1$). Concerning the productivity and cost parameters I have chosen $\theta_M = 0.38$, $\theta_H = 0.52$ and 40% for the relative costs of medium skill education, exactly as in one of the examples in section 4.3. For λ_M and λ_H we chose either 0.9 or 0.1, meaning we analyse different combinations of well and badly substitutable medium and high skill workers. Regarding the population development we examine as in the basic model in chapter 2 an either constant or exponentially increasing or decreasing population size. The rate of population growth is again $\gamma = 0.0072$ or $\gamma = -0.0072$. Each plot in figure 4.6 now illustrates the three population scenarios for one combination of λ_M and λ_H. The left upper plot shows the case of both elasticities of substitution being close to one and we see that the optimal education rates are similar to the left panel in figure 4.2. The dashed lines show the optimal controls for an increasing population and we basically get the three results

- Medium skill education shifts towards older ages due to lower education rates in the middle ages and higher in the oldest ages.
- Second-level education stays nearly unaffected by demographic changes.
- High skill education increases for all ages above 10.

For a decreasing population size (dotted lines) the results are just vice-versa. As all controls are basically identical in the first years for all scenarios, the number of young medium and high skilled workers increases (for a growing population). To maintain the optimal mix of workers of different age-groups, the social planer increases the high skill educational efforts for older ages to raise their numbers too. It seems like for this parameter constellation it is more important to keep the age composition of high skilled workers, since the lower medium skill education rates at the middle ages lead to an even more pronounced excess of young workers in this sector. Nevertheless these results are quite positive, because we see, that population growth leads to higher educational efforts per capita and therefore to a better educated society as a whole.

The solid lines in the lower left panel of figure 4.6 show the effects of medium skilled labour becoming badly substitutable across age. Again it is intuitive that the medium skill education rate increases sharply approaching the age of 0. Since medium skilled workers are badly substitutable across age, the social planer already needs a significant amount of them at the youngest ages. As only an insignificant amount of workers enters the labour market already equipped with this skill, the medium skill education rate has to increase noticeably.[4]. On the other hand high skill education remains almost unchanged by this shift in substitutability, but the rates of second-level education increase for the first 10 to 15 years, what might be due to a surplus of medium skilled workers after the first few years. Analysing the effect of demographic

[4]We obtained similar results in the basic model in chapter 2

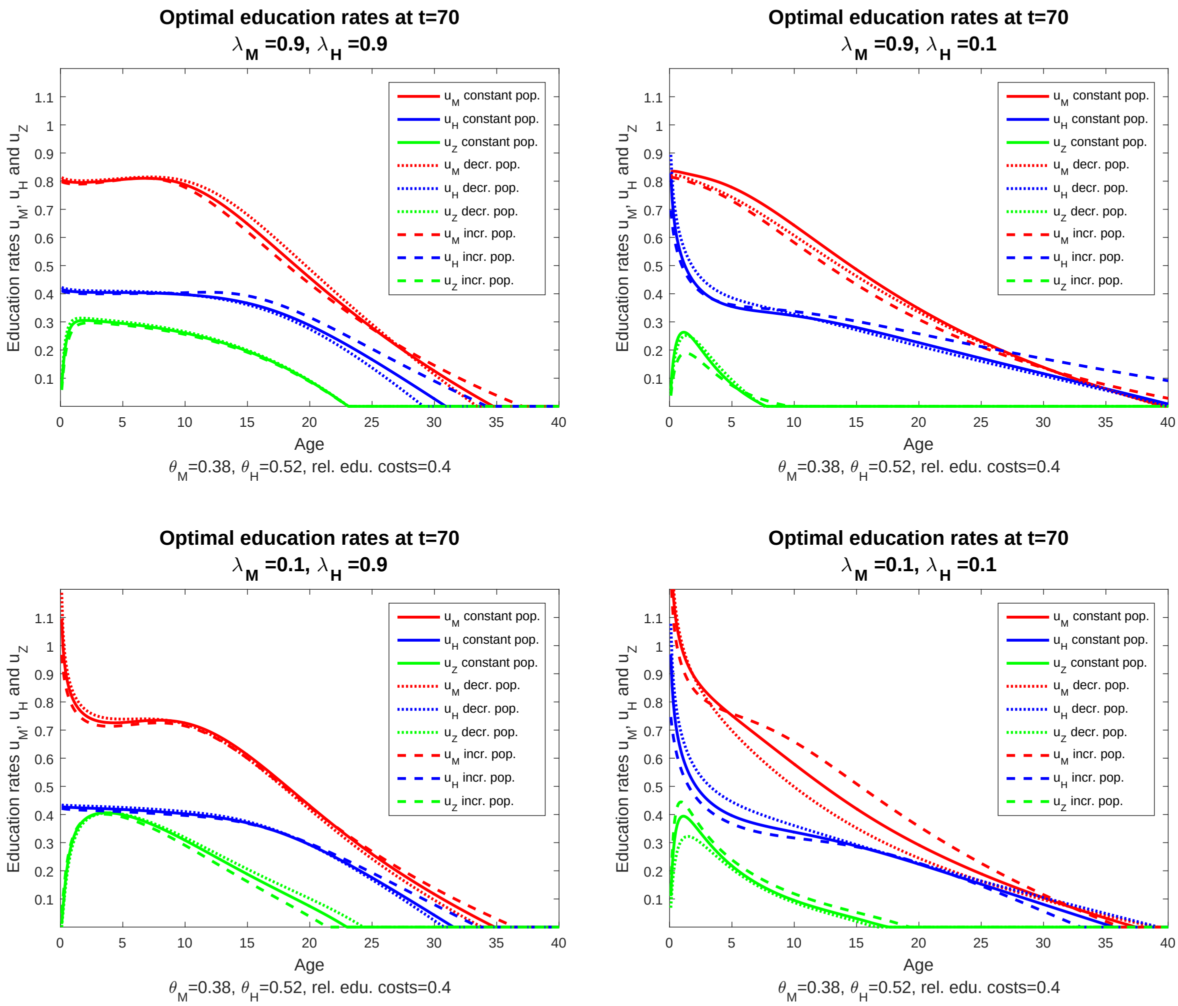


Figure 4.6: Optimal education rates in case of a change of population size

changes we can see that the increase of higher education for a growing population becomes less pronounced than in the last case and is now only significant for ages above 20. Also second-level education rates decrease for ages above 5 if the social planer faces a rising population size. The slight increase in medium skill education rates for the oldest ages is present in this case too, but now for the youngest ages a small decrease can be observed. For a shrinking population we obtain just the opposite results for all education types. Summarising we have seen, that for badly substitutable medium skilled workers an increasing population size does not lead to such a positive societal reaction as in the last case with both skills being good substitutes across age.

In the upper right panel the optimal rates for this time badly substitutable high skilled workers are illustrated, while medium skilled labour is again well substitutable. The interpretation of the sharply increasing high skill education for the youngest ages is the same as for medium skill education in the last scenario, but this time the resulting high number of young high skilled individuals leads to the disappearance of second-level education for people with more than 10 years of working experience. If additionally the population is increasing, the second-level education rates for the youngest ages drop further and high-skill education gets strengthened at the middle and older ages, yet again to maintain the overall age composition of high skilled workers. For a decreasing population size on the other hand the necessary higher number of high skilled people is not reached through an increase of second-level education, but directly through an increasing higher education rate. For medium skill education we end up with the somewhat surprising result, that in both population scenarios (increasing and decreasing) the respective education rates are lower than for the constant population case.

This is a surprising result, we haven't obtained before. Trying to interpret the result economically we first have to state, that under demographic changes the social planer primarily aims to re-establish the age-specific composition of high skilled labour, because medium skilled workers are better substitutable across age. To achieve this in case of an increasing population the medium skill education rates for the younger ages drop to reduce the number of candidates for the two step education process right in the beginning. For the middle ages the rates are lower to indirectly increase the number of people receiving higher education. For a decreasing population on the other hand the latter mentioned effect supports higher education at the younger ages. During the middle ages the medium skill education rates converge to the rates of the constant population case, so for ages above 20 the difference is insignificant.

The most distinct effects of demographic change can be found in the case of both types of educated labour being badly substitutable. The corresponding plot is depicted in the bottom right panel of figure 4.6. For an increasing population size both medium and higher education rates are lower in the first years, but while medium education becomes more frequently used for all groups older than 5 years, the high skill education rates stay on a lower level (compared to the constant population case). The diminishing number of high skilled workers is at least partially counterbalanced by a higher rate of second-level education for all age-groups, but the total effect on the share of high skilled labour can not be determined by solely inspecting the education rates. If in the other case the population size is shrinking, the results are just complementary to the case of population growth. We obtain higher rates of high skill education, while second-level education decreases. The medium education at the same time is significantly lower for most age groups and only insignificantly larger for the youngest.

To summarise the conducted analysis we have seen, that the effects of demographic changes are more complicated to interpret economically than in the basic model and it seems like multiple mechanisms affect the optimal education rates and their impacts are hard to separate. Also, as shown in the upper right panel, an increasing or diminishing population does not always directly lead to opposite reactions of the social planer, because he has more alternatives to choose from to end up at the desired results. Of course most of the time, when a model gets extended the results become multifaceted, and we significantly extended the basic model in chapter 2 by one additional state and two additional controls, but the effects these adaptations had are still remarkable.

4.6 Discussion

As it should always be the aim to make an economic model as realistic as possible, while keeping it simple enough to be analytically solvable and meaningful economically interpreted, we now want to discuss if the additional difficulties the extended models brought with them are overcompensated by their higher degree of realism and generality.

This question directly leads to the first advantage of the extended models. One target of introducing an heterogeneous control model with age as a continuous variable was to obtain a more generally formulated model. So why shouldn't we aim for the same, when it comes to the modelling of skills or education? Of course assuming three skill levels is far away from the most general case, what would be a continuous skill level, but the introduction of the third level alone significantly increased the realism of the model. As already stated in the motivation of chapter 3 most empirical data is presented or can naturally be collapsed into three levels of skill or education. This also makes the extended model better suitable to fit with actual data. To fit the basic two skill level model, one would have to choose a more arbitrary threshold collapsing the actual data into two skill levels.

On the other hand we were able to obtain more analytical results for the basic model and numerical experiments can never completely replace analytical derivations. Additionally the effects of demographic changes were more clear cut and easier to interpret economically. However we were still able to transfer the most basic analytical results to the extended models and received intuitive outputs in the sensitivity analyses. The extended models are also able to reflect mechanisms, which were not covered by the basic model due to its simpler formulation. The increasing number of parameters on the one hand makes it harder to find appropriate values for all of them and adds another dimension, which can affect the solution and should be kept in mind. On the other hand the additional degree of freedom makes it possible to fit the model better to existing data.

Taking all these arguments into account we end up with the basic model being easier to handle and more appropriate to derive general conclusion about the impacts of different factors. The extended models then again are more suitable to reflect real data due to their higher degree of realism and might be more useful for deriving outcomes for more explicitly formulated questions and are therefore more policy relevant.

In the end, all the insights we got from the sensitivity analyses and the changes in the education policies after the introduction of second-level education very well justify the extensions of the basic model. Together with further adaptations mentioned in the conclusions in chapter 6 the model should become more and more realistic and should be of interest for further research, e.g. case studies with the model being fitted to a certain country or being implement into a bigger more general economic model.

5 Theoretical background for age-structured optimal control models

Mathematical tools for the solution of age-structured optimal control models can be found in Feichtinger et al. (2003). In their paper the authors developed a new Pontryagin's type maximum principle for a general form of age structured optimal control models, since the classic theory of Pontryagin about dynamic optimal control models is not applicable to these kind of models.

5.1 The Model

The authors analyse models of the following standardised form.

$$\min_{u(\cdot,\cdot),v(\cdot),w(\cdot)} \int_0^\omega l(a, y(T,a))da + \int_0^T \int_0^\omega L\Big(t, a, y(t,a), p(t,a), q(t), u(t,a), v(t), w(a)\Big) dadt \tag{5.1}$$

subject to the following differential, integral and boundary equations:

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) y(t,a) = f(t, a, y(t,a), p(t,a), q(t), u(t,a)) \tag{5.2}$$

$$p(t,a) = \int_0^\omega g(t, a, a', y(t,a'), u(t,a'))da' \tag{5.3}$$

$$q(t) = \int_0^\omega h(t, a, y(t,a), p(t,a), q(t), u(t,a))da \tag{5.4}$$

$$y(0,a) = y^0(a, w(a)) \tag{5.5}$$

$$y(t,0) = \varphi(t, q(t), v(t)) \tag{5.6}$$

and for the controls:

$$u(t,a) \in U \quad , \quad v(t) \in V \quad , \quad w(a) \in W$$

with U, V, W being subsets of finite-dimensional linear normed spaces (e.g. Euclidean vector spaces). Furthermore they consider the following properties regarding the dimensions for all variables

Time parameter	$t \in [0,T] \quad , \quad a \in [0,\omega] \quad , \quad D := [0,T] \times [0,\omega]$
States	$y : D \mapsto \mathbb{R}^m \quad , \quad p : D \mapsto \mathbb{R}^n \quad , \quad q : [0,T] \mapsto \mathbb{R}^r$
Controls	$u : D \mapsto U \quad , \quad v : [0,T] \mapsto V \quad , \quad w : [0,\omega] \mapsto W$
Other functions	$l : [0,\omega] \times \mathbb{R}^m \mapsto \mathbb{R} \quad , \quad L : D \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^r \times U \times V \times W \mapsto \mathbb{R}$
	$f : D \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^r \times U \mapsto \mathbb{R}^m$
	$g : D \times [0,\omega] \times R^m \times U \mapsto \mathbb{R} \quad , \quad h : D \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^r \times U \mapsto \mathbb{R}^r$
	$y^0 : [0,\omega] \times W \mapsto \mathbb{R}^m \quad , \quad \varphi : [0,T] \times \mathbb{R}^r \times V \mapsto \mathbb{R}^m$

I will now skip the standing assumptions about characteristics of each function (measurability, continuity, Lipschitz-continuity, ...) since they are quite technical and focus on the optimality conditions and their applications rather than on the proof of their optimality.[1]

The equations (5.1) to (5.6) can be interpreted in the following way:

- The term $\int_0^\omega l(a, y(T,a))da$ represents the residual value of the state $y(T,\cdot)$ at the end of the time horizon. For example within a problem regarding the optimal investment structure for machines it could describe the value, which can be generated by selling all remaining machines at the end of the time horizon.

