



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

*Operations
Research and
Control Systems*



Optimal education in an age-structured model under changing labor demand and supply

Alexia Prskawetz, Tsvetomir Tsachev, Vladimir M. Veliov

Research Report 2010-09

March, 2010

Operations Research and Control Systems
Institute of Mathematical Methods in Economics
Vienna University of Technology

Research Unit ORCOS
Argentinierstraße 8/E105-4,
1040 Vienna, Austria
E-mail: orcos@eos.tuwien.ac.at

Optimal education in an age-structured model under changing labor demand and supply

Alexia Prskawetz*

Tsvetomir Tsachev**

Vladimir M. Veliov†

* Institute for Mathematical Methods in Economics (Research Unit Economics), Vienna University of Technology, Argentinierstrasse 8, A-1040 (afp@econ.tuwien.ac.at), and Vienna Institute of Demography, Austrian Academy of Sciences, Vienna, Austria

**Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev str., block 8, 1113 Sofia, Bulgaria (tsachev@math.bas.bg).

† Institute for Mathematical Methods in Economics (Research Unit Operations Research and Control Systems), Vienna University of Technology, Argentinierstrasse 8, A-1040 Vienna, Austria (veliov@tuwien.ac.at).

Abstract

We introduce a model of the optimal education policy at the macro level allowing for heterogeneity of the workforce with respect to its age and qualification skills. Within this framework we study the optimal education rate in the context of changes in the labor demand (as represented by the elasticity of substitution across ages and qualification) and labor supply (as represented by a change in the population growth rates). Applying an age-structured optimal-control model we derive features of the optimal age specific education rate. Our results show that the relation between the elasticities of substitution of labor across ages plays a crucial role for the way the demographic changes affect (both in the short and in the long run) the optimal educational policy. We also show that under imperfect substitutability across age and qualification groups the optimal educational policy is adjusted in advance to any change in the labor supply.

Key words: human capital, age-composition of labor, age-structured model, optimal control

JEL: C61, J23, J24

1 Introduction

Changes in the age distribution of the workforce—as caused by changes in the cohort size of entering labor flows—and its implications for labor market outcomes such as wages and unemployment have been extensively discussed in the empirical economic literature (Freeman 1979; Katz and Murphy 1992; Murphy and Welch 1992; Welch 1979). Besides the age structure, the skill level constitutes a further important heterogeneity of workers in the labor market. Labor market outcomes for different skill groups (in particular wage dispersion) are extensively studied (e.g. Katz and Autor 1999).

In a seminal paper Card and Lemieux 2001 reconcile the work of Welch 1979 and of Katz and Autor 1999. They introduce an aggregate production function that accounts for imperfect substitutability of workers across age and education. Their aim is to explain the empirical fact that the rise in education related wage differentials in the US from 1959-1995 is mainly due to a rise in the college-high school wage gap for younger men while the gap for older men has remained fairly constant over the same time period (1959-1995). Within their framework the authors show "that the increase in the college-high school wage gap over the past two decades is attributable to steadily rising relative demand for college-educated labor, coupled with a dramatic slowdown in the rate of growth of the relative supply of college-educated workers".

As argued in Rojas 2005, the implications of these empirical studies have not yet been integrated into formal macroeconomic models except in Lam 1989 and in Kremer and Thomson 1998. While Lam considered the change in age structure on life-cycle wage profiles in a stable population, Kremer and Thomson study the role of imperfect substitution across workers of different age to explain the speed of convergence of per capita output between countries.

The assumption that workers of different ages are not perfectly substitutable across age within educational groups is also empirically verified by Stapleton and Young 1988. According to them there is empirical evidence that the elasticity of substitution across ages of the unskilled labor is higher than for skilled labor. Recently Roger and Wasmer (2009) also indicated the importance to control for age and skill heterogeneity for explaining labor productivity. The fact of imperfect substitutability across age and education implies that there is an optimal age-education mix of the workforce (in terms of output maximization). Our aim is to provide a model of the optimal education/training policy, i.e. optimal assignment of resources to transform unskilled into skilled labor, at the macro level allowing for heterogeneity of the workforce with respect to its age and qualification skills. Within this framework we study the optimal education/training rate in the context of alternative labor demand and labor supply effects. The degree of substitutability across workers of different age and education characterizes the labor demand pattern. Labor supply is determined by the inflow of low and high skilled workers. Since we assume zero mortality and full employment labor supply is determined by a change in the demographic factors in our model.

Our paper is motivated by the fact that population ageing, as caused by decreasing fertility and increasing survival to older ages, implies a change in the age and educational composition of the workforce and, hence, requires a change in the optimal age-specific education policy. Obviously the

substitutability of workers across age and across qualifications, together with increasing demand for educated workers and technological progress, constitute key factors explaining the relation between demographic change and educational investment at the macro level.

According to the theory of optimal life cycle human capital investment, human capital accumulation concentrates at the beginning of life (Weiss 1986). An economy characterized by population ageing might therefore exhibit an 'older vintage' of human capital. It is therefore of interest how a social planner will react in terms of its age-specific human capital investment under conditions of labor force ageing and faced with alternative labor demand patterns.

So far, formal models on population ageing and human capital formation, concentrate on explaining educational activities at the individual level, e.g. Heijdra and Romp 2009 and Boucekkine et al. 2002, and its implication for the long run economic growth. Increased longevity will increase educational activities at the individual level and may boost economic growth at the aggregate level. In this literature, the pattern of labor demand and its interrelationship with changing demographic supply, is not considered. In contrast to these models we start at the macro-level and consider the optimal age-specific education policies if workers of different age and skill are not perfectly substitutable in the work process and if labor supply is determined by the prevailing demographic structure.¹

In summary, our paper is intended to raise and address the following questions:

- (i) Compared to the case of perfectly substitutable labor across ages and across education, how will the optimal age-specific educational policy change under conditions of imperfectly substitutable labor across age and across qualification groups.
- (ii) Compared to the case of a stationary population, how will an increasing or decreasing population (hence change in the labor supply) affect the optimal educational policy.
- (iii) Will the long-run and short-run effects of a change in the demand and supply factors of labor differ? In particular, will a change in demographic factors be anticipated in the optimal educational policy in advance of the time point where the actual change takes place?

The main methodological instrument employed in the paper in order to answer the above questions is that of intertemporal dynamic optimization in a model in which time and age change continuously. The dynamics of the human capital is then described by an *age-structured system of differential equations*, in contrast to overlapping generations (OLG) models involving a presumably small number of co-existing generations. The paper exhibits several advantages of the continuous-age framework. First, the obtained results are independent of the number of age groups (in contrast to OLG models where the number of age groups may have an effect on the results). Second, working in continuous time/age allows to use more efficiently standard analytic tools such as derivatives and integrals. Third, some of the results (that in Section 4, in particular) cannot be established

¹Hence, while the social planner at the macro level takes into account changes in the supply of labor, an individual considers the demographic structure as constant. Moreover when an individual invests into age specific education he/she ignores that he/she thereby also influences the skill distribution at the macroeconomy at any instant of time. An obvious externality therefore arises since the skill distribution together with the prevailing labor demand parameters will impinge on the returns to education for each individual.

at all in the OLG framework unless a rather fine differentiation of age groups is used (above 10 age groups for the result in Section 4, for example.) Finally, the continuous time/age model is advantageous also from a numerical point of view, since it allows for a more profound time/age aggregation (discretization) than the a priori discretization involved in the OLG models (see e.g. Section 5 in Veliov, 1997).