- $\int_0^T \int_0^\omega L(t,a,y(t,a),p(t,a),q(t),u(t,a),v(t),w(a))dadt$ contains the accumulated value over the considered time span. In general the function L depends on all state and control variables, because it covers the costs, that arise through the use of the controls and the corresponding state, but also the resulting benefits at time t and for each age-group a. Integrating over all age-groups and every time point in the considered time horizon finally leads to the value, which should be considered for optimisation.

- Equation (5.2) describes the dynamics of the state variable. Since time and age move at the same speed $\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) y(t,a)$ captures the change of the state as time and age simultaneously pass for a marginal unit. It is intuitive, if one thinks of the mortality of a population. To analyse the size of a cohort one has to respect that as time changes, so does the age of the cohort.

- The function $g(\cdot)$ in equation (5.3) can be viewed as a kernel function. It can measure effects cohorts can have on each other. E.g. in models about the optimal drug control like Almeder et al. (2004) $p(t,a)$ describes the reputation a drug has at time t in a cohort at age a. A drug is less likely to be used by teenagers (has a lower reputation) if their parents generation is using it to a high amount and more likely (higher reputation) if

[1] The conditions can be found in Feichtinger et al. (2003, p. 51).

people of the same or slightly higher age are taking it. This effect can be captured by the function $g(\cdot)$.

- Equation (5.4) describes just one aggregated value for the whole population at time t. Examples could simply be the total population size or the total fertility of a population.
- y^0 represents the composition of the population[2] at the beginning of the time horizon. In most models these starting values are assumed to be exogenous $y(0,a) = y^0(a)$, $\forall a \in [0,\omega]$, but Feichtinger et al. (2003) allowed for an additional control in their model influencing them. An example for the case of a controllable starting composition is a new firm, which can decide how many machines of which age they want to start their business with.
- $\varphi(\cdot)$ in equation (5.6) defines the size of the cohort of age 0 at each time-point t, which in general can be controlled. In human populations the number of individuals entering with age 0, is subject to the total fertility of the population, but could also be affected through migration. Considering a population of machines, $y(t,0)$, the number of new machines, which have not yet been used for production, is likely to be defined by the investments in new machines.

I also want to make a slightly technical remark about equation (5.2), which should be stressed in my opinion. $\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) y(t,a)$ should be viewed as a single operator applied to $y(t,a)$, as it directly describes the derivative in the direction $(1,1)$. It can be a huge restriction using the sum of the partial derivatives $\left(\frac{\partial y(t,a)}{\partial t} + \frac{\partial y(t,a)}{\partial a}\right)$ instead, because this would indicate that the solution $y(t,a)$ has to be continuously differentiable with respect to both arguments. Staying with the first term uses the big advantage of the parallelism of time and age and leads to the fact that the solution only has to be differentiable along the characteristics with $t - a = const.$ More formally this can be found in definition 1 in Feichtinger et al. (2003, p. 51).

5.2 The necessary optimality conditions

As already mentioned the authors derive necessary optimality conditions, but not sufficient ones. This implies all theorems below only hold under the following condition.

Condition 1

There exists an optimal solution $(\hat{y}, \hat{p}, \hat{q}, \hat{u}, \hat{v}, \hat{w}) \in L^\infty(D, \mathbb{R}^m) \times L^\infty(D, \mathbb{R}^n) \times L^\infty([0,T], \mathbb{R}^r) \times L^\infty(D, U) \times L^\infty([0,T], V) \times L^\infty([0,\omega], W)$

To make the formulas below better readable and the notation more convenient the function arguments, which are fixed at an optimal "hat" value, are skipped, e.g.

$$f(t,a) := f(t, a, \hat{y}(t,a), \hat{y}(t,a), \hat{p}(t,a), \hat{q}(t), \hat{u}(t,a)) \qquad \text{resp.}$$

[2]The term should always be seen in a broader sense, since we could also use it to describe a set of machines, as they, like a human population, can also be distinguished by their age.

$$f(t,a,u) := f(t,a,\hat{y}(t,a),\hat{y}(t,a),\hat{p}(t,a),\hat{q}(t),u).$$

As in the classic Pontryagin maximum principle, the co-state or adjoint variables play an important role. (ξ,η,ζ) shall be the adjoint variables for (y,p,q) and have the same domains and image spaces as y, p and q respectively, but while (y,p,q) are column-vector functions, (ξ,η,ζ) are row-vector function.

Definition 1 (Adjoint system)

The co-state functions (ξ,η,ζ) solve the following adjoint system

$$-\left(\frac{\partial}{\partial t}+\frac{\partial}{\partial a}\right)\xi(t,a) = \nabla_y L(t,a) + \xi(t,a)\nabla_y f(t,a) + \\ + \zeta(t)\nabla_y h(t,a) + \int_0^\omega \eta(t,a')\nabla_y g(t,a',a)da' \tag{5.7}$$

$$\xi(T,a) = \nabla_y l(a,\hat{y}(T,a)) \tag{5.8}$$

$$\xi(t,\omega) = 0 \tag{5.9}$$

$$\eta(t,a) = \nabla_p L(t,a) + \xi(t,a)\nabla_p f(t,a) + \zeta(t)\nabla_p h(t,a) \tag{5.10}$$

$$\zeta(t) = \xi(t,0)\nabla_q\varphi(t) + \int_0^\omega \nabla_q L(t,a) + \xi(t,a)\nabla_q f(t,a) + \zeta(t)\nabla_q h(t,a)da \tag{5.11}$$

To characterise the optimal solution $(\hat{u},\hat{v},\hat{w})$ we first define three Hamiltonian functions.

Definition 2 (Hamiltonians)

The initial, boundary and distributed Hamiltonian functions are defined by

Initial $$H_0(a,w) := \xi(0,a)y^0(a,w) + \int_0^T L(s,a,w)ds \tag{5.12}$$

Boundary $$H_b(t,v) := \xi(t,0)\varphi(t,v) + \int_0^\omega L(t,b,v)db \tag{5.13}$$

Distributed $$H(t,a,u) := L(t,a,u) + \xi(t,a)f(t,a,u) + \tag{5.14}$$

$$+ \int_0^\omega \eta(t,a')g(t,a',a,u)da' + \zeta(t)h(t,a,u) \tag{5.15}$$

In the classic theory of optimal control problems (see Feichtinger and Hartl (1986, p. 16-19)) the Hamilton function is introduced to state the necessary optimality conditions in a more compact way. This classic Hamilton function has a relatively intuitive interpretation, because it captures the trade-off between the direct effect of the controls on the objective function and the indirect effect on the objective value through the effects on the state variables.

As we have in general the two controls $w(a)$ and $v(t)$ on the boundaries[3], which aren't present in the classic theory, we obtain two additional Hamilton functions (the initial and the boundary in definition 2). Nevertheless all three Hamiltonians in definition 2 can be interpreted similar to the classic Hamilton function.

- The distributed Hamiltonian is the sum of the 4 effects the control $u(t,a)$ has at each time t for each age a.
  - $L(t,a,u)$ is the value of the objective function at (t,a), so this term captures the instantaneous effect of the control on the objective function reps. the objective values.
  - $f(t,a,u)$ describes the change of the state-variable $y(t,a)$, when a certain value for the control u is chosen. This value is multiplied with the shadow value of the state-variable resulting in the whole term describing the long-run effect of the control choice on the objective value through the state variable.
  - The integral term $\int_0^\omega \eta(t,a')g(t,a',a,u)da'$ results from "adding up" the effects of $u(t,a')$ on the state-variable $p(t,a)$, while again multiplying it with its shadow-value to obtain the effect on the objective value.
  - The fourth term can basically interpreted the same way as the second one.
- The initial Hamiltonian sums up the effect of the initial control $w(a)$ for each age a. This effect is again the sum of two different parts
  - The term $\xi(0,a)y^0(a,w)$ results from calculating the initial state for a chosen control w and then multiplying it with the shadow-value of the state at $t=0$ to obtain the indirect impact on the objective function.
  - The integral $\int_0^T L(s,a,w)ds$ describes the aggregated direct effect on the objective value the initial control $w(a)$ has over the whole time horizon $[0,T]$.
- The boundary Hamiltonian can be treated similarly to the initial Hamiltonian with the time-point t being chosen fixed, but the impact of the boundary control $v(t)$ on all ages is considered.

[3] I.e. the controls influencing the initial states of all ages at $t=0$ and the boundary states for $a=0$ at every time point t

Now we can finally state the maximum principle.

Theorem 1 (Pontryagin's maximum principle)

Under condition (1) the adjoint system of definition (1) has a unique solution ξ, η, ζ *and for a.e.* $t_0 \in [0, T]$*,* $a_0 \in [0, \omega]$ *and* $(t, a) \in D$ *holds*

$$\frac{\partial H_0(a_0, \hat{w}(a_0))}{\partial w}(w - \hat{w}(a_0)) \geq 0 \qquad \forall w \in W$$

$$\frac{\partial H_b(t_0, \hat{v}(t_0))}{\partial v}(v - \hat{v}(t_0)) \geq 0 \qquad \forall v \in V$$

$$H(t, a, u) - H(t, a, \hat{u}(t, a)) \geq 0 \qquad \forall u \in U \tag{5.16}$$

The proof can be found in Feichtinger et al. (2003) and marks the main and biggest part of their paper. How to apply this theorem to a certain model is shown in the other chapters. Still we want to shortly summarise the intuition behind the optimality conditions.

The first two conditions in theorem 1 can be interpreted as the scalar product of the gradient of the respective Hamiltonian with the vector connecting the optimal control with another arbitrary feasible control. The scalar product being positive means, that the two vectors form a sharp angle (thinking about Euclidean vector spaces). This means differing marginally from the optimal control $\hat{w}(a_0)$ (respectively $\hat{v}(t_0)$) in the direction of another feasible control would lead to an increase of the Hamiltonian. This cannot be optimal since we are considering a minimization problem. One should note, that if $\hat{w}(a_0)$ lies in the interior of W the gradient of the initial Hamiltonian has to be equal to zero to fulfil the optimality conditions. (The same holds of course for $\hat{v}(t_0)$ and V.)

The third equation (5.16) just shows that the optimal control $\widehat{u}(t, a)$ has to minimize the distributed Hamiltonian in the set of all feasible controls for every time-point t and every age a (one should not be mislead by the name maximum principle, because we still defined the standard problem as a minimization problem). It should also be mentioned that the two different types of optimality conditions imply, that an optimal control $(\hat{u}, \hat{w}, \hat{v})$ consists of

- a globally (on U) minimizing distributed control $\hat{u}$ for each time and age,
- but only locally optimal initial and boundary controls $\hat{w}$ and $\hat{v}$.

6 | Conclusions

We will now summarise the main findings of each model and the overall results, before making some suggestions about possible extensions or adaptations for further research. In chapter 1 we motivated the introduction of age and skill heterogeneity in an economic model and emphasised the importance and the benefits of treating age as a continuous variable. In chapter 2 we reviewed the results of the model in Prskawetz et al. (2012) with two skill levels. We proofed that in case of perfectly substitutable labour the population development has no effect on the optimal education policies and were able to derive a result, which is non standard in the literature about education and human capital accumulation. Namely it is possible that the optimal education rates are increasing in the earlier ages and we were able to establish sufficient conditions under which this phenomenon occurs. We also managed to replicate the results of the numerical experiments of the authors for the most part and thereby showed that the reactions of the social planer to changes in the population dynamics strongly depend on the elasticities of substitution with respect to age of the two skill levels. As an additional interesting outcome we showed that the social planer anticipates demographic changes and reacts to them before they even set in.

In chapter 3 we introduced a new skill level, which also implied the existence of two different educational investment possibilities. We based this extension on the observation, that three skill levels are a more natural classification and better resemble the available empirical data. Within the framework we were able to transfer the independence on the population size for the case of perfect substitutability analytically and also obtained optimal education rates, which are increasing in age, in numerical simulations. To analyse the importance of the different education types and how they are used, we conducted a sensitivity analysis with varying the costs arising from the education types and the productivities of the different skill levels. For this purpose we conducted several numerical simulations for different parameter sets and deduced a number of mechanisms, which influence the age-specific intensities and the type of education being in focus. The qualitative structure of the solutions remained when the elasticity of substitution with respect to skill was reduced.

The main innovation in chapter 4 was the implementation of a new educational investment strategy for the social planer, which enabled medium skilled worker to become high skilled through further training. We thereby first examined in which parameter constellations the application of this second-level education was profitable and we obtained that this is not the case for the more unrealistic investigated combinations. On the other hand for a large range of realistic scenarios the new control is not only used substantially in the optimal education policy, but also leads to significant gains in the long-run social benefits. We also found some very interesting connections between the aggregated values of the different education types and the mean and standard deviation of the age, at which the education is applied. Regarding the effects of population growth (or decrease) in our extended model, we again deduced a pronounced influence of the elasticities of substitution with respect to age, but the specific changes in the social optimal education rates in each scenario are more manifold than in the basic model and the reasoning behind the differences is more complex.

Overall the analysis of the models has shown, that there are various crucial factors, which influence the social optimal allocation of educational efforts. While we have derived, that the elasticities of substitution have a significant impact on the age structure of the optimal education rates, the cost and productivity structures affect the allocation of educational efforts between the different types of education. Chapter 5 finally summarised the necessary optimality conditions for age-structured optimal control models, which Feichtinger et al. (2003) were able to derive. We should again point out the high degree of generality of their problem formulation, which opens the possibility of various further adaptations and extension. To conclude I will present some options, which might be of interest for further research.

To account for the flaw of the production function, that the elasticity of substitution with respect to age does not depend on the age difference of two workers one could introduce a fuzzy CES production function. This adaptation is more explicitly explained in appendix A.2 and should not lead to major technical difficulties, while possibly providing intriguing new results. On the other hand the production function of our models only depends on labour as an input factor. Adding capital as an input should therefore add realism to the model and lead to interesting results. Nevertheless I expect this step to be technical more challenging as a whole new sector has to be modelled and new investment decisions have to be included. Another important factor for production, especially in long-run economic growth models, is technological progress. Introducing this would also make it possible to model the skill depreciation rate[1] as an endogenous variable. Technological progress could then have a two-sided effect, leading to higher production on the one hand, but increasing the skill depreciation rates at the same time, so more and continuous training has to be applied to keep up with the technology.

As far as population development is concerned the number of people entering the labour market at each time point could be endogenized by introducing fertility rates of the population. We can also suspect intriguing results, as the fertility of women at a certain time point only affects the labour market approximately 15-20 years later, when the children are old enough to enter the labour force. Another demographic force, which would be interesting to adjust for, is migration. One might argue whether migration, if included, should be controllable within the model or not, but in both cases the introduction leads to people entering the labour market with “ages” unequal to zero, which might have fascinating effects on population dynamics and economic development. Simon (2013) and Skritek (2015) already analysed optimal age-specific migration policies in their works, but both of them do not contain optimal education strategies from a social planer point of view.

[1] Which we assumed to be equal to 0 in all our simulations.

Last but not least it might be convenient to more precisely model the educational process. Explicitly defining the total time a person can spend in either production or education, would let us not only more precisely cover the costs of education, but also let us analyse how the optimal time allocation of an individual looks like. Then we would be able to answer if it is better to enjoy a full-time education over a shorter time period before starting to work full-time, or to stay for a longer time in a situation with a part-time job and part-time education simultaneously.

Bibliography

Wittgenstein centre data explorer, 2015. URL http://witt.null2.net/shiny/wic/.

Christian Almeder, Jonathan P. Caulkins, Gustav Feichtinger, and Gernot Tragler. An age-structured single-state drug initiation model–cycles of drug epidemics and optimal prevention programs. *Socio-Economic Planning Sciences*, 38(1):91–109, 2004.

Gustav Feichtinger and Richard F. Hartl. *Optimale Kontrolle ökonomischer Prozesse: Anwendungen des Maximumprinzips in den Wirtschaftswissenschaften.* Walter de Gruyter, 1986.

Gustav Feichtinger, Gernot Tragler, and Vladimir M. Veliov. Optimality conditions for age-structured control systems. *Journal of Mathematical Analysis and Applications*, 288(1): 47–68, 2003.

UNESCO Institute for Statistics. *International standard classification of education: ISCED 2011.* UNESCO Institute for Statistics, 2011. URL http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf.

UNESCO Institute for Statistics. UIS.STAT, February 2016. URL http://data.uis.unesco.org/.

Ivan Frankovic, Michael Kuhn, and Stefan Wrzaczek. Medical care within an olg economy with realistic demography. Technical report, ECON WPS-Vienna University of Technology Working Papers in Economic Theory and Policy, 2016.

Giora Hanoch. Cresh production functions. *Econometrica: Journal of the Econometric Society*, pages 695–712, 1971.

Matthias Kredler. Experience vs. obsolescence: A vintage-human-capital model. *Journal of Economic Theory*, 150:709–739, 2014.

Elke Loichinger and Daniela Weber. Trends in working life expectancy in europe. *Journal of aging and health*, 28(7):1194–1213, 2016.

Wolfgang Lutz and KC Samir. Global human capital: Integrating education and population. *Science*, 333(6042):587–592, 2011.

Dimiter Philipov, Anne Goujon, and Paola Di Giulio. Ageing dynamics of human-capital-specific population: A demographic perspective. *Demographic Research*, 31(44):1311–1336, 2014.

Alexia Prskawetz, Thomas Fent, and Ross Guest. Workforce aging and labor productivity: The role of supply and demand for labor in the G7 countries. *Population and Development Review*, 34:298–323, 2008.

Alexia Prskawetz, Tsvetomir Tsachev, and Vladimir M. Veliov. Optimal education in an age-structured model under changing labor demand and supply. *Macroeconomic Dynamics*, 16 (02):159–183, 2012.

Christa Simon. Effects of age-specific immigration; an optimal control approach, 2013. Wien, Techn. Univ., Diss., 2013.

Bernhard Skritek. On the optimal control of heterogeneous systems, 2015. Wien, Techn. Univ., Diss., 2015.

Vladimir M. Veliov. Newton's method for problems of optimal control of heterogeneous systems. *Optimization Methods and Software*, 18(6):689–703, 2003.

List of Figures

Appendices

A | CES production function

A.1 Elasticity of substitution

The elasticity of substitution is an often appearing and frequently used characteristic of production or utility functions. Despite being so present in the literature, it deserve a short reconsideration of its derivation, due to its remarkable impact on the solutions and general importance in our models. For illustration purposes I am going to use a production function using two inputs: capital and labour.

$$Y = F(K, L)$$

Such a production function has isoquants describing which different input combinations can be used to produce the same amount of output. The marginal rate of substitution

$$MRS_{K,L} = -\frac{dL}{dK}(\overline{K}, \overline{L}) = \frac{\frac{\partial Y(\overline{K},\overline{L})}{\partial K}}{\frac{\partial Y(\overline{K},\overline{L})}{\partial L}} = \frac{Y_K(\overline{K}, \overline{L})}{Y_L(\overline{K}, \overline{L})}$$

defines in which proportion the inputs can be exchanged at a given input combination $(\overline{K}, \overline{L})$ while staying at the same output level, i.e. staying on the same isoquant. The elasticity of substitution now describes in simple words, how much percent the ratio of the input changes, if the marginal rate of substitution changes by 1%.

$$\sigma_{L,K} = \frac{\frac{\Delta(L/K)}{(L/K)}}{\frac{\Delta MRS_{K,L}}{MRS_{K,L}}} = \frac{d\left(\frac{L}{K}\right)}{d\left(\frac{Y_K}{Y_L}\right)} \cdot \frac{\left(\frac{Y_K}{Y_L}\right)}{\left(\frac{L}{K}\right)} = \frac{d\ln(L/K)}{d\ln(\frac{Y_K}{Y_L})} = -\frac{d\ln(K/L)}{d\ln\left(\frac{Y_K}{Y_L}\right)} =$$

$$= -\frac{d\ln(K/L)}{d\ln\left(\frac{Y_K}{Y_L}\right)} = -\left(\frac{d\ln\left(\frac{Y_K}{Y_L}\right)}{d\ln(K/L)}\right)^{-1}$$

In case of the production function of the basic model $Y(t) = \left(\theta_L(t)\tilde{L}(t)^\rho + \theta_H(t)\tilde{H}(t)^\rho\right)^{\frac{1}{\rho}}$, we obtain

$$\frac{\partial Y(t)}{\partial \tilde{L}(t)} = Y_{\tilde{L}}(t) = \frac{1}{\rho}Y(t)^{\frac{1-\rho}{\rho}}\rho\theta_L(t)\tilde{L}(t)^{\rho-1}$$

and therefore for the marginal rate of substitution

$$MRS_{\tilde{L},\tilde{H}} = -\frac{d\tilde{H}}{d\tilde{L}} = \frac{Y_{\tilde{L}}(t)}{Y_{\tilde{H}}(t)} = \frac{\theta_L(t)}{\theta_H(t)} \left(\frac{\tilde{L}(t)}{\tilde{H}(t)} \right)^{\rho-1}$$

Finally the elasticity of substitution can be calculated by

$$\sigma_{\tilde{H},\tilde{L}} = -\left(\frac{\partial \ln(Y_{\tilde{L}}(t)/Y_{\tilde{H}}(t))}{\partial \ln(\tilde{L}(t)/\tilde{H}(t))} \right)^{-1} = -\left[\left(\frac{\partial}{\partial \ln(\tilde{L}(t)/\tilde{H}(t))} \right) \left(\ln\left(\frac{\theta_L(t)}{\theta_H(t)} \right) + (\rho-1)\ln\left(\frac{\tilde{L}(t)}{\tilde{H}(t)} \right) \right) \right]^{-1} =$$

$$= -(\rho-1)^{-1} = \frac{1}{1-\rho}$$

For the sub-aggregates $\tilde{L}(t)$, $\tilde{M}(t)$ and $\tilde{H}(t)$ the constant elasticity of substitution is slightly more difficult to show, since we have an integral structure in this case. In economic papers the derivative of an integral function, e.g. $\tilde{L}(t) = \left(\int_0^\omega \pi_L(s) L(t,s)^{\lambda_L} ds \right)^{1/\lambda_L}$ with respect to its inputs is often mathematically not correctly stated. For notational purposes we will consider a fixed time point t and omit the time argument in the following. Kredler (2014) describes very good that one has to derive the Fréchet derivative of the function $\tilde{L} = f(L(\cdot))^{1/\lambda_L}$. f represents here the integral operator

$$f: \quad L_\infty \to \mathbb{R} \qquad \text{with} \qquad f(L(\cdot)) = \int_0^\omega \pi(s) L(s)^{\lambda_L} ds$$

The Frechet derivative $v(L(\cdot))$ of the operator f is itself an operator, i.e. $v: \; L_\infty \times L_\infty \to \mathbb{R}$ and since f is an integral operator it holds for the derivative

$$v(L(\cdot))(h) = \int_0^\omega \pi(s) \lambda_L L(s)^{\lambda_L - 1} h(s) ds$$

Setting $h(s) = h \cdot \mathbb{1}_{[s_0, s_0+\varepsilon]}$ we can see why using the mathematically incorrect approach of just differentiating the integrand leads to correct results. So for further calculations we define (using additionally the chain rule for differentiation)

$$\tilde{L}_{L(t,s)}(t) = \frac{\partial \tilde{L}(t)}{\partial L(t,s)} := \frac{1}{\lambda_L} \tilde{L}(t)^{\frac{1-\lambda_L}{\lambda_L}} \pi(s) \lambda_L L(t,s)^{\lambda_L - 1} = \tilde{L}(t)^{\frac{1-\lambda_L}{\lambda_L}} \pi(s) L(t,s)^{\lambda_L - 1}$$

With this definition we can again calculate the elasticity of substitution between two age-groups

$$\sigma_{s_1,s_2} = -\left(\frac{\partial \ln(\tilde{L}_{L(t,s_2)}(t) / \ln(\tilde{L}_{L(t,s_1)}(t)}{\partial \ln(L(t,s_2)/L(t,s_1))} \right)^{-1} = \frac{1}{1-\lambda_L} \tag{A.1}$$

following the analogue calculation as for $\sigma_{\tilde{H},\tilde{L}}$.

A.2 CRESH and fuzzy CES production functions

As mentioned in chapter 2 the production function, or to be precise the age-aggregates have the flaw, that the elasticity of substitution is independent of the age difference between two workers (see equation A.1). This makes the production function as a whole in fact relatively simple and easier to trace analytically, but on the other hand also unrealistic. As a solution one could use one of the adaptations for production functions in Prskawetz et al. (2008). I will present the CRESH and the fuzzy CES production function as extension of the classic CES here. Both functions are discussed by the authors with age as a discrete variable, but I will directly illustrate their continuous counterparts. We will use the following notations for the calculations.