Optimality conditions for the continuous-time/age dynamic optimization problem are presented and discussed in Section 3. The conditions are derived in Appendix 2 using the general result in Feichtinger et al. 2003.

While a (continuous) age structure has been introduced into the neoclassical theory of optimal investment in several recent contributions (e.g. Barucci and Gozzi 2001; Feichtinger et al. 2005, 2006), we are only aware of the paper by Christiaans 2003 and our own work Prskawetz and Veliov 2007 that augmented the classical theory of labor demand by modelling the age structure of the workers. The vintage structure of human capital has been integrated into a model of growth and technological diffusion by Chari and Hopenhayn 1991. The authors assume that different vintages coexist each time period and operate with technology specific skilled and unskilled human capital. The equilibrium distribution of skilled workers across the vintages is derived endogenously depending on consumers optimal supply of unskilled and skilled labor force and vintage specific optimal demand for skilled and unskilled workers. Similar to our model, new and old human capital may not be perfectly substitutable in the production process. While Chari and Hopenhayn only allow for young unskilled and old skilled or unskilled labor, and differentiate between vintages of production, we ignore the latter assumption but allow for continuous age specific labour. More recently the vintage structure of human capital has also been introduced into a model of endogenous growth by Boucekkine et al. 2002.

In the model we present in Section 2 the aggregate production depends on the age and skill specific supply of labor. A general production function of CES type is employed, which reflects the imperfect substitutability between different age groups and between labor with different qualification.

In Section 4 we prove that in case of perfect substitutability of workers across age and qualification, the optimal educational rate is independent of demographic changes. Furthermore we show that it might be non-monotonic in age, that is, it might be profitable to postpone education to somewhat older ages in order to benefit from the cost-free “learning by doing” and to take advantage from the more advanced knowledge that one can obtain at a later time.

The situation substantially changes in case of non-perfect substitutability of workers across age and qualification, and this is a key point in the paper. In Section 5 we show that the effect of demographic changes on the formation of human capital critically depends on the elasticity of substitution across ages and across qualifications. In addition, this effect may be qualitatively different in the short versus the long run. Changes in the age-specific supply of workers together with the prevailing demand structure of the economy—as reflected by the parameters of the elasticity of substitution—determine the evolution of the human capital. The analysis clearly exhibits the importance of the relation between the elasticities of substitution across age and qualification, which may lead to

different (opposite) impacts of the demographic change on the optimal educational policy. Moreover, it is shown that a change in the labor supply influences the optimal education policy already in advance of the time point when the change takes place.

In Section 6 we study the path of the wage differential between high and low skilled workers over time under the assumption that workers of higher skill are less substitutable across ages compared to lower skilled workers and for different scenarios of the population growth rates. The final section concludes the paper.

The data specifications, the derivation of the optimality conditions and some longer proofs are given in three appendices.

2 The model

Below $t \in [0, T]$ denotes the time, T is the end of the (presumably large) planning horizon. In our model the workers will be distinguished by their “active” age, s . It is assumed that all individuals start working at the same age $s = 0$ (which corresponds to, say, 20 years of biological age) and retire at age $s = \omega$. We distinguish the workers also by their skills, considering for simplicity two levels of qualification: low-skilled and high-skilled workers. We denote by $L(t, s)$ the amount of low-skilled workers of age s at time t , and similarly, by $H(t, s)$ – the amount of high-skilled workers resulting from the government’s investment in human capital.

Upgrading of low-skilled workers into high-skilled workers takes place with a rate $l(t, s)u(t, s) + e(s)$. Here $u(t, s)$ denotes the educational rate at time t for workers of age s . The function $l(t, s)$ reflects the dependence of the learning abilities of the workers on time and age, and $e(s)$ represents the learning by doing which depends on the years spent working. At the same time, due to the technological progress or other reasons, high-skilled workers of age s may lose their skills with a rate $\delta(t, s)$.

The dependence of the learning abilities on age is investigated in Pfeiffer and Reuß 2007. As the authors demonstrate, though older persons might already have high levels of skills as opposed to younger persons, their learning ability is slower as compared to a younger person. They also show that skill depreciation accelerates with age. Besides the age-specific dependence of learning and depreciation, learning abilities may also depend on time due to the technological progress.

The equations for the dynamics of the stock of low- and high-skilled workers are therefore

$$\begin{aligned} L_t + L_s &= \delta(t, s)H(t, s) - e(s)L(t, s) - l(t, s)u(t, s)L(t, s), & L(t, 0) &= L_0(t), \\ H_t + H_s &= -\delta(t, s)H(t, s) + e(s)L(t, s) + l(t, s)u(t, s)L(t, s), & H(t, 0) &= H_0(t), \\ & & L(0, s), H(0, s) &- \text{given initial data.} \end{aligned}$$

The left hand side, $L_t + L_s = \lim_{h \rightarrow 0} (L(t+h, s+h) - L(t, s))/h$, represents the change in one unit of time of the low-skilled labor that is of age s at time t . This change is composed of downgrading high-skilled to low-skilled workers (decay of human capital) at the rate $\delta(t, s)$ and of upgrading low-skilled to high-skilled workers at the rate $e(s)$ due to costless on the job learning by doing and at the rate $l(t, s)u(t, s)$ due to costly education. Similarly, the left hand side, $H_t + H_s$ represents the change in one unit of time of the high-skilled labor. This change is composed of the same components as the change of the low-skilled labor, but with the opposite sign.

At age $s = 0$ the number of those who enter the work force at time t is $L_0(t)$ for the low-skilled and $H_0(t)$ for the high-skilled. We assume throughout the paper that $H_0(t)$ is for each t relatively small compared with $L_0(t)$. This reflects the fact that all but a few high school graduates enter the work force as low-skilled workers and have to undergo additional education/training in order to become high-skilled.

We assume also that there is no unemployment and there is no mortality at working ages, therefore the sum $L(t, s) + H(t, s)$ equals the total working age population, which is determined by the exogenously given total inflow, $N_0(t)$. Thus $L_0(t) + H_0(t) = N_0(t)$. This allows to exclude the variable $H(t, s)$ from the model and to pass to a single differential equation for $L(t, s)$. However, for the purposes of better transparency of the exposition we shall work with both equations for $L(\cdot, \cdot)$ and $H(\cdot, \cdot)$.

The price of the per capita education $u(t, s)$ is $p(s, u(t, s))$, where $p(\cdot, \cdot)$ is a given function of (s, u) . The total cost $P(t)$ of the educational effort for the society at time t is therefore represented as

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

We allow for imperfect substitutability across age groups for both low-skilled and high-skilled labour. If $\pi_L(\cdot)$ and $\pi_H(\cdot)$ are the respective relative efficiency parameters (assumed to be fixed over time), the two sub-aggregates of low-skilled and high-skilled labour at time t , $\tilde{L}(t)$ and $\tilde{H}(t)$, are given by the following two CES functions

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

Here $\lambda_L \in (-\infty, 1]$ and $\lambda_H \in (-\infty, 1]$ give the respective partial elasticity of substitution $\frac{1}{1-\lambda_i}$ for $i = L, H$. In the limiting case of perfect substitutability across age groups λ_L and λ_H are equal to 1 and the aggregate of low-skilled and skilled-labor is simply a weighted sum of age-specific supply.