- $L : [0, \omega] \to \mathbb{R}$ is the input-function. $L(i)$ is the number of workers of age i.
- $p : [0, \omega] \to \mathbb{R}$ is the price function with $p(i)$ describing the costs for a worker of age i.
- $Y = F(L)$ is the final output in functional dependence of the inputs.

CRESH production function

CRESH production functions were first presented in Hanoch (1971). Their name origins from the fact, that they are function with **C**onstant **R**atios of **E**lasticity of **S**ubstitution, which are also **H**omothetic or **H**omogenous. These types of production functions are implicitly defined by

$$\int_0^\omega \alpha(i) \left(\frac{L(i)}{Y} \right)^{\rho(i)} = 1$$

with $\alpha(i)$ the productivity of a worker of age i and $\rho(i)$ indicating the degree of flexibility of workers with age i. If we additionally define $s(i)$, the factor share of workers of age i, and $a(i)$ through

$$a(i) = \frac{1}{1 - \rho(i)} \qquad \text{and} \qquad s(i) = \frac{p(i)L(i)}{\int_0^\omega p(j)L(j)dj}$$

it can be shown[1], that the elasticity of substitution between workers of two different ages i and j can be calculated by

$$\sigma(i, j) = \frac{a(i)a(j)}{\int_0^\omega s(k)a(k)dk} \tag{A.2}$$

Hence a CRESH production function can be used to model different substitutabilities between different ages. As the parameter $a(i)$ increases for a fixed i, this age-group becomes better substitutable with any other age-group. For example in the low-skill sector age-groups in the middle of the working life should be in general relatively good substitutable with both younger and older workers, since they still have a part of the physical abilities of the younger workers, but also already collected years of working experience and come close to the knowledge of older workers. As a result $a(i)$ should be chosen relatively high for the middle ages. Substituting young workers with older workers and vice-versa will not work out well, since their assets are completely different (physical abilities vs. working experience).

[1] The proof for the discrete function can be found in Hanoch (1971, p. 697 ff.)

However due to its implicit definition the CRESH production function is difficult to implement in our optimal control framework and might by more useful in static age-structured optimization problems.

Fuzzy CES production function

The fuzzy CES production function takes a different approach. Starting with the classic CES production function, the inputs $L(i)$ are replaced by their "fuzzy" representative $\hat{L}(i)$ resulting in

$$Y = \left(\int_0^\omega \alpha(i) \hat{L}(i)^\rho di \right)^{1/\rho}$$

The fuzzy $\hat{L}(i)$ now try to capture the fact, that workers of similar age-groups are far better substitutable than workers with a big age difference. Therefore $\hat{L}(i)$ is assumed to be

$$\hat{L}(i) = \int_0^\omega k(i,j) L(j) dj$$

This means, that within $\hat{L}(i)$ all $L(k)$ are perfectly substitutable. We also see that each $L(i)$ now has not only influence on his corresponding $\hat{L}(i)$, but also on the other $\hat{L}(j)$ in its relatively close environment. $k(i,j)$ is a kernel function, which contains the information about the range of age-groups, which are better substitutable with an age-group i and affect $\hat{L}(i)$. We will require some properties of this kernel function:

- The bigger the age difference to i, the smaller should be the influence of this age-group on $\hat{L}(i)$. This translates into

$$k(i,n) \leq k(i,m) \leq k(i,i) \quad \forall\, n \leq m \leq i$$

$$k(i,n) \leq k(i,m) \leq k(i,i) \quad \forall\, i \leq m \leq n$$

- Consistency with the classic CES-function. If we require

$$\int_0^\omega k(i,j) dj = 1$$

  and reduce the support of $j \mapsto k(i,j)$ to smaller intervals around i, the kernel converges to the Dirac-distribution and it holds that

$$\hat{L}(i) = \int_0^\omega k(i,j) L(j) dj \;\longrightarrow\; \delta_i(L) = L(i)$$

  So the classic CES production function can be interpreted as a limit of fuzzy CES functions.

Simple examples for feasible kernel functions are

- The indicator function $k(i,j) = \frac{1}{2a}\chi_{[i-a,i+a]}$ with a being the age range for better substitutability.

- ”Hat”-functions with

$$k(i,j) = \begin{cases} \frac{1}{2a} \cdot \frac{j-(i-a)}{a} & \text{for } j \in [i-a, i] \\ \frac{1}{2a} \cdot \frac{(i+a)-j}{a} & \text{for } j \in (i, i+a] \\ 0 & \text{otherwise} \end{cases}$$

After all fuzzy CES production functions are more suitable to be included in our framework than the CRESH functions. I will give a quick idea how the production function of the basic model of chapter 2 could be adapted. First we introduce

$$\hat{L}(t,s) = \int_0^\omega k(s,s')L(t,s')ds'$$

with $k(s,s')$ being one of the kernel functions shown above. Then we replace $L(t,s)$ by $\hat{L}(t,s)$ in the equation for $\tilde{L}(t)$. Overall the production sector can be summarised as follow:

$$Y(t) = \Big(\theta_L(t)\tilde{L}(t)^\rho + \theta_H(t)\tilde{H}(t)^\rho\Big)^{1/\rho}$$

$$\tilde{L}(t) = \left(\int_0^\omega \pi_L(s)\hat{L}(t,s)^{\lambda_L} ds\right)^{1/\lambda_L} \qquad \hat{L}(t,s) = \int_0^\omega k_L(s,s')L(t,s')ds'$$

$$\tilde{H}(t) = \left(\int_0^\omega \pi_H(s)\hat{H}(t,s)^{\lambda_H} ds\right)^{1/\lambda_H} \qquad \hat{H}(t,s) = \int_0^\omega k_H(s,s')H(t,s')ds'$$

As we can see this adaptation results in two new distributed state variables in the optimisation problem. Aside from the increasing dimension of the problem, this should not lead to any major technical difficulties. Also the fuzzy CES adaptation does not have to be solely used for age aggregates, but could also be applied to model different substitutabilities between different skill levels. Summarising I did not include these adaptations, since it would have gone beyond the scope of this master thesis, but nevertheless this generalisation should be of major interest for further research.

B | Derivation of the optimality conditions for the basic and the extended model

B.1 Proof for proposition 1 in the basic model

Model transformation

To obtain the optimality conditions, we have to transform the model into the form as given in chapter 5. Therefore we have to introduce new state-variables $\overline{L}(t)$ and $\overline{H}(t)$

$$\overline{L}(t) = \widetilde{L}^{\lambda_L} \quad , \quad \overline{H}(t) = \widetilde{H}^{\lambda_H}$$

This is necessary since the state variable has to be represented by an integral with respect to s. Now we can collect all control variables in a single control $u(t,s)$, the distributed state variables characterised by PDEs in $y(t,s)$ and the aggregated state variables in $q(t)$.[1] We also present the functional form of $h(t,s,y(t,s),u(t,s))$, which will also appear later on.

$$y(t,s) = \begin{pmatrix} L(t,s) \\ H(t,s) \end{pmatrix}$$

$$q(t) = \begin{pmatrix} \overline{L}(t) \\ \overline{H}(t) \\ P(t) \end{pmatrix} = \int_0^\omega \begin{pmatrix} \pi_L(s)L(t,s)^{\lambda_L} \\ \pi_H(s)H(t,s)^{\lambda_H} \\ p(s,u_M(t,s))L(t,s) \end{pmatrix} ds = \int_0^\omega h(t,s,y(t,s),u(t,s))ds$$

In the next step we easily get the function $f(t,s,y(t,s),u(t,s))$, which describes the transition dynamics, and the corresponding initial and boundary conditions.

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) y(t,s) = f(t,s,y(t,s),u(t,s)) = \begin{pmatrix} \delta(t,s)H(t,s) - e(s)L(t,s) - l(t,s)u(t,s)L(t,s) \\ -\delta(t,s)H(t,s) + e(s)L(t,s) + l(t,s)u(t,s)L(t,s) \end{pmatrix}$$

$$y(0,s) = \begin{pmatrix} L_b(s) \\ H_b(s) \end{pmatrix} \quad , \quad y(t,0) = \begin{pmatrix} L_0(t) \\ H_0(t) \end{pmatrix}$$

In the objective function the new state-variables $\overline{L}(t)$ and $\overline{H}(t)$ appear. Note that exponents of the variables understandably had to change. We also integrate over s and divide by ω, what

[1] In our model we have no variables corresponding to $p(t,s)$ in chapter 5. As a result $p(t,s)$, the function $g(t,s,s',y(t,s'),u(t,s'))$ and the co-state variable $\eta(t,s)$ can be omitted and will not appear in the equations.

results in the original objective function, since all terms are independent of s. Consequently we receive the standard form.

$$\int_0^T \int_0^\omega \frac{1}{\omega} \cdot e^{-rt} \left(\left(\theta_L(t)\overline{L}(t)^{\frac{\rho}{\lambda_L}} + \theta_H(t)\overline{H}(t)^{\frac{\rho}{\lambda_H}} \right)^{1/\rho} - P(t) \right) dsdt = \int_0^T \int_0^\omega \mathcal{L}(t,s,q(t))dsdt$$

To avoid confusion I am using $\mathcal{L}(t,s,q(t))$ instead of $L(t,s,q(t))$ for the function under the integral since we already use $L(t,s)$ for the number of low skilled workers.

Adjoint system

Now we have transformed all equation into the needed framework of chapter 5 and can introduce the adjoint variables

$$\xi(t,s) = \Big(\xi_L(t,s), \xi_H(t,s)\Big) \qquad , \qquad \zeta(t) = \Big(\zeta_{\overline{L}}(t), \zeta_{\overline{H}}(t), \zeta_P(t)\Big)$$

and we can explicitly write down the adjoint system

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\xi(t,s) = \underbrace{\nabla_y \mathcal{L}(t,s,q(t))}_{=0} + \xi(t,s)\nabla_y f(t,s,y(t,s),u(t,s)) + \zeta(t)\nabla_y h(t,s,y(t,s),u(t,s)) =$$

$$= \Big(\xi_L(t,s), \xi_H(t,s)\Big) \begin{pmatrix} -e(s) - l(t,s)u(t,s) & \delta(t,s) \\ e(s) + l(t,s)u(t,s) & -\delta(t,s) \end{pmatrix} +$$

$$+ \Big(\zeta_{\overline{L}}(t), \zeta_{\overline{H}}(t), \zeta_P(t)\Big) \begin{pmatrix} \pi_L(s)\lambda_L L(t,s)^{\lambda_L - 1} & 0 \\ 0 & \pi_H(s)\lambda_H H(t,s)^{\lambda_H - 1} \\ p(s,u(t,s)) & 0 \end{pmatrix} =$$

$$= \begin{pmatrix} (-e(s) - l(t,s)u(t,s))\,\xi_L(t,s) + (e(s) + l(t,s)u(t,s))\,\xi_H(t,s) + \pi_L(s)\lambda_L L(t,s)^{\lambda_L - 1}\zeta_{\overline{L}}(t)p(s,u(t,s))\zeta_P(t) \\ \delta(t,s)\,(\xi_L(t,s) - \xi_H(t,s)) + \pi_H(s)\lambda_H H(t,s)^{\lambda_H - 1}\zeta_{\overline{H}}(t) \end{pmatrix}^T$$

Simplifying and treating the two components separately leads to

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\xi_L(t,s) = \Big(-e(s) - l(t,s)u(t,s)\Big) \cdot \Big(\xi_L(t,s) - \xi_H(t,s)\Big) + \pi_L(s)\lambda_L L(t,s)^{\lambda_L - 1}\zeta_{\overline{L}}(t) + p(s,u(t,s))\zeta_P(t) \tag{B.1}$$

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\xi_H(t,s) = \delta(t,s)\Big(\xi_L(t,s) - \xi_H(t,s)\Big) + \pi_H(s)\lambda_H H(t,s)^{\lambda_H - 1}\zeta_{\overline{H}}(t) \tag{B.2}$$

with the boundary conditions

$$\xi_L(T,s) = \xi_H(T,s) = 0 \qquad \forall s \in [0,\omega]$$

$$\xi_L(t,\omega) = \xi_H(t,\omega) = 0 \qquad \forall t \in [0,T]$$

Also the $\zeta(t)$ can be explicitly expressed

$$\zeta(t) = \int_0^\omega \nabla_q \mathcal{L}(t,s,q(t)) + \underbrace{\nabla_q f(t,s,y(t,s),u(t,s))}_{=0} \, ds$$

$$= \int_0^\omega \begin{pmatrix} \frac{1}{\omega} \cdot e^{-rt} \cdot \frac{1}{\rho} Y(t)^{1-\rho} \cdot \frac{\rho}{\lambda_L} \theta_L(t) \overline{L}(t)^{\frac{\rho-\lambda_L}{\lambda_L}} \\ \frac{1}{\omega} \cdot e^{-rt} \cdot \frac{1}{\rho} Y(t)^{1-\rho} \cdot \frac{\rho}{\lambda_H} \theta_H(t) \overline{H}(t)^{\frac{\rho-\lambda_H}{\lambda_H}} \\ \frac{1}{\omega} \cdot e^{-rt} \cdot (-1) \end{pmatrix}^T ds = e^{-rt} \cdot \begin{pmatrix} Y(t)^{1-\rho} \cdot \frac{1}{\lambda_L} \theta_L(t) \overline{L}(t)^{\frac{\rho-\lambda_L}{\lambda_L}} \\ Y(t)^{1-\rho} \cdot \frac{1}{\lambda_H} \theta_H(t) \overline{H}(t)^{\frac{\rho-\lambda_H}{\lambda_H}} \\ (-1) \end{pmatrix}^T$$

We can now also change back from the variables $\overline{L}(t)$ to $\widetilde{L}(t)$:

$$\zeta_{\overline{L}}(t) = e^{-rt} \cdot Y(t)^{1-\rho} \cdot \frac{1}{\lambda_L} \theta_L(t) \overline{L}(t)^{\frac{\rho-\lambda_L}{\lambda_L}} = e^{-rt} \cdot Y(t)^{1-\rho} \cdot \frac{1}{\lambda_L} \theta_L(t) \widetilde{L}(t)^{\rho-\lambda_L}$$

$$\zeta_{\overline{H}}(t) = e^{-rt} \cdot Y(t)^{1-\rho} \cdot \frac{1}{\lambda_H} \theta_H(t) \overline{H}(t)^{\frac{\rho-\lambda_H}{\lambda_H}} = e^{-rt} \cdot Y(t)^{1-\rho} \cdot \frac{1}{\lambda_H} \theta_H(t) \widetilde{H}(t)^{\rho-\lambda_H}$$

$$\zeta_P(t) = \qquad = -\, e^{-rt}$$

We will now insert the expressions for $\left(\zeta_{\overline{L}}(t), \zeta_{\overline{H}}(t), \zeta_P(t)\right)$ into the PDEs (B.1) and (B.2) and focus on the investigation of $\xi(t)$. This step is possible, since $\zeta(t)$ is explicitly given[2]. However we also do not loose any meaningful economic interpretations through the elimination of $\zeta(t)$, as it contains just the shadow-prices of the transformed variables (e.g. $\overline{L}(t)$) (which themselves are hard to interpret), and not the original variables (e.g. $\widetilde{L}(t)$).

Inserting the terms as described results in the equations

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \xi_L(t,s) = \Big(-e(s) - l(t,s)u(t,s)\Big) \cdot \Big(\xi_L(t,s) - \xi_H(t,s)\Big) + \\ + e^{-rt} \cdot Y(t)^{1-\rho} \cdot \theta_L(t)\pi_L(s) \cdot \widetilde{L}(t)^{\rho-\lambda_L} L(t,s)^{\lambda_L - 1} - e^{-rt} p(s, u(t,s)) \tag{B.3}$$

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \xi_H(t,s) = \delta(t,s)\Big(\xi_L(t,s) - \xi_H(t,s)\Big) + \\ + e^{-rt} \cdot Y(t)^{1-\rho} \cdot \theta_H(t)\pi_H(s)\widetilde{H}(t)^{\rho-\lambda_H} H(t,s)^{\lambda_H - 1} \tag{B.4}$$

Switch to current-values

To eliminate the term e^{-rt} and receive economically more interesting results, we switch from the present value to the current value adjoint variables[3].

Definition 3

[2]This not naturally, since $\zeta(t)$ is defined through an integral in general.
[3]Elimination e^{-rt} also removes the explicit dependence of the PDE of t.

For an optimal control model exhibiting a discounting term e^{-rt} in the objective function, we define the current value adjoint variables $\mu(t,s)$ and $\nu(t,s)$ as follows

$$\mu_L(t,s) := e^{rt}\xi_L(t,s) \qquad , \qquad \mu_H(t,s) := e^{rt}\xi_H(t,s)$$

$$\nu_{\overline{L}}(t) := e^{rt}\zeta_{\overline{L}}(t) \qquad , \qquad \nu_{\overline{H}}(t) := e^{rt}\zeta_{\overline{H}}(t) \qquad , \qquad \nu_P := e^{rt}\zeta_P(t)$$

The current value adjoint variables satisfy the following PDEs ($i = L, H$)

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\mu_i(t,s) = -e^{rt}\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\xi_i(t,s) - r\cdot e^{rt}\xi_i(t,s) = -e^{rt}\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\xi_i(t,s) - r\mu_i(t,s)$$

Using the already derived terms for $-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\xi_i(t,s)$ we obtain

$$\begin{aligned} -\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\mu_L(t,s) &= \Big(-e(s) - l(t,s)u(t,s)\Big)\cdot\Big(\mu_L(t,s) - \mu_H(t,s)\Big) + \\ &\quad + Y(t)^{1-\rho}\cdot\theta_L(t)\pi_L(s)\cdot\widetilde{L}(t)^{\rho-\lambda_L}L(t,s)^{\lambda_L-1} - p(s,u(t,s)) - r\mu_L(t,s) \end{aligned} \tag{B.5}$$

$$\begin{aligned} -\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\mu_H(t,s) &= \delta(t,s)\Big(\mu_L(t,s) - \mu_H(t,s)\Big) + \\ &\quad + Y(t)^{1-\rho}\cdot\theta_H(t)\pi_H(s)\widetilde{H}(t)^{\rho-\lambda_H}H(t,s)^{\lambda_H-1} - r\mu_H(t,s) \end{aligned} \tag{B.6}$$

Optimality conditions for control $u(t,s)$

We now can turn to the optimality conditions regarding the control $u(t,s)$. This time we directly change to the current-value terms and define the current-value Hamiltonian $\mathcal{H}(t,s,u)$

$$\mathcal{H}(t,s,u) = e^{rt}\cdot\mathcal{L}(t,s,u) + \mu(t,s)f(t,s,u) + \nu(t)h(t,s,u) =$$

$$= \frac{1}{\omega}\cdot\left(\left(\theta_L(t)\overline{L}(t)^{\frac{\rho}{\lambda_L}} + \theta_H(t)\overline{H}(t)^{\frac{\rho}{\lambda_H}}\right)^{1/\rho} - P(t)\right) + \mu(t,s)f(t,s,u) + \nu(t)\begin{pmatrix}\pi_L(s)L(t,s)^{\lambda_L}\\ \pi_H(s)H(t,s)^{\lambda_H}\\ p(s,u(t,s))L(t,s)\end{pmatrix} \tag{B.7}$$

Using this function and the results of theorem 1 we obtain, that the optimal control has the following properties[4]: $u(t,s)$ maximizes the following function (resulting from eliminating all terms independent of $u(t,s)$)

$$\begin{aligned} F(u(t,s)) &= \underbrace{\nu_P(t)}_{=-1}p(s,u(t,s))L(t,s) + \mu(t,s)\begin{pmatrix}-l(t,s)u(t,s)L(t,s)\\ l(t,s)u(t,s)L(t,s)\end{pmatrix} = \\ &= -p(s,u(t,s))L(t,s) + \Big[\mu_H(t,s) - \mu_L(t,s)\Big]l(t,s)u(t,s)L(t,s) \end{aligned} \tag{B.8}$$

[4]Note that in chapter 5 we are analysing a *minimization* problem, but our model *maximizes* profits minus costs. As a result the optimal controls $u_M(t,s)$ and $u_H(t,s)$ have to maximize the current-value Hamiltonian

If we assume that for all $s \in [0,\omega]$ $p(s,u_M)$ is continuous differentiable with respect to u and additionally take into account the convex structure of the cost function, we can characterise the optimal controls by taking the derivative of equations (B.8) and setting it equal to zero.

$$F'(u(t,s)) = -\frac{\partial p(s,u(t,s))}{\partial u} L(t,s) + \Big[\mu_H(t,s) - \mu_L(t,s)\Big] l(t,s) L(t,s) \overset{!}{=} 0 \tag{B.9}$$

We should stress, that this first-order optimality condition is only necessary if $u(t,s)$ is strictly positive. If e.g. the optimal rate $u(t,s)=0$, it might hold that $F'(u(t,s)) < 0$, but $u(t,s)$ still maximizes the Hamiltonian. After all we should not be surprised by negative gradients in the optimal solution. The mathematical correct formulation, which encounters these facts, can be written as follows

$$F'(u(t,s)) \leq 0, \qquad F'(u(t,s)) \cdot u(t,s) = 0 \tag{B.10}$$

Summarising the optimality conditions

Rearranging and simplifying the equation we obtain the final optimality conditions, which we are going to work with in the following. First the PDEs for the co-state variables can be even further reduced by introducing $\Delta(t,s) = \mu_H(t,s) - \mu_L(t,s)$, the difference between the current-value shadow prices of high- and low-skilled workers

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \Delta(t,s) = \Big(r + e(s) + \delta(t,s) + l(t,s)u(t,s)\Big) \cdot \Delta(t,s) - p(s,u(t,s)) - (f_H(t,s) - f_L(t,s)) \tag{B.11}$$

$$\Delta(T,s) = 0, \quad \Delta(t,\omega) = 0 \qquad \forall t \in [0,T]\ \forall s \in [0,\omega] \tag{B.12}$$

with

$$f_L(t,s) = Y(t)^{1-\rho} \cdot \theta_L(t)\pi_L(s) \cdot \widetilde{L}(t)^{\rho-\lambda_L} L(t,s)^{\lambda_L - 1}$$

and $f_H(t,s)$ analogue. Introducing $\Delta(t,s)$ also reduces the FOC for the control $u(t,s)$ to

$$-\frac{\partial p(s,u(t,s))}{\partial u} + l(t,s)\Delta(t,s) \leq 0$$

$$\left(-\frac{\partial p(s,u(t,s))}{\partial u} + l(t,s)\Delta(t,s)\right) \cdot u(t,s) = 0$$

B.2 Derivation of th optimality conditions for the extended model in chapter 3

In this section we will derive the optimality conditions for the extended model of chapter 3. The derivations are similar to the ones for the basic model, but slightly more difficult, due to the additional state variable and control. The derivations are now only shortly described and for the exact explanations of the procedure I refer to Appendix B.1.

Model transformation

As in the basic model we carry out a model transformation to fit the standard form of chapter 5. We introduce

$$\overline{L}(t) = \widetilde{L}^{\lambda_L} \quad , \quad \overline{M}(t) = \widetilde{M}^{\lambda_M} \quad , \quad \overline{H}(t) = \widetilde{H}^{\lambda_H}$$

Since the control is now two dimensional (for each (t, s)), the distributed states three dimensional and the aggregated states four dimensional, we obtain a different formula for the function $h(t, s, y(t, s), u(t, s))$

$$u(t,s) = \begin{pmatrix} u_M(t,s) \\ u_H(t,s) \end{pmatrix} \qquad y(t,s) = \begin{pmatrix} L(t,s) \\ M(t,s) \\ H(t,s) \end{pmatrix}$$

$$q(t) = \begin{pmatrix} \overline{L}(t) \\ \overline{M}(t) \\ \overline{H}(t) \\ P(t) \end{pmatrix} = \int_0^{\omega} \begin{pmatrix} \pi_L(s)L(t,s)^{\lambda_L} \\ \pi_M(s)M(t,s)^{\lambda_M} \\ \pi_H(s)H(t,s)^{\lambda_H} \\ \Big(p_M(s,u_M(t,s)) + p_H(s,u_H(t,s))\Big)L(t,s) \end{pmatrix} ds = \int_0^{\omega} h(t,s,y(t,s),u(t,s))ds$$

Again we define the function $f(t, s, y(t, s), u(t, s))$, which describes the transition dynamics, and the corresponding initial and boundary conditions. The function f is relatively more complicated in the extended model, since in the basic model the two flows had the same value, but only different sign.

$$\Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\Big)y(t,s) = f(t,s,y(t,s),u(t,s)) =$$

$$= \begin{pmatrix} \delta_M(t,s)M(t,s) + \delta_H(t,s)H(t,s) - e(s)L(t,s) - l_M(t,s)u_M(t,s)L(t,s) - l_H(t,s)u_H(t,s)L(t,s) \\ -\delta_M(t,s)M(t,s) + e(s)L(t,s) + l_M(t,s)u_M(t,s)L(t,s) \\ -\delta_H(t,s)H(t,s) + l_H(t,s)u_H(t,s)L(t,s) \end{pmatrix}$$

$$y(0,s) = \begin{pmatrix} L_0(s) \\ M_0(s) \\ H_0(s) \end{pmatrix} \quad , \quad y(t,0) = \begin{pmatrix} L_b(t) \\ M_b(t) \\ H_b(t) \end{pmatrix}$$

As in the basic model the objective function changes and now contains the new state-variables

$$\int_0^T \int_0^\omega \frac{1}{\omega} \cdot e^{-rt} \left(\left(\theta_L(t)\overline{L}(t)^{\frac{\rho}{\lambda_L}} + \theta_M(t)\overline{M}(t)^{\frac{\rho}{\lambda_M}} + \theta_H(t)\overline{H}(t)^{\frac{\rho}{\lambda_H}} \right)^{1/\rho} - P(t) \right) dsdt = \int_0^T \int_0^\omega \mathcal{L}(t,s,q(t))dsdt$$