We assume that the production technology depends only upon labor. The aggregate output at time t is given by a CES function of the two sub-aggregates of low-skilled and high-skilled labor, in which the technological level is represented by the two efficiency parameters $\theta_L(t)$ and $\theta_H(t)$:

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

Here again $\rho \in (-\infty, 1]$ gives the partial elasticity of substitution $\sigma_W = \frac{1}{1-\rho}$ between high-skilled and low-skilled labor. Note that the marginal product of labor for a given age-education group depends on the group's own supply of labor and the aggregate supply of labor in the education category.

The net revenue of the society at time t is the aggregate output $Y(t)$ minus the cost of education, $P(t)$, i.e. $Y(t) - P(t)$.

The formal problem of a central planner is to maximize the accumulated discounted net revenue by choosing optimally the educational rate $u(t, s)$.² Discounting the future with a rate $r \geq 0$ we come up with the following dynamic optimization problem with state variables $L(t, s)$, $H(t, s)$, $\tilde{L}(t)$, $\tilde{H}(t)$ and $P(t)$, and control variable $u(t, s)$:

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

subject to

$$L_t + L_s = \delta(t, s)H(t, s) - e(s)L(t, s) - l(t, s)u(t, s)L(t, s), \quad L(t, 0) = L_0(t), \quad (2)$$

$$H_t + H_s = -\delta(t, s)H(t, s) + e(s)L(t, s) + l(t, s)u(t, s)L(t, s), \quad H(t, 0) = H_0(t), \quad (3)$$

$L(0, s)$, $H(0, s)$ – given initial data,

$$\tilde{L}(t) = \left(\int_0^\omega \pi_L(s)(L(t, s))^{\lambda_L} ds \right)^{1/\lambda_L}, \quad (4)$$

$$\tilde{H}(t) = \left(\int_0^\omega \pi_H(s)(H(t, s))^{\lambda_H} ds \right)^{1/\lambda_H}, \quad (5)$$

$$P(t) = \int_0^\omega p(s, u(t, s))L(t, s) ds, \quad (6)$$

$$u(t, s) \geq 0. \quad (7)$$

Since our model does not include the education as a separate sector (cf. Lucas 1988; Boucekine and Ruiz-Tamarit 2008), the control u is directly interpreted as on-the-job training. However, the results are not limited to this interpretation. Indeed, let the workers of different qualification be perfectly substitutable ($\rho = 1$) and let the unskilled workers of different ages also be perfectly substitutable ($\lambda_L = 1$). Then an amount of labor, uL , of age s allocated at time t to education instead of production has a total cost consisting of the opportunity cost $\pi_L(s)\theta_L uL$ (from lost production) and the cost of education. Both costs may be included in $p(s, u)L$, therefore our model completely covers the case of college/university education in the case of linear production function.

²Note that compared to our previous model at the firm level (Prskawetz and Veliov 2007) the current model allows for a much more flexible and empirically relevant production function that includes imperfect substitutability across age and education. While at the firm level the hiring and firing of workers constituted—in addition to the educational investment—a control variable, the educational rate is the only control variable in the social planner model since labor inflow is solely determined by demographic developments in the current model.

In the case of non-perfect elasticity of substitution either across ages or across qualifications the learning should be (strictly speaking) interpreted as costly “training”.

The next assumptions will be supposed to hold throughout the paper.

Standing Assumptions: All exogenous functions are continuous as well as the derivatives that appear later on in the text, excluding the population inflow data $L_0(t)$ and $H_0(t)$, which are assumed only piecewise continuous. The cost function $p(s, u)$ is monotonically increasing and strongly convex with respect to u . The initial and the boundary data $L(0, s)$, $H(0, s)$, $L_0(t)$ and $H_0(t)$, as well as the efficiency coefficients θ_L , θ_H , π_L and π_H are strictly positive.

The strong convexity assumption for $p(s, \cdot)$ provides a substantial mathematical convenience, as usual. At the same time it can be economically justified: the bigger the fraction of the low-skilled workers of a certain age which are involved in education, the larger the per capita educational cost due to the heterogeneity of people with respect to their abilities (the best are taken first in the educational process).

For a precise definition of the notion of a solution to system (2)–(6) and the appropriate space settings we refer to Webb 1985 and to Feichtinger et al. 2003.

3 Optimality conditions

The necessary optimality conditions for problem (1)–(7) follow from the Pontryagin type maximum principle obtained in Feichtinger et al. 2003. A sketch of the derivation is given in Appendix 2, where the following result is obtained.

Proposition 1 *If $(L, H, \tilde{L}, \tilde{H}, P, u)$ is a solution of the optimal control problem (1)–(7) then the equation*

$$\begin{aligned} \Delta_t + \Delta_s &= (r + e(s) + \delta(t, s) + l(t, s)u(t, s))\Delta - p(s, u(t, s)) - f(t, s), \\ \Delta(t, \omega) &= 0, \quad \Delta(T, s) = 0 \end{aligned} \quad (8)$$

with

$$f(t, s) = Y(t)^{1-\rho} \theta_H(t) \pi_H(s) \tilde{H}(t)^{\rho-\lambda_H} H(t, s)^{\lambda_H-1} - Y(t)^{1-\rho} \theta_L(t) \pi_L(s) \tilde{L}(t)^{\rho-\lambda_L} L(t, s)^{\lambda_L-1}$$

has a unique solution $\Delta(t, s)$ and for (almost) every $(t, s) \in [0, T] \times [0, \omega]$ the optimal $u(t, s)$ maximizes the function $\Delta(t, s)l(t, s)u - p(s, u)$ over all $u \geq 0$.

Here $\Delta(t, s)$ represents the difference between the marginal values (“shadow prices”) of skilled and non-skilled labor, that is, the marginal benefit of transforming one unit of low-skilled labor into

one unit of high-skilled. In the above equation $(r + \delta(t, a))\Delta(t, s)$ represents the opportunity cost, f is the age-specific marginal gain in productivity from this transforming. The remaining terms $(e(s) + l(t, s)u(t, s))\Delta(t, s) - p(s, u(t, s))$ in the right-hand side of (8) represent the fact that by being skilled the educational cost $p(s, u)$ is saved, but with it the worker loses the opportunity of becoming skilled later at the rate $e(s) + l(t, s)u(t, s)$.

According to the standing assumptions the derivative $p_u(s, u)$ exists and is invertible with respect to u . Denote by $z \rightarrow p_u^{-1}(s, z)$ the inverse and define

$$p_{u+}^{-1}(s, z) = \begin{cases} p_u^{-1}(s, z) & \text{if } z > p_u(s, 0), \\ 0 & \text{elsewhere.} \end{cases}$$

Then the unique maximizer of $\Delta(t, s)l(t, s)u - p(s, u)$ subject to $u \geq 0$ can be written as

$$u(t, s) = p_{u+}^{-1}(s, l(t, s)\Delta(t, s)). \quad (9)$$

Notice that if $l(t, s)\Delta(t, s) > p_u(s, 0)$ for some (t, s) then $u(t, s) > 0$ and (9) reads as $p_u(s, u(t, s)) - l(t, s)\Delta(t, s) = 0$. Otherwise $u(t, s) = 0$. For example, with the usual specification $p(s, u) = b(s)u + 0.5cu^2$ the optimal control takes the form

$$u(t, s) = \begin{cases} (\Delta(t, s) - b(s))/c & \text{if } \Delta(t, s) > b(s), \\ 0 & \text{elsewhere.} \end{cases} \quad (10)$$

Thus educational effort is applied only for those ages for which the marginal cost of education at $u = 0$ is exceeded by the benefit of the education, measured by the difference between the shadow prices of skilled and unskilled labor.