Adjoint system

The adjoint variables are now obviously of higher dimension too

$$\xi(t,s) = \Big(\xi_L(t,s), \xi_M(t,s), \xi_H(t,s)\Big) \qquad , \qquad \zeta(t) = \Big(\zeta_{\overline{L}}(t), \zeta_{\overline{M}}(t), \zeta_{\overline{H}}(t), \zeta_P(t)\Big)$$

And the PDEs take the following form

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \xi(t,s) = \underbrace{\nabla_y \mathcal{L}(t,s,q(t))}_{=0} + \xi(t,s)\nabla_y f(t,s,y(t,s),u(t,s)) + \zeta(t)\nabla_y h(t,s,y(t,s),u(t,s)) =$$

$$= \Big(\xi_L(t,s), \xi_M(t,s), \xi_H(t,s)\Big) \begin{pmatrix} -e(s) - l_M(t,s)u_M(t,s) - l_H(t,s)u_H(t,s) & \delta_M(t,s) & \delta_H(t,s) \\ e(s) + l_M(t,s)u_M(t,s) & -\delta_M(t,s) & 0 \\ l_H(t,s)u_H(t,s) & 0 & -\delta_H(t,s) \end{pmatrix} +$$

$$+ \Big(\zeta_{\overline{L}}(t), \zeta_{\overline{M}}(t), \zeta_{\overline{H}}(t), \zeta_P(t)\Big) \begin{pmatrix} \pi_L(s)\lambda_L L(t,s)^{\lambda_L - 1} & 0 & 0 \\ 0 & \pi_M(s)\lambda_M M(t,s)^{\lambda_M - 1} & 0 \\ 0 & 0 & \pi_H(s)\lambda_H H(t,s)^{\lambda_H - 1} \\ p_M(s,u_M(t,s)) + p_H(s,u_H(t,s)) & 0 & 0 \end{pmatrix} =$$

$$= \begin{pmatrix} \left(-e(s) - l_M(t,s)u_M(t,s) - l_H(t,s)u_H(t,s)\right)\xi_L(t,s) + \left(e(s) + l_M(t,s)u_M(t,s)\right)\xi_M(t,s) + l_H(t,s)u_H(t,s)\xi_H(t,s) + \\ \delta_M(t,s)\left(\xi_L(t,s) - \xi_M(t,s)\right) + \pi_M(s)\lambda_M M(t,s)^{\lambda_M-1}\zeta_{\overline{M}}(t) \\ \delta_H(t,s)\left(\xi_L(t,s) - \xi_H(t,s)\right) + \pi_H(s)\lambda_H H(t,s)^{\lambda_H-1}\zeta_{\overline{H}}(t) \end{pmatrix}$$

$$\left. + \pi_L(s)\lambda_L L(t,s)^{\lambda_L - 1}\zeta_{\overline{L}}(t) + \Big(p_M(s,u_M(t,s)) + p_H(s,u_H(t,s))\Big)\zeta_P(t) \right)^T$$

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \xi_L(t,s) = \Big(-e(s) - l_M(t,s)u_M(t,s) - l_H(t,s)u_H(t,s)\Big)\xi_L(t,s) +$$
$$+ \Big(e(s) + l_M(t,s)u_M(t,s)\Big)\xi_M(t,s) + l_H(t,s)u_H(t,s)\xi_H(t,s) + \pi_L(s)\lambda_L L(t,s)^{\lambda_L - 1}\zeta_{\overline{L}}(t) +$$
$$+ \Big(p_M(s,u_M(t,s)) + p_H(s,u_H(t,s))\Big)\zeta_P(t) \tag{B.13}$$

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \xi_M(t,s) = \delta_M(t,s)\Big(\xi_L(t,s) - \xi_M(t,s)\Big) + \pi_M(s)\lambda_M M(t,s)^{\lambda_M - 1}\zeta_{\overline{M}}(t) \tag{B.14}$$

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \xi_H(t,s) = \delta_H(t,s)\Big(\xi_L(t,s) - \xi_H(t,s)\Big) + \pi_H(s)\lambda_H H(t,s)^{\lambda_H - 1}\zeta_{\overline{H}}(t) \tag{B.15}$$

with the boundary conditions

$$\xi(T,s) = (0,0,0) \quad \forall s \in [0,\omega] \qquad \text{and} \qquad \xi(t,\omega) = (0,0,0) \quad \forall t \in [0,T]$$

$\zeta(t)$ still can be explicitly expressed. The aggregated co-states are thereby not affected by the more complicated transition dynamic and we obtain the same terms for $\zeta_{\overline{L}}(t)$, $\zeta_{\overline{H}}(t)$ and $\zeta_P(t)$ as in the basic model, but additionally have $\zeta_{\overline{M}}(t)$ with

$$\zeta_{\overline{M}}(t) = e^{-rt} \cdot Y(t)^{1-\rho} \cdot \frac{1}{\lambda_M} \theta_M(t) \overline{M}(t)^{\frac{\rho-\lambda_M}{\lambda_M}} \quad = \quad e^{-rt} \cdot Y(t)^{1-\rho} \cdot \frac{1}{\lambda_M} \theta_M(t) \widetilde{M}(t)^{\rho-\lambda_M}$$

As in the basic model we insert the expressions for $\Big(\zeta_{\overline{L}}(t), \zeta_{\overline{M}}(t), \zeta_{\overline{H}}(t), \zeta_P(t)\Big)$ into the PDEs (B.13) - (B.15) and focus on the investigation of $\xi(t)$. At the same time we directly make another step and change to the current-value variables (see definition 3)

$$\begin{aligned} -\Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\Big)\mu_L(t,s) = &\Big(- e(s) - l_M(t,s)u_M(t,s) - l_H(t,s)u_H(t,s)\Big)\mu_L(t,s) + \\ &+ \Big(e(s) + l_M(t,s)u_M(t,s)\Big)\mu_M(t,s) + l_H(t,s)u_H(t,s)\mu_H(t,s) + \\ &+ Y(t)^{1-\rho} \cdot \theta_L(t)\pi_L(s) \cdot \widetilde{L}(t)^{\rho-\lambda_L} L(t,s)^{\lambda_L - 1} - \Big(p_M(s,u_M(t,s)) + p_H(s,u_H(t,s))\Big) - r\mu_L(t,s) \end{aligned} \tag{B.16}$$

$$-\Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\Big)\mu_M(t,s) = \delta_M(t,s)\Big(\mu_L(t,s) - \mu_M(t,s)\Big) + Y(t)^{1-\rho} \cdot \theta_M(t)\pi_M(s)\widetilde{M}(t)^{\rho-\lambda_M} M(t,s)^{\lambda_M-1} - r\mu_M(t,s) \tag{B.17}$$

$$-\Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\Big)\mu_H(t,s) = \delta_H(t,s)\Big(\mu_L(t,s) - \mu_H(t,s)\Big) + Y(t)^{1-\rho} \cdot \theta_H(t)\pi_H(s)\widetilde{H}(t)^{\rho-\lambda_H} H(t,s)^{\lambda_H-1} - r\mu_H(t,s) \tag{B.18}$$

Optimality conditions for control $u(t,s)$

For the optimality conditions for $u(t,s)$ we need the current-value Hamiltonian

$$\begin{aligned} \mathcal{H}(t,s,u) &= e^{rt} \cdot \mathcal{L}(t,s,u) + \mu(t,s)f(t,s,u) + \nu(t)h(t,s,u) = \\ &= \frac{1}{\omega} \cdot \left(\Big(\theta_L(t)\overline{L}(t)^{\frac{\rho}{\lambda_L}} + \theta_M(t)\overline{M}(t)^{\frac{\rho}{\lambda_M}} + \theta_H(t)\overline{H}(t)^{\frac{\rho}{\lambda_H}}\Big)^{1/\rho} - P(t) \right) + \\ &+ \mu(t,s)f(t,s,u) + \nu(t) \begin{pmatrix} \pi_L(s)L(t,s)^{\lambda_L} \\ \pi_M(s)M(t,s)^{\lambda_M} \\ \pi_H(s)H(t,s)^{\lambda_H} \\ \Big(p_M(s,u_M(t,s)) + p_H(s,u_H(t,s))\Big)L(t,s) \end{pmatrix} \end{aligned} \tag{B.19}$$

Theorem 1 shows us that the optimal $u_M(t,s)$ and $u_H(t,s)$ respectively maximize the following functions (resulting from eliminating all terms of the Hamiltonian independent of $u_M(t,s)$ and $u_H(t,s)$ respectively)

$$F_1(u_M(t,s)) = \underbrace{\nu_P(t)}_{=-1} p_M(s,u_M(t,s))L(t,s) + \mu(t,s) \begin{pmatrix} -l_M(t,s)u_M(t,s)L(t,s) \\ l_M(t,s)u_M(t,s)L(t,s) \\ 0 \end{pmatrix} =$$

$$= -p_M(s, u_M(t,s))L(t,s) + \Big[\mu_M(t,s) - \mu_L(t,s)\Big] l_M(t,s)u_M(t,s)L(t,s) \quad \text{(B.20)}$$

$$F_2(u_H(t,s)) = \underbrace{\nu_P(t)}_{=-1} p_H(s, u_H(t,s))L(t,s) + \mu(t,s) \begin{pmatrix} -l_H(t,s)u_H(t,s)L(t,s) \\ 0 \\ l_H(t,s)u_H(t,s)L(t,s) \end{pmatrix} =$$

$$= -p_H(s, u_H(t,s))L(t,s) + \Big[\mu_H(t,s) - \mu_L(t,s)\Big] l_H(t,s)u_H(t,s)L(t,s) \quad \text{(B.21)}$$

If we assume that for all $s \in [0, \omega] : p_M(s, u_M)$ and $p_H(s, u_H)$ are continuous differentiable with respect to u_M and u_H respectively and additionally take into account the convex structure of these costs function, we can characterise the optimal controls similar to the basic model

$$F_1'(u_M(t,s)) \leq 0, \qquad F_1'(u_M(t,s)) \cdot u_M(t,s) = 0 \quad \text{(B.22)}$$

$$F_2'(u_H(t,s)) \leq 0, \qquad F_2'(u_H(t,s)) \cdot u_H(t,s) = 0 \quad \text{(B.23)}$$

Summarising the optimality conditions

Now we can summarize the optimality conditions for the extended model. Below the equations are a little bit simplified, but we cannot reduce them further to equations for the differences in the shadow prices as in the basic model.The PDEs for the co-state variables can be written as

$$\begin{aligned} \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \mu_L(t,s) &= \Big(e(s) + l_M(t,s)u_M(t,s) + l_H(t,s)u_H(t,s) + r\Big)\mu_L(t,s) - \\ &- \Big(e(s) + l_M(t,s)u_M(t,s)\Big)\mu_M(t,s) - l_H(t,s)u_H(t,s)\mu_H(t,s) + \\ &+ \Big(p_M(s, u_M(t,s)) + p_H(s, u_H(t,s))\Big) - f_L(t,s) \quad \text{(B.24)} \end{aligned}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \mu_M(t,s) = \delta_M(t,s)\Big(\mu_M(t,s) - \mu_L(t,s)\Big) + r\mu_M(t,s) - f_M(t,s) \quad \text{(B.25)}$$

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right) \mu_H(t,s) = \delta_H(t,s)\Big(\mu_H(t,s) - \mu_L(t,s)\Big) + r\mu_H(t,s) - f_H(t,s) \quad \text{(B.26)}$$

$$\mu_i(T,s) = 0 \quad \mu_i(t,\omega) = 0 \qquad \forall i \in \{L, M, H\} \; \forall t \in [0,T] \; \forall s \in [0,\omega]$$

with

$$f_L(t,s) = Y(t)^{1-\rho} \cdot \theta_L(t)\pi_L(s) \cdot \widetilde{L}(t)^{\rho-\lambda_L} L(t,s)^{\lambda_L - 1}$$

and $f_M(t,s))$ and $f_H(t,s)$ analogue and the FOC for the controls are

$$-\frac{\partial p_M(s, u_M(t,s))}{\partial u_M} + \Big(\mu_M(t,s) - \mu_L(t,s)\Big) l_M(t,s) \leq 0 \quad \text{(B.27)}$$

$$\Big(-\frac{\partial p_M(s,u_M(t,s))}{\partial u_M} + \Big(\mu_M(t,s) - \mu_L(t,s)\Big) l_M(t,s)\Big) \cdot u_M(t,s) = 0 \tag{B.28}$$

$$-\frac{\partial p_H(s,u_H(t,s))}{\partial u_H} + \Big(\mu_H(t,s) - \mu_L(t,s)\Big) l_H(t,s) \leq 0 \tag{B.29}$$

$$\Big(-\frac{\partial p_H(s,u_H(t,s))}{\partial u_H} + \Big(\mu_H(t,s) - \mu_L(t,s)\Big) l_H(t,s)\Big) \cdot u_H(t,s) = 0 \tag{B.30}$$

B.3 Derivation of the optimality conditions for the "Lehre mit Matura" model

Introducing the flow between medium and high-skilled workers clearly also affects the optimality conditions. The control obtains another dimension, while the dimensions of the other variables stay the same compared to the extended model in chapter 3. To derive the optimality conditions we also perform the same steps as before beginning with the model transformation

$$\overline{L}(t) = \widetilde{L}^{\lambda_L} \quad , \quad \overline{M}(t) = \widetilde{M}^{\lambda_M} \quad , \quad \overline{H}(t) = \widetilde{H}^{\lambda_H}$$

The control is now three dimensional (for each (t,s)) and also the formula for the education expenditures has changed.

$$u(t,s) = \begin{pmatrix} u_M(t,s) \\ u_H(t,s) \\ u_Z(t,s) \end{pmatrix} \qquad y(t,s) = \begin{pmatrix} L(t,s) \\ M(t,s) \\ H(t,s) \end{pmatrix}$$

$$q(t) = \begin{pmatrix} \overline{L}(t) \\ \overline{M}(t) \\ \overline{H}(t) \\ P(t) \end{pmatrix} = \int_0^\omega \begin{pmatrix} \pi_L(s) L(t,s)^{\lambda_L} \\ \pi_M(s) M(t,s)^{\lambda_M} \\ \pi_H(s) H(t,s)^{\lambda_H} \\ \Big(p_M(s,u_M(t,s)) + p_H(s,u_H(t,s))\Big) L(t,s) + p_Z(s,u_Z(t,s)) M(t,s) \end{pmatrix} ds =$$

$$= \int_0^\omega h(t,s,y(t,s),u(t,s)) ds$$

The PDE for the low-skilled workers stays the same, while we have to include the new flow between the medium and high skilled workers.

$$\Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\Big) y(t,s) = f(t,s,y(t,s),u(t,s)) =$$

$$= \begin{pmatrix} \delta_M(t,s)M(t,s) + \delta_H(t,s)H(t,s) - e(s)L(t,s) - l_M(t,s)u_M(t,s)L(t,s) - l_H(t,s)u_H(t,s)L(t,s) \\ -\delta_M(t,s)M(t,s) + e(s)L(t,s) + l_M(t,s)u_M(t,s)L(t,s) - l_Z(t,s)u_Z(t,s)M(t,s) \\ -\delta_H(t,s)H(t,s) + l_H(t,s)u_H(t,s)L(t,s) + l_Z(t,s)u_Z(t,s)M(t,s) \end{pmatrix}$$

$$y(0,s) = \begin{pmatrix} L_0(s) \\ M_0(s) \\ H_0(s) \end{pmatrix} \quad , \quad y(t,0) = \begin{pmatrix} L_b(t) \\ M_b(t) \\ H_b(t) \end{pmatrix}$$

Since nothing has changed about the aggregated state variables the objective function stays the same:

$$\int_0^T \int_0^\omega \frac{1}{\omega} \cdot e^{-rt} \left(\left(\theta_L(t)\overline{L}(t)^{\frac{\rho}{\lambda_L}} + \theta_M(t)\overline{M}(t)^{\frac{\rho}{\lambda_M}} + \theta_H(t)\overline{H}(t)^{\frac{\rho}{\lambda_H}} \right)^{1/\rho} - P(t) \right) dsdt = \int_0^T \int_0^\omega \mathcal{L}(t,s,q(t))dsdt$$

Adjoint system

Deriving the adjoint system for

$$\xi(t,s) = \Big(\xi_L(t,s), \xi_M(t,s), \xi_H(t,s)\Big) \qquad , \qquad \zeta(t) = \Big(\zeta_{\overline{L}}(t), \zeta_{\overline{M}}(t), \zeta_{\overline{H}}(t), \zeta_P(t)\Big)$$

we see that essentially only the PDE for $\xi_M(t,s)$, the adjoint variable of the medium skilled workers, is affected by the newly introduced control.

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\xi(t,s) = \underbrace{\nabla_y \mathcal{L}(t,s,q(t))}_{=0} + \xi(t,s)\nabla_y f(t,s,y(t,s),u(t,s)) + \zeta(t)\nabla_y h(t,s,y(t,s),u(t,s)) =$$

$$= \Big(\xi_L(t,s), \xi_M(t,s), \xi_H(t,s)\Big) \begin{pmatrix} -e(s) - l_M(t,s)u_M(t,s) - l_H(t,s)u_H(t,s) & \delta_M(t,s) & \delta_H(t,s) \\ e(s) + l_M(t,s)u_M(t,s) & -\delta_M(t,s) - l_Z(t,s)u_Z(t,s) & 0 \\ l_H(t,s)u_H(t,s) & l_Z(t,s)u_Z(t,s) & -\delta_H(t,s) \end{pmatrix} +$$

$$+ \Big(\zeta_{\overline{L}}(t), \zeta_{\overline{M}}(t), \zeta_{\overline{H}}(t), \zeta_P(t)\Big) \begin{pmatrix} \pi_L(s)\lambda_L L(t,s)^{\lambda_L - 1} & 0 & 0 \\ 0 & \pi_M(s)\lambda_M M(t,s)^{\lambda_M - 1} & 0 \\ 0 & 0 & \pi_H(s)\lambda_H H(t,s)^{\lambda_H - 1} \\ p_M(s,u_M(t,s)) + p_H(s,u_H(t,s)) & p_Z(s,u_Z(t,s)) & 0 \end{pmatrix} =$$

$$= \begin{pmatrix} (-e(s) - l_M(t,s)u_M(t,s) - l_H(t,s)u_H(t,s))\,\xi_L(t,s) + (e(s) + l_M(t,s)u_M(t,s))\,\xi_M(t,s) + l_H(t,s)u_H(t,s)\xi_H(t,s) + \\ \delta_M(t,s)\,(\xi_L(t,s) - \xi_M(t,s)) - l_Z(t,s)u_Z(t,s)\,(\xi_M(t,s) - \xi_H(t,s)) + \pi_M(s)\lambda_M M(t,s)^{\lambda_M - 1}\zeta_{\overline{M}}(t) + p_Z(s,u_Z(t,s))\zeta_P(t) \\ \delta_H(t,s)\,(\xi_L(t,s) - \xi_H(t,s)) + \pi_H(s)\lambda_H H(t,s)^{\lambda_H - 1}\zeta_{\overline{H}}(t) \end{pmatrix}$$

$$+\pi_L(s)\lambda_L L(t,s)^{\lambda_L - 1}\zeta_{\overline{L}}(t) + \Big(p_M(s,u_M(t,s)) + p_H(s,u_H(t,s))\Big)\zeta_P(t) \Bigg)^T$$

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\xi_L(t,s) = \Big(-e(s) - l_M(t,s)u_M(t,s) - l_H(t,s)u_H(t,s)\Big)\xi_L(t,s) + $$
$$+ \Big(e(s) + l_M(t,s)u_M(t,s)\Big)\xi_M(t,s) + l_H(t,s)u_H(t,s)\xi_H(t,s) + \pi_L(s)\lambda_L L(t,s)^{\lambda_L - 1}\zeta_{\overline{L}}(t) +$$
$$+ \Big(p_M(s,u_M(t,s)) + p_H(s,u_H(t,s))\Big)\zeta_P(t) \tag{B.31}$$

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\xi_M(t,s) = \delta_M(t,s)\Big(\xi_L(t,s) - \xi_M(t,s)\Big) - l_Z(t,s)u_Z(t,s)\Big(\xi_M(t,s) - \xi_H(t,s)\Big) +$$
$$+ \pi_M(s)\lambda_M M(t,s)^{\lambda_M - 1}\zeta_{\overline{M}}(t) + p_Z(s,u_Z(t,s))\zeta_P(t) \tag{B.32}$$

$$-\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\right)\xi_H(t,s) = \delta_H(t,s)\Big(\xi_L(t,s) - \xi_H(t,s)\Big) + \pi_H(s)\lambda_H H(t,s)^{\lambda_H - 1}\zeta_{\overline{H}}(t) \tag{B.33}$$

with the boundary conditions

$$\xi(T,s) = (0,0,0) \qquad \forall s \in [0,\omega] \qquad \text{and} \qquad \xi(t,\omega) = (0,0,0) \qquad \forall t \in [0,T]$$

Also $\zeta(t)$ does not change through the newly added control and we can use the same terms as in appendix B.2. As in the derivations before, we insert the terms for $\zeta(t)$ in the PDEs (B.31) - (B.33). This leads to the following equations for the current-value adjoint variables

$$\begin{aligned} -\Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\Big)\mu_L(t,s) = &\Big(- e(s) - l_M(t,s)u_M(t,s) - l_H(t,s)u_H(t,s)\Big)\mu_L(t,s) + \\ &+ \Big(e(s) + l_M(t,s)u_M(t,s)\Big)\mu_M(t,s) + l_H(t,s)u_H(t,s)\mu_H(t,s) + \\ &+ Y(t)^{1-\rho}\cdot\theta_L(t)\pi_L(s)\cdot\widetilde{L}(t)^{\rho-\lambda_L}L(t,s)^{\lambda_L-1} - \Big(p_M(s,u_M(t,s)) + p_H(s,u_H(t,s))\Big) - r\mu_L(t,s) \end{aligned} \tag{B.34}$$

$$\begin{aligned} -\Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\Big)\mu_M(t,s) = &\delta_M(t,s)\Big(\mu_L(t,s) - \mu_M(t,s)\Big) - l_Z(t,s)u_Z(t,s)\Big(\mu_M(t,s) - \mu_H(t,s)\Big) + \\ &+ Y(t)^{1-\rho}\cdot\theta_M(t)\pi_M(s)\widetilde{M}(t)^{\rho-\lambda_M}M(t,s)^{\lambda_M-1} - p_Z(s,u_Z(t,s)) - r\mu_M(t,s) \end{aligned} \tag{B.35}$$

$$-\Big(\frac{\partial}{\partial t} + \frac{\partial}{\partial s}\Big)\mu_H(t,s) = \delta_H(t,s)\Big(\mu_L(t,s) - \mu_H(t,s)\Big) + Y(t)^{1-\rho}\cdot\theta_H(t)\pi_H(s)\widetilde{H}(t)^{\rho-\lambda_H}H(t,s)^{\lambda_H-1} - r\mu_H(t,s) \tag{B.36}$$

Optimality conditions for control $u(t,s)$

The optimality conditions are again calculated by using the current value Hamiltonian

$$\begin{aligned} \mathcal{H}(t,s,u) &= e^{rt}\cdot\mathcal{L}(t,s,u) + \mu(t,s)f(t,s,u) + \nu(t)h(t,s,u) = \\ &= \frac{1}{\omega}\cdot\bigg(\Big(\theta_L(t)\overline{L}(t)^{\frac{\rho}{\lambda_L}} + \theta_M(t)\overline{M}(t)^{\frac{\rho}{\lambda_M}} + \theta_H(t)\overline{H}(t)^{\frac{\rho}{\lambda_H}}\Big)^{1/\rho} - P(t)\bigg) + \\ &+ \mu(t,s)f(t,s,u) + \nu(t)\begin{pmatrix} \pi_L(s)L(t,s)^{\lambda_L} \\ \pi_M(s)M(t,s)^{\lambda_M} \\ \pi_H(s)H(t,s)^{\lambda_H} \\ \Big(p_M(s,u_M(t,s)) + p_H(s,u_H(t,s))\Big)L(t,s) + p_Z(s,u_Z(t,s))M(t,s) \end{pmatrix} \end{aligned} \tag{B.37}$$

The additional control $u_Z(t,s)$ leads to an additional optimality condition, but doesn't affect the optimality conditions from the previous extension. We obtain the functions

$$\begin{aligned} F_1(u_M(t,s)) &= \underbrace{\nu_P(t)}_{=-1}p_M(s,u_M(t,s))L(t,s) + \mu(t,s)\begin{pmatrix} -l_M(t,s)u_M(t,s)L(t,s) \\ l_M(t,s)u_M(t,s)L(t,s) \\ 0 \end{pmatrix} = \\ &= -p_M(s,u_M(t,s))L(t,s) + \Big[\mu_M(t,s) - \mu_L(t,s)\Big]l_M(t,s)u_M(t,s)L(t,s) \end{aligned} \tag{B.38}$$

$$F_2(u_H(t,s)) = \underbrace{\nu_P(t)}_{=-1} p_H(s, u_H(t,s))L(t,s) + \mu(t,s) \begin{pmatrix} -l_H(t,s)u_H(t,s)L(t,s) \\ 0 \\ l_H(t,s)u_H(t,s)L(t,s) \end{pmatrix} =$$

$$= -p_H(s, u_H(t,s))L(t,s) + \Big[\mu_H(t,s) - \mu_L(t,s)\Big] l_H(t,s)u_H(t,s)L(t,s) \quad \text{(B.39)}$$

$$F_3(u_Z(t,s)) = \underbrace{\nu_P(t)}_{=-1} p_Z(s, u_Z(t,s))M(t,s) + \mu(t,s) \begin{pmatrix} 0 \\ -l_Z(t,s)u_Z(t,s)M(t,s) \\ l_Z(t,s)u_Z(t,s)M(t,s) \end{pmatrix} =$$

$$= -p_Z(s, u_Z(t,s))M(t,s) + \Big[\mu_H(t,s) - \mu_M(t,s)\Big] l_Z(t,s)u_Z(t,s)M(t,s) \quad \text{(B.40)}$$

We can break the optimality conditions down to the following equation, because all assumptions about differentiability and convexity are still satisfied.

$$F_1'(u_M(t,s)) \leq 0, \qquad F_1'(u_M(t,s)) \cdot u_M(t,s) = 0 \quad \text{(B.41)}$$

$$F_2'(u_H(t,s)) \leq 0, \qquad F_2'(u_H(t,s)) \cdot u_H(t,s) = 0 \quad \text{(B.42)}$$

$$F_3'(u_Z(t,s)) \leq 0, \qquad F_3'(u_Z(t,s)) \cdot u_Z(t,s) = 0 \quad \text{(B.43)}$$

The summary of all equations used to calculate the optimal solutions can be found in chapter 4.