Substituting u from (9) in (8) we obtain that

$$\begin{aligned} \Delta_t + \Delta_s &= (r + e(s) + \delta(t, s) + l(t, s)p_{u+}^{-1}(s, l(t, s)\Delta))\Delta - p(s, p_{u+}^{-1}(s, l(t, s)\Delta)) - f(t, s), \\ \Delta(t, \omega) &= 0, \quad \Delta(T, s) = 0. \end{aligned} \quad (11)$$

It is important to notice that the function $p_{u+}^{-1}(s, \cdot)$ is Lipschitz continuous, therefore (8), together with the side conditions $\Delta(t, \omega) = 0$ and $\Delta(T, s) = 0$, uniquely determines the solution $\Delta(t, s)$, hence also the optimal control u by (9).

4 The case of perfect substitutability of labor

In this section we consider the simplest case, in which problem (1)–(7) becomes linear with respect to the state variables³: the case of perfectly substitutable labor across ages and across qualifications, $\rho = \lambda_L = \lambda_H = 1$.

³We stress that the problem is still nonlinear due to the bilinear dependence on (L, u) .

Proposition 2 *The optimal educational rate $u(t, s)$ is independent of the initial data $L(0, s)$, $H(0, s)$, and of the low- and high-skilled workers inflows $L_0(t)$ and $H_0(t)$. If all data are time-invariant (except for $L_0(t)$ and $H_0(t)$) then $u(t, s) = u(s)$ is also time-invariant in the time-interval $[0, T - \omega]$.*

The proposition implies, in particular, that a demographic change (represented by $L_0(t)$ and $H_0(t)$)⁴ would not have any influence on the optimal educational rate. Moreover, in the stationary case the end of the time-horizon T may influence the optimal control no longer than one generation time before the end of the horizon, that is, on $[T - \omega, T]$ only.

Proof. To prove the first claim of the proposition we note that now $f(t, s) = \theta_H(t)\pi_H(s) - \theta_L(t)\pi_L(s)$ in (8) and the right-hand side of (8) and the side conditions do not depend on L and H , therefore on the data $L(0, s)$, $H(0, s)$, $L_0(t)$ and $H_0(t)$. Since Δ does not depend on the demographic data, so does u because of (9).

In the time invariant case we have $f(t, s) = f(s)$. Let us denote $z(\tau; a) = \Delta(\tau - a, \omega - a)$, where $\tau \in [0, T]$ is a parameter and $a \in [0, \min\{\tau, \omega\}]$. Then, denoting $a' = \omega - a$, we have

$$-\frac{d}{da}z = (r + \delta(a') + e(a') + l(a')p_{u+}^{-1}(a', l(a')z))z - f(a') - p(a', p_{u+}^{-1}(a', l(a')z)).$$

Clearly, the right-hand side is independent of τ , and since we have $z(\tau, 0) = \Delta(\tau, \omega) = 0$, the solution $z(\tau; a) = z(a)$ is independent of τ . For $t \in [0, T - \omega]$ and $s \in [0, \omega]$ we have

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

and the right-hand side is independent of t (since $t + \omega - s \in [0, T]$ if $t \in [0, T - \omega]$). Using (9) we complete the proof. *Q.E.D.*

Several papers on optimal education and human capital formation (cf. Boucekkine et al. 2002 and Weiss 1986 for a review of the theoretical literature on optimal life-cycle human capital investment) assume or conclude that the educational efforts are optimally allocated in youngest ages and decrease with age. We established numerically that this is not always true in our model, namely, the optimal educational rate, $u(s)$, may strictly increase at certain ages. Below we analyze mathematically the reason for this effect and give some economic explanations. To make the analysis more transparent we present it in the case of time-invariant data and a quadratic cost function $p(u)$. Moreover, it is clear that a decreasing learning ability $l(s)$ encourages learning in young ages and cannot be a reason for increasing learning with age. Therefore we assume the learning ability to be constant.

⁴We can denote a change in the initial level of low-skilled and high-skilled labor a demographic change since we abstract from an endogenous modelling of the labor market and assume full employment.

Proposition 3 Assume that all data (except for $L_0(t)$ and $H_0(t)$) are time-invariant and continuous in s , $l(s) = l$, $p(u) = bu + \frac{c}{2}u^2$. Let there exist some s for which $u(s) > 0$ and such that one of the following conditions holds:

$$\begin{aligned} & \text{(i) } d(s) - \frac{b}{c}l \geq 0 \quad \& \quad \theta_H \pi_H(s) - \theta_L \pi_L(s) < \frac{b^2}{2c}; \\ \text{or} & \\ & \text{(ii) } d(s) - \frac{b}{c}l < 0 \quad \& \quad \theta_H \pi_H(s) - \theta_L \pi_L(s) < \frac{b}{l}d(s) - \frac{c}{2l^2}(d(s))^2, \end{aligned}$$

where $d(s) = r + \delta(s) + e(s)$. Then there exists $s_0 \in (0, \omega)$ where u is differentiable and $u'(s_0) > 0$.

Proof. According to (10), in the time-invariant case

$$u(s) = \max\{0, \frac{1}{c}(l\Delta(s) - b)\}, \quad (12)$$

$$\Delta'(s) = (r + \delta(s) + e(s) + lu(s)) \Delta(s) - f(s) - \left(bu(s) + \frac{c}{2}(u(s))^2\right), \quad (13)$$

where as before $f(s) = \theta_H \pi_H(s) - \theta_L \pi_L(s)$. Hence, the optimal control is a continuous function. Let a be the minimal number such that $u(s) = 0$ on $(a, \omega]$. Since u is not identically zero, we have $a > 0$. Clearly $u(s) > 0$ on some maximal (to the left) interval (a_0, a) . Then

$$u(s) = \frac{1}{c}(l\Delta(s) - b) \quad \text{for } s \in (a_0, a),$$

In particular, u is differentiable on (a_0, a) .

Assume that $u'(s) \leq 0$ on (a_0, a) . Apparently this implies $a_0 = 0$. Moreover, $\Delta(s)$ must satisfy $\Delta'(s) \leq 0$ on $(0, a)$. On this interval Δ satisfies the equation (resulting from (13) after substitution of $u(s)$ from (12))

$$\begin{aligned} \Delta'(s) &= \left(r + \delta(s) + e(s) + \frac{l}{c}(l\Delta(s) - b)\right) \Delta(s) - \\ & \quad f(s) - \frac{b}{c}(l\Delta(s) - b) - \frac{1}{2c}(l\Delta(s) - b)^2. \end{aligned} \quad (14)$$

Rearranging the terms we obtain

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

where

$$q_2 = \frac{l^2}{2c}, \quad q_1(s) = r + \delta(s) + e(s) - \frac{bl}{c}, \quad q_0 = f(s) - \frac{b^2}{2c}. \quad (15)$$