C | Numerical Optimization

In this section I will present the algorithm used to derive the numerical results for the basic model in chapter 2 and for the extended models in chapter 3 and 4. The code is based on the work of Frankovic et al. (2016), and was by courtesy of the authors made available to me for the use in my thesis. Nevertheless I needed to adapt it to fit the problem formulation of the models we are analysing and to overcome some numerical challenges.

C.1 The algorithm

The basic approach of the algorithm resembles the gradient descend method and contains the following summarised steps

1. Initialisation with an arbitrary chosen control $u_0(t,s)$ and setting $u_{new} = u_0$.
2. Calculating the corresponding state and co-state variables for $u_{best} := u_{new}$.
3. Calculation of the gradient of the Hamiltonian $\frac{\partial H}{\partial u}$ evaluated at this momentarily best solution.
4. Deriving the "optimal" step-size x_{new}.
5. Calculating a new control $u_{new} = u_{best} + x_{new} \cdot \frac{\partial H}{\partial u}$
6. Repeating the steps 2-5 until an abort criterion is fulfilled.

Below I will give a more detailed explanation and interpretation of each of these steps.

Step 1: Initialisation

The optimisation starts with the arbitrary chosen control $u_0(t,s)$. The level of the control should be chosen realistically to increase the convergence speed or even "assure" convergence of the algorithm at all.[1] For example the starting control can be chosen as constant both over age and time at a reasonable level. It might be beneficial for the calculation time, if one has a rough idea about the qualitative form of the optimal solution and to use this information for the starting solution u_0. This can be obtained by applying the algorithm with a relatively big step-size and transferring this solution to a finer grid.

[1]For a detailed analysis of numerical solutions for heterogeneous control systems see Veliov (2003).

Step 2: Calculation of the states and co-states

Given a control $u_{best}(t, s)$ and data for the initial and boundary conditions we can calculate the state variables $L(t, s)$ and $H(t, s)$ (resp. $L(t, s)$, $M(t, s)$ and $H(t, s)$) as well as the corresponding aggregated variables and the value of the objective function, which we call y_0 in the following.[2] Despite the development of the state variables being given by PDEs, we can apply solving methods for ODEs, because along the characteristic lines illustrated in figure 2.2 an ODE system is fulfilled.[3] To solve the ODEs I chose Heun's method, which is an explicit second order Runge-Kutta method with the Butcher tableau in table C.1 After having calculated

0	0	0
1	1	0
	1/2	1/2

Table C.1: Butcher tableau for Heun's method

all state and aggregated variables we are able to derive the values for the co-state variables $\mu_L(t, s)$ and $\mu_H(t, s)$ with the equations (B.5) and (B.6) (resp. $\mu_L(t, s)$, $\mu_M(t, s)$ and $\mu_H(t, s)$ in equations (B.24)-(B.26)). Again we solve the PDEs along the characteristic lines using Heun's method given the initial and boundary data for the co-states. It should be mentioned, that the ODEs for the co-states have to be solved backwards (in time and age), because the boundary values are given just for the complementary boundary of the time-age space compared to the state variables.

Step 3: The gradient of the Hamiltonian

With the co-state variables we can now calculate the gradient of the Hamiltonian as given by the term in equation (B.9) (resp. equations (B.20) and (B.21)). The gradient defines the direction in which we are going to search the next approximative solution u_{new}. In the classic gradient descent a line-search is applied to find the maximum of the value function along the direction of the gradient by taking the step-size x as the only one-dimensional decision variable. In static optimisation problems the objective function can "simply" be evaluated for a new control $u = u_{best} + x \cdot \nabla H$ and the step-size x can be altered continuously. However in our dynamic problem, we have to solve a couple of PDE systems to obtain the corresponding objective value y for a single step-size x. Therefore the search for the optimal step-size is adjusted as described in step 4.

Step 4: The optimal step-size

Since we are not able to apply classic one-dimensional optimisation methods to find the optimal step-size, we basically do not search for the best solution in direction of the gradient, but for a

[2]See problem 1, problem 2 or problem 3 for the formulas.

[3]Since the the slope of the characteristics is equal to one, we should use the same step-size for the discretisation in both dimensions age and time. This results in a grid with (t_j, s_i) and (t_{j+1}, s_{i+1}) being on the same characteristic line.

reasonable better solution in direction of the gradient. Figure C.1 illustrates the basic scheme used to derive the new approximative solution, i.e. the triple

$$\text{(control, step-size, objective value)} = (u_{new}, x_{new}, y_{new})$$

starting with the triple $(u_{best}, x_{best}, y_{best})$

The general idea is to calculate two new controls u_1 and u_2 (for step-sizes x_1 and x_2) and interpolate the three points $(0, y_0)$, (x_1, y_1) and (x_2, y_2) using a quadratic function $y(x)$. This leads to the following possible scenarios

1. The resulting quadratic function $y(x)$ is convex and it holds that $y_2 > y_1 > y_0$: In this case u_2 is chosen as the new approximative solution and x_2 is the new basic step-size.

2. $y(x)$ is concave: It can be shown that this function exhibits its maximum at the point

$$x_3 = \frac{x_2^2(y_0 - y_1) + x_1^2(y_2 - y_0)}{2(x_1(y_2 - y_0) + x_2(y_0 - y_1))}$$

   Now the corresponding control u_3 and objective value y_3 are calculated and compared to the up to this point best solution.

3. $y(x)$ is convex and the condition $y_2 > y_1 > y_0$ does not hold: By construction of the scheme it follows that $y_0 > y_1$ and $y_0 > y_2$, what means that the starting solution u_0 is still the best ($u_{new} = u_{best}$ and $y_{new} = y_{best}$). This fact denotes that the step-size is too big to get improvement along the direction of the gradient ∇H and therefore the step-size will be changed to $x_{new} = x_3 = \frac{1}{4}x_1$.

4. The concave function attains its maximum at a negative step-size: This can also only be the case if $y_0 > y_1$ and $y_0 > y_2$ and the arguments of the last point also apply here. It results in $((u_{new}, x_{new}, y_{new}) = (u_{best}, \frac{1}{4} \cdot x_{best}, y_{best}))$

Step 5: Defining the new approximative solution

As explained in step 4 the new approximative solution can be determined by the scheme in figure C.1.

Repeating the steps 2-5

With the new approximative solution $(u_{new}, x_{new}, y_{new})$ the step 2-5 can be repeated. Of course the steps 2 and 3 can be skipped if their was no change in the approximative solution u_{new} since the last iteration, as not only the states, co-states and aggregated variables would stay the same, but also the gradient of the Hamiltonian is not affected in this case. The steps 2-5 are now repeated until at least one of the following two abort criteria is fulfilled

- The number of iterations exceeds a predetermined upper limit.
- A chosen norm (like the L_1-Norm) of the gradient of the Hamiltonian is smaller than a predefined boundary ε.

The last solution u_{new} is then taken as the final numerical approximation for the optimal solution $\hat{u}(t,s)$.

C.2 Remarks

In the discussion above we should not forget, that we have non-negativity constraints for our control. So after calculating a new control $u_i = u_0 + x_i \cdot \nabla H$ we have to ensure, that all entries are non-negative, which is simply done by setting all negative values to zero. This is also the reason, why the gradient of the Hamiltonian can exhibit large negative values for an almost optimal solution, what must be taken into account, when calculating the abort criterion.

For certain cases of imperfect substitutability I have also further modified the algorithm to a small extend:

- When the number of workers entering the labour market for a certain skill group is close to zero and the partial elasticity of substitution with respect to age of this skill group is also close to zero, the optimal education rate tends to diverge to infinity when approaching the age of zero.
- In the discrete numerical algorithm this leads to the gradient for the first age point being several powers of ten times the maximum of all other points in the time-age space. Simply applying the algorithm in this case leads to the step-size converging to zero without obtaining any improvement in the solution.
- Therefore I have limited the gradient at the first age-point to be at maximum 10^k times the maximum of all other age-points. (For example k is set equal to $k = 1$ or $k = 2$.)
- As a result the approximation for the first age-group might not be as good as possible, but the downside of this fact is offset by the improvement of the solutions for a big number of other age-groups.

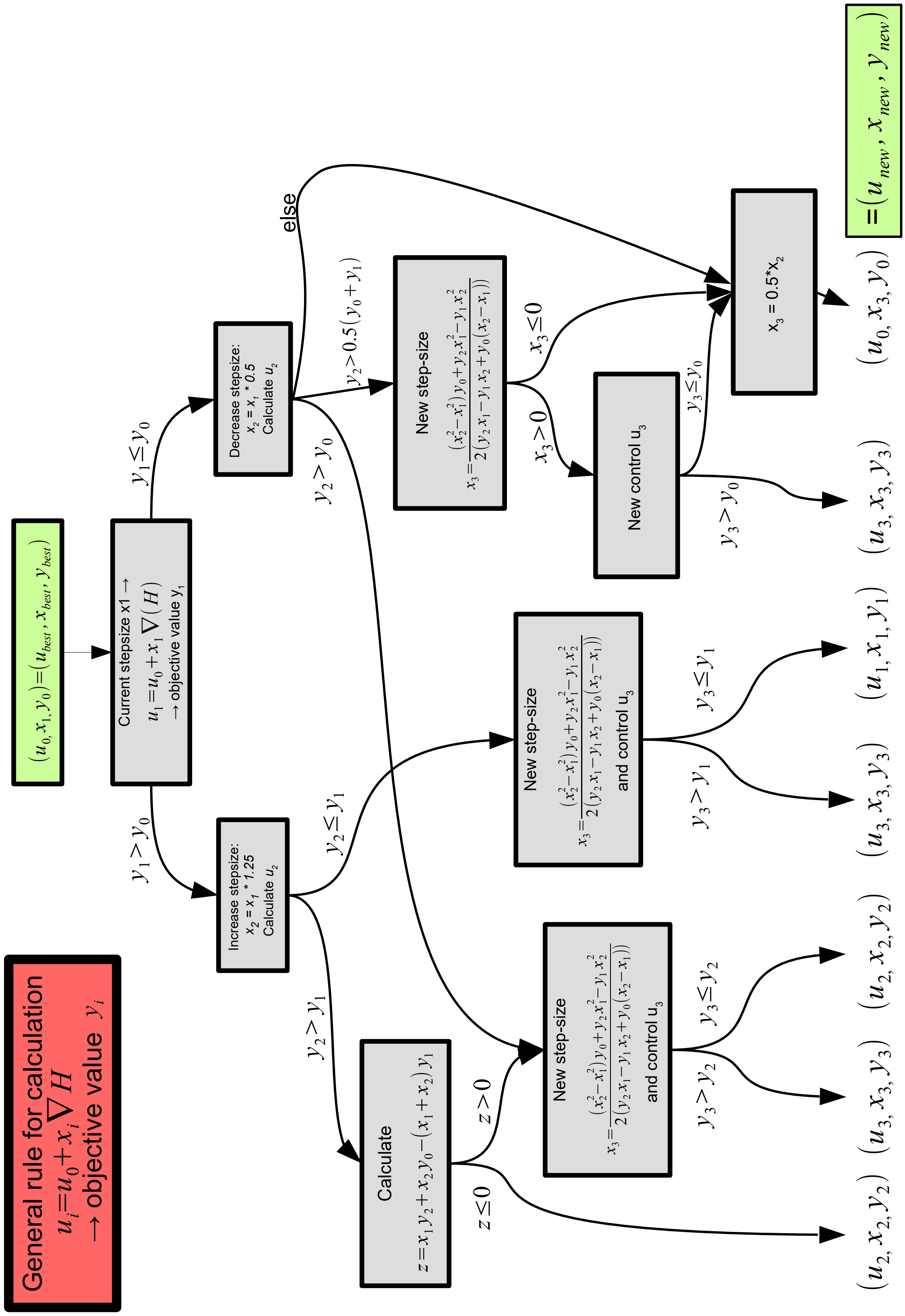


Figure C.1: Finding the next approximative solution, the corresponding objective value and the new step-size along the direction of the gradient.