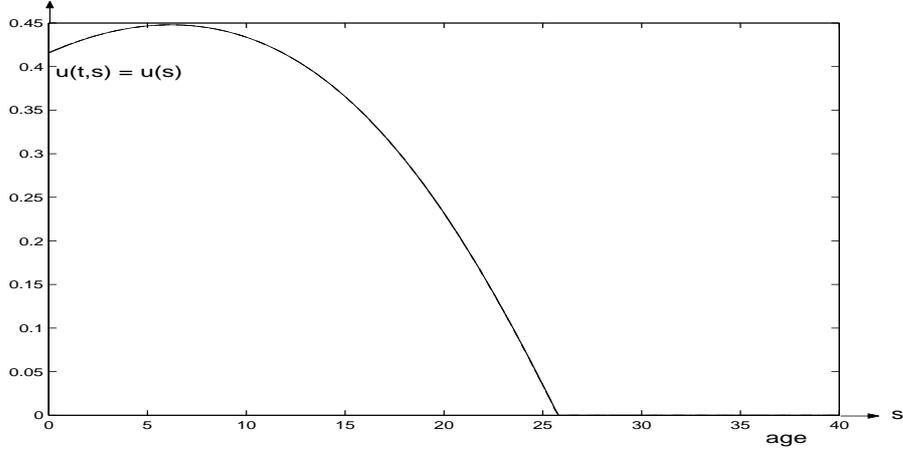


Figure 1: Age distribution of the optimal training rate.

Since $\Delta'(s) \leq 0$, then for every $s \in (0, a)$ the quadratic form $q_2x^2 + q_1(s)x - q_0(s)$ takes a non-positive value for some $x \geq 0$. This means that

$$\min_{x \geq 0} \{q_2x^2 + q_1(s)x - q_0(s)\} \leq 0.$$

Calculating this minimum we obtain two possibilities.

Case (i): If $-\frac{q_1(s)}{2q_2} \leq 0$ then $q_0(s) \geq 0$,

Case (ii): If $-\frac{q_1(s)}{2q_2} > 0$ then $(q_1(s))^2 + 4q_0(s)q_2 \geq 0$.

Substituting from (15) one obtains that Case (i) contradicts assumption (i) and Case (ii) contradicts the alternative assumption (ii) of the proposition. This completes the proof. *Q.E.D.*

The sufficient conditions for non-monotonic behaviour of the optimal educational rate are fulfilled if

(i) $d(s) = r + \delta(s) + e(s)$ is sufficiently large for some s (larger than bl/c) and $f(s) = \theta_H\pi_H(s) - \theta_L\pi_L(s)$ is sufficiently small for this s (smaller than $b^2/2c$).

(ii) $d(s)$ is not that large, but $f(s)$ is small enough, now depending on the value of $d(s)$.

In economic terms, an increase of the educational rate with age would happen if for some age the sum of depreciation (r), dequalification ($\delta(s)$) and “learning by doing” ($e(s)$) rates is large relative to the productivity differential f . If in a certain age interval the productivity differential is small, this may lead to postponement of learning since in the short run the returns to education are small. The returns to education may be reduced by the fact that people can lose their costly qualification (with rate $\delta(s)$), and can make use of the costless “learning by doing” (with rate $e(s)$). In addition, higher depreciation rate diminishes the role of the length of the time-interval in which the worker exercises his qualification, therefore increases the chances for non-monotone learning.

We have numerically found the optimal educational policy for the perfect substitutability case (the respective data is presented in Appendix 1). Figure 1 shows that increasing learning with age may happen even when $\delta = e = 0$. We mention that a higher value of δ can be associated with a higher rate of technological progress. Then higher technological progress may lead to postponement of learning to older ages in order to take advantage of the more advanced knowledge.

5 Effects of the demographic factor in case of imperfect substitutability of labor

5.1 Long run effects

The demographic factor is represented in our model by the exogenous inflow of new labor, $N_0(t)$, split into $L_0(t) + H_0(t) = N_0(t)$. Since we assume full employment a change in the demographic factor is equivalent to a change in labor supply. In order to investigate the impact of a demographic change on the optimal educational policy we compare three scenarios: constant, increasing, and decreasing population, by choosing $L_0(t) = L_0 e^{\gamma t}$, $H_0(t) = H_0 e^{\gamma t}$, where γ is zero, positive, or negative growth rate, respectively. Since all the equations in the model and the objective function are homogeneous of first order, the optimal solution is independent of the size of the population. Hence, not the different population sizes are responsible for the different solutions in the three scenarios that we encounter below, rather *the differences in the age-distribution* of the populations with different growth rates.

The elasticity of substitution across ages for low and for high skilled labor are determined by λ_L and λ_H , respectively. We focus our investigation in this section on how the demographic factor influences the optimal learning rates in different ages in the case of non-perfect substitutability of workers, taking $\lambda_L < 1$ and/or $\lambda_H < 1$. Therefore, in order to shorten the formulas we assume perfect substitutability across qualifications: $\rho = 1$. Moreover, we assume that all data (except for $L_0(t)$ and $H_0(t)$) are time invariant.

We change the variables $L^\gamma(t, s) = e^{-\gamma(t-s)}L(t, s)$, $H^\gamma(t, s) = e^{-\gamma(t-s)}H(t, s)$ in equations (2), (3) (in which the boundary data $L_0(t)$ and $H_0(t)$ are specified as explained above). In fact, L^γ and H^γ satisfy exactly the same equations (2), (3), only the boundary conditions become constant: L_0 and H_0 , respectively. Also the initial conditions change correspondingly. We next discuss the long run behaviour of the optimal solution in which the initial data are irrelevant.

Substituting (L, H) by (L^γ, H^γ) also in equations (4)–(6) and in the objective function (1) we obtain the following equivalent problem:

$$\max \int_0^T e^{-(r-\gamma)t} \left[\theta_L \tilde{L}^\gamma(t) + \theta_H \tilde{H}^\gamma(t) - P^\gamma(t) \right] dt \quad (16)$$

subject to

$$L_t^\gamma + L_s^\gamma = \delta(s)H^\gamma(t, s) - e(s)L^\gamma(t, s) - l(s)u(t, s)L^\gamma(t, s), \quad L^\gamma(t, 0) = L_0, \quad (17)$$

$$H_t^\gamma + H_s^\gamma = -\delta(s)H^\gamma(t, s) + e(s)L^\gamma(t, s) + l(s)u(t, s)L^\gamma(t, s), \quad H^\gamma(t, 0) = H_0, \quad (18)$$

$$L^\gamma(0, s), \quad H^\gamma(0, s) - \text{given data,}$$

$$\tilde{L}^\gamma(t) = \left(\int_0^\omega e^{-\gamma\lambda_L s} \pi_L(s) (L^\gamma(t, s))^{\lambda_L} ds \right)^{1/\lambda_L}, \quad (19)$$

$$\tilde{H}^\gamma(t) = \left(\int_0^\omega e^{-\gamma\lambda_H s} \pi_H(s) (H^\gamma(t, s))^{\lambda_H} ds \right)^{1/\lambda_H}, \quad (20)$$

$$P^\gamma(t) = \int_0^\omega e^{-\gamma s} p(s, u(t, s)) L^\gamma(t, s) ds, \quad (21)$$

$$u(t, s) \geq 0. \quad (22)$$

Thus we obtain that (L, H, u) solves our initial problem with an exponentially changing population ($L_0(t) = L_0 e^{\gamma t}$, $H_0(t) = H_0 e^{\gamma t}$) if and only if (L^γ, H^γ, u) solves the same problem for a constant population ($L_0(t) = L_0$, $H_0(t) = H_0$), but with modified data:

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

The above simple transformation deserves some comments. First of all we established that the optimal educational policy for an exponentially changing population is exactly the same as for a stationary population with the above modified data. Assume for a moment that $\gamma > 0$. The modified per capita cost of learning, $p^\gamma(s, u)$, decreases with age compared to $p(s, u)$. This does not imply that there would be more learning at old ages, since the efficiency coefficients $\pi_L^\gamma(s)$ and $\pi_H^\gamma(s)$ decrease with s , too. However, we point out that $\pi_L^\gamma(s)$ and $\pi_H^\gamma(s)$ decrease with age with different rates. For instance, if $\lambda_H < \lambda_L$, then the efficiency of the high-skilled labor decreases with age less than that of low-skilled labor. This observation leads to the suggestion that for $\gamma > 0$ and $\lambda_H \ll \lambda_L$ there would be a relative shift of learning from younger to older ages. This means, that the normalized distribution of the optimal learning, $\nu(t, s) = u(t, s) / \int_0^\omega u(t, \sigma) d\sigma$, will be shifted to older ages compared with the stationary case $\gamma = 0$. The situation is similar for a decreasing population ($\gamma < 0$), only ‘‘increase’’ and ‘‘decrease’’ should be interchanged. This is seen on Figure 2, where $\lambda_H = 0.1$, $\lambda_L = 0.9$. Both the absolute (left plot) and the normalized (right plot) learning rate shift to older ages for the increasing population, and to younger ages for the decreasing population.

The situation is opposite if $\lambda_H \gg \lambda_L$. This is clearly supported by Figure 3, which shows that the learning age-density, ν shifts to older ages if the population decreases, and to young ages if it increases (the right plot gives $\nu(40, \cdot)$). The left plot represents the optimal learning $u(t = 40, s)$ for the three populations showing that in addition to the age-shift there is an increase of learning at all ages for the decreasing population.

Intuitively, this change in the age-specific learning rate can be explained as follows. In case of an increasing population the additional labor entering the market would lead to a shift from the

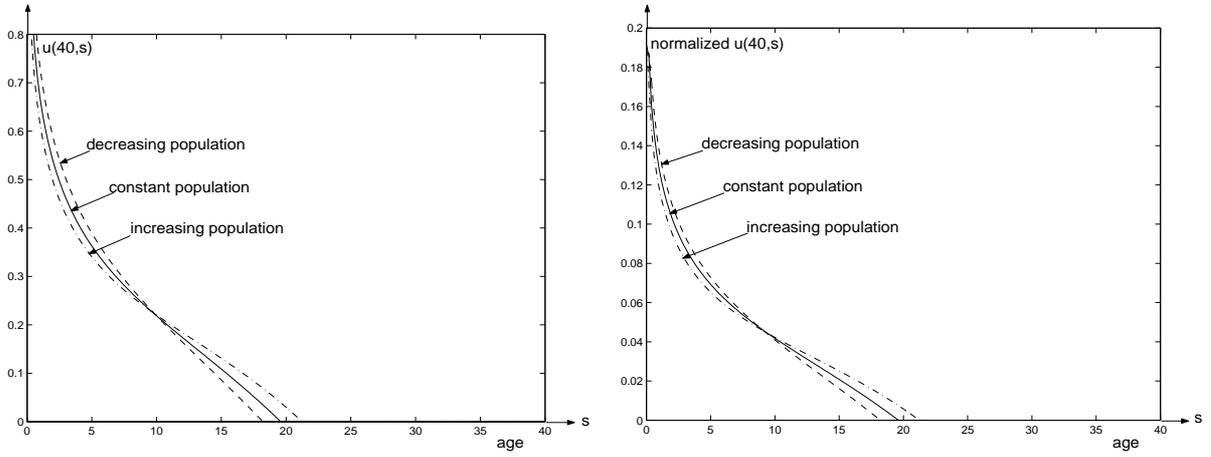


Figure 2: Optimal training rate u (left) and the normalized age-density ν (right) at $t = 40$ for the three scenarios. Here $\lambda_L = 0.9$, $\lambda_H = 0.1$.

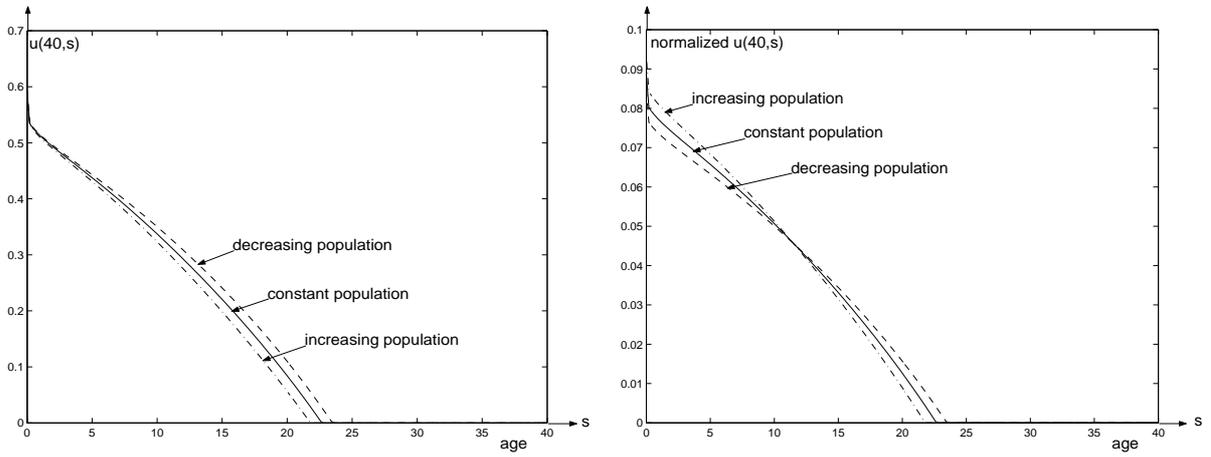


Figure 3: Optimal training rate u (left) and the normalized age-density ν (right) at $t = 40$ for the three scenarios. Here $\lambda_L = 0.1$, $\lambda_H = 0.9$.

optimal age composition for the stationary case towards an excessive amount of young unskilled labor. If the educational rate remains unchanged, then this shift will result in a shift towards lower ages of the age distribution also of skilled labor. Thus keeping the educational rate the same as for the stationary population leads to excessive amount of young workers in both qualification groups. Now we have to distinguish two cases. If $\lambda_H \ll \lambda_L$, then the distortion of the optimal age distribution of high-skilled labor would be the dominant trouble due to the smaller elasticity of substitution. A counteraction to restore the age-balance of the high-skilled labor is to increase the educational rate for older ages. On the contrary, if $\lambda_H \gg \lambda_L$, then it is more important to restore the age distribution of low-skilled labor, since their elasticity of substitution is smaller. The way to do this is to increase learning, more for young ages, which is to decrease the young low-skilled labor by educating it. In case of a decreasing population the arguments are exactly reverse.

Although the theoretical arguments that we presented above are supported by the numerical experiments, they do not provide a formal proof of the established dependence of the age-distribution of learning on the demographic factor and on the elasticities of substitution. However, we are able to obtain a rigorous result (proved in Appendix 2) in a somewhat simplified situation. Namely, we assume that $\delta(s) = e(s) = 0$, $\lambda_H = 1$, $\lambda_L < 1$, $p(s, u) = \frac{\varepsilon}{2}u^2$. Also, we assume here that the control constraint $u \geq 0$ is not binding, which, due to the specific form of $p(s, u)$, is equivalent to assuming that for each (t, s) the shadow price of a high-skilled worker is larger than the shadow price of a low-skilled one. Let $u^\gamma(t, s)$ be the optimal control. Since in this subsection we are interested in the long run behavior, we consider the optimal steady-state control $u^\gamma(s) = \lim_{t \rightarrow +\infty} u^\gamma(t, s)$. To investigate how it depends on γ we define

$$\Gamma(s) = \left. \frac{du^\gamma}{d\gamma}(s) \right|_{\gamma=0}.$$

Proposition 4 *The function Γ has the following form: there exists $\bar{s} < \omega$ such that $\Gamma(s) \geq 0$ on $[0, \bar{s}]$ and $\Gamma(s) < 0$ on (\bar{s}, ω) . It may happen that $\bar{s} < 0$, that is, $\Gamma(s) < 0$ for all $s \in (0, \omega)$.*

The interpretation is, that in an increasing population with non-perfect substitutability of non-skilled labor and perfect substitutability of skilled labor there will be more learning in young ages and less in old ages than for the stationary population. The higher learning in young ages, however, may be absent. Clearly, the result is completely consistent with those on Figure 3.

The other “extreme” case, $\lambda_L = 1$, $\lambda_H < 1$ is technically more complicated and we do not present its analysis. As Figure 2 (left plot) shows the effect of the demography on learning at older ages is just the opposite in this case, compared with the case $\lambda_L = 0.1$, $\lambda_H = 0.9$. However, for young ages the increasing population learns less in both cases.

The situation complicates even more if both the high-skilled and the low-skilled labor are not perfectly substitutable across ages. Here the relation between λ_L and λ_H plays a role for the qualitative dependence of the optimal learning rate on the demography in the same direction as in

the two “extreme” cases considered above. However, if both λ_L and λ_H are strictly smaller than one, and if the difference between them is “small”, then the particular data $\theta_L, \theta_H, \pi_L(s)$ and $\pi_H(s)$ may have a decisive role for the qualitative impact of the demography on the optimal educational policy.

5.2 Short run and anticipation effects of a demographic change

In this section we investigate (theoretically and numerically) the short run effects of a demographic change on the optimal learning rate. That is, if a stationary population begins to increase or decrease, how this demographic change would influence the optimal educational rate shortly after the change takes place. Moreover, we establish that in case of imperfect substitutability of labor (either across ages or across qualifications) a change in the demographic factor influences the optimal education policy not only afterwards but also before the change takes place. That is, contrary to the case of perfect substitutability of labor, here the expectation of a future change in the supply of labor influences the human capital building even before the labor market becomes affected by this change.

We start with the three scenarios from the previous section—constant, increasing, and decreasing population—by choosing a constant growth rate γ , where γ is zero, positive, or negative, respectively. We then assume that till the year $\bar{t} = 20$ the population is stationary in all the three scenarios, while starting at time $\bar{t} = 20$ the three populations grow differently: $N_0(t) = N_0 e^{\gamma(t-\bar{t})}$ for $t > \bar{t}$. We shall compare the optimal learning rate in the three scenarios at time $t \in [\bar{t}, \bar{t} + \omega)$ (that is, shortly after the growing/shrinking population enters the labor market), but also at time $t < \bar{t}$, when the labor market is still not affected by the demographic change, but is already aware of this future change in labor supply. In the numerical results presented below⁵ the constant size population is simulated numerically before time $t = 0$, and the end time T is sufficiently large (we have used $T = 180$) so that the optimal solution for the constant scenario is close to the steady state (does not change with time in the interval $(0, 30)$). We have chosen $\lambda_L = 0.9, \lambda_H = 0.1$, (alternatively $\lambda_L = 0.1, \lambda_H = 0.9$), $\gamma = 0$ or $\gamma = \pm 0.0072$ in the three scenarios, the rest of the data is as specified in Appendix 1.

Figure 4 presents the age distribution of the optimal education rate at $t = 18$ (two years before the demographic change starts to influence the labor market). The expectation of an increasing/decreasing population leads to a change in the optimal education rate before (in fact, also shortly after) the change at time $\bar{t} = 20$. The direction of change is different for different combinations of λ_L and λ_H . We see on the left plot of Figure 4 that if $\lambda_H = 0.1 < 0.9 = \lambda_L$, then the expectation of increasing population leads to higher education rate before the time of change,

⁵ The numerical solution is itself a challenging issue (this applies also to the previous sections). In our numerical analysis we use the general solver for age-structured optimal control problems developed by the third author, which is very briefly described in Feichtinger et al. 2004. A detailed description will be given in a forthcoming paper of the third author.

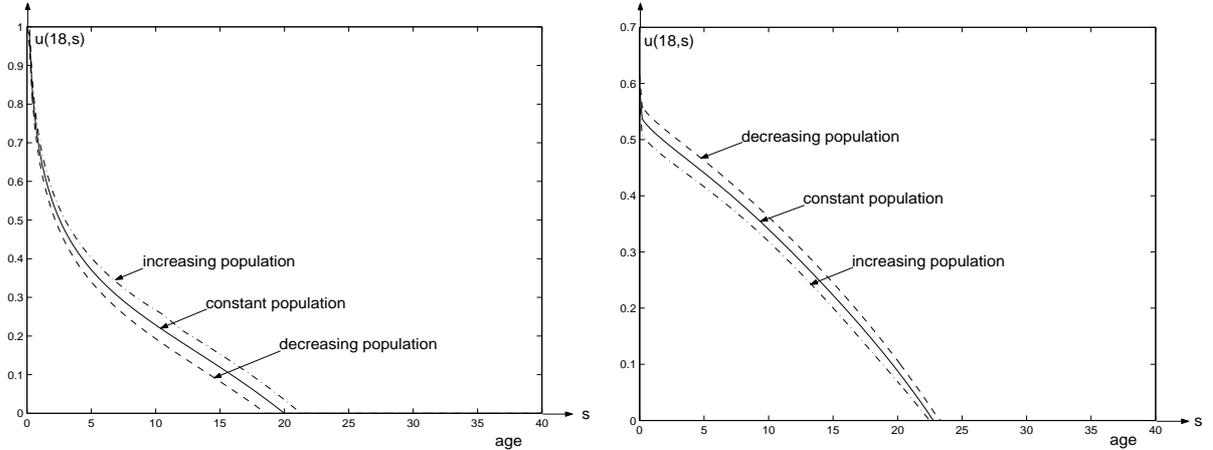


Figure 4: Age distribution of the optimal training rate in the anticipation phase ($t = 18$) for the three scenarios and for $\lambda_L = 0.9$, $\lambda_H = 0.1$ (left plot) and $\lambda_L = 0.1$, $\lambda_H = 0.9$ (right plot).

while for the decreasing population the education rate is lower. The situation is just the opposite for $\lambda_L = 0.1 < 0.9 = \lambda_H$ (the right plot).

The anticipation effect is not restricted just to the few years before \bar{t} . In Figure 5 we plot the time-path of the aggregate per capita education effort, defined as $\int_0^\omega L(t, s)u(t, s) ds / \int_0^\omega N(t, s) ds$. Since the population before \bar{t} is of the same size for all t , this is a relevant indicator for the educational effort. Clearly, the educational effort is highest for the population for which the labor market is expected to expand at \bar{t} and this happens in the whole plotted interval (at least 20 years before the demographic change results in a change of the labor supply).

The anticipation effect, discussed above, is rigorously established in the next proposition (proved in Appendix 2). To simplify somewhat the consideration we compare the following two scenarios: the constant inflow case, $N_0(t) = N_0$, and

$$N_0(t) = N_0^*(t) = \begin{cases} N_0 & \text{for } t \leq \bar{t}, \\ N_0^* & \text{for } t > \bar{t}, \end{cases}$$

where $N_0^* > N_0$ (the case $N_0^* < N_0$ can be treated in the same way). We shall assume that the upward jump in the population at (economically active) age $s = 0$ is affecting only the low-skilled workers: $N_0^*(t) = L_0^*(t) + H_0^*(t)$ where

$$L_0^*(t) = \begin{cases} L_0 & \text{for } t \leq \bar{t}, \\ L_0^* & \text{for } t > \bar{t}, \end{cases}$$

with $L_0^* > L_0$, and $H_0^*(t) \equiv H_0 > 0$ for $t \in [0, T]$. Although the scenario N_0^* is different from the scenario with exponentially increasing population $N_0(t) = N_0 e^{\gamma(t-\bar{t})}$, $\gamma > 0$, considered in the

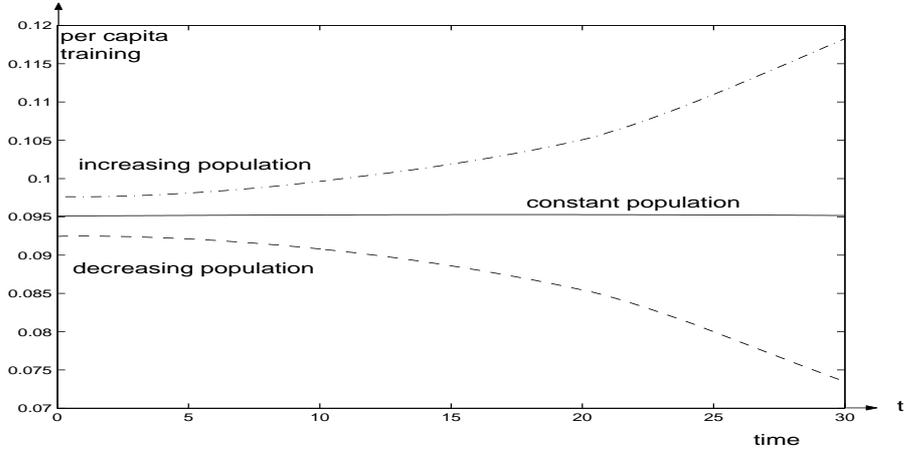


Figure 5: Per capita training effort in the anticipation phase ($t \in [0, 20]$) for the three scenarios, and for $\lambda_L = 0.9$, $\lambda_H = 0.1$.

numerical experiment, both scenarios have the same effect on the population change shortly after \bar{t} : in both cases the age distribution of the population shifts to lower ages in the time interval $[\bar{t}, \bar{t} + \omega)$. As we have already mentioned in Subsection 5.1, it is this shift that determines the policy change (before or after the demographic change) rather than the difference in the absolute size of the population, which plays no role.

Here we restrict the consideration to the case of imperfect substitutability between different qualifications ($\rho < 1$), while $\lambda_L = \lambda_H = 1$. The case of imperfect substitutability across ages is analytically more difficult and is enlightened only by the numerical results.

Denote by $u(t, s)$, $L(t, s)$, $H(t, s)$ the optimal path for the scenario with stationary demography, and by $u^*(t, s)$, $L^*(t, s)$, $H^*(t, s)$ – the optimal path for the $N_0^*(t)$ scenario for the demography.

Proposition 5 *Assume the following: (i) $\bar{t} > \omega$; (ii) there exists $s_0 \in (0, \omega)$ such that $u(t, s) > 0$ for $s \in [0, s_0]$ and $t < \bar{t}$. Then there exist a nonempty set $(t_1, t_2) \times (s_1, s_2)$ with $t_2 < \bar{t}$ such that $u^*(t, s) > u(t, s)$ on this set.*

We have thus established, theoretically and numerically, that a change of the learning rate takes place in the case of increasing/decreasing population even before the demographic change starts affecting the labor market. What is the economic reason for this effect?

A. Let us consider first the case of imperfect substitutability of high-skilled and low-skilled labor ($\rho < 1$), treated by Proposition 5. Since the additional labor that enters the labor market after the beginning of the demographic change at time \bar{t} is of low-skilled labor, the optimal balance between

high-skilled and low-skilled labor will be violated by an exceeding amount of low-skilled labor. Due to the imperfect substitutability across qualifications, in order to restore the optimal balance, the learning rate should be increased. Due to the increasing marginal cost of learning, the cost per unit of learning rate would become higher than before the demographic change, if the learning rate for $t < \bar{t}$ were not increased. On the other hand, the young high-skilled labor at time t before and close to \bar{t} would remain high-skilled labor also after \bar{t} . Then at the optimum a part of the additional learning rate after the demographic change would be shifted to the years before the change, in order to take advantage of the lower learning cost. This is the positive anticipation effect.

B. In case of imperfect substitutability of labor across ages ($\lambda_L < 1$, $\lambda_H < 1$, $\rho = 1$) the same intuitive reasoning applies as we discussed in Subsection 5.1 on the long run behavior of the optimal learning rate. Then the anticipation effect (Figure 4 represents actually the anticipation phase) follows from the same reasoning as in part **A**: shifting some of the educational efforts to times before \bar{t} (smoothing in this way the learning rate) decreases the costs of education.

In the case when λ_H and λ_L are both less than one and relatively close to each other, the same comment applies as in the end of Subsection 5.1.

The above analysis clearly exhibits the importance of the relation between the elasticity of substitution across ages of high-skilled and of low-skilled labor, which may lead to different (opposite) effects of the demographic changes on the optimal educational policy.

6 Changing returns to education and its relation to demographic change

In this section we define unskilled workers as those with at most high school and skilled as those with at least a college.

Assume now that the labor force is efficiently utilized. This means that the wages equal the marginal productivities. Then, after the optimization problem is solved, we can calculate the resulting age s and time t specific wages $w_L(t, s)$ and $w_H(t, s)$ from the respective marginal productivities.

Assuming that wages equal marginal productivity of labor and following the approach as in Card and Lemieux 2001 the age and time specific differential in the logarithmic wage of high- and low-skilled workers can be expressed as follows:

$$\begin{aligned} r(s, t) = \log w_H - \log w_L = & \log \left(\frac{\theta_H}{\theta_L} \right) + \log \left(\frac{\pi_H}{\pi_L} \right) + \frac{1}{\sigma_Y} (\log \tilde{L}(t) - \log \tilde{H}(t)) \\ & + \frac{1}{\sigma_H} (\log \tilde{H}(t) - \log H(t, s)) - \frac{1}{\sigma_L} (\log \tilde{L}(t) - \log L(t, s)), \end{aligned}$$

where $\sigma_L = 1/(1 - \lambda_L)$, $\sigma_H = 1/(1 - \lambda_H)$ and $\sigma_Y = 1/(1 - \rho)$ are the elasticities of substitution across ages of low-skilled and high-skilled labor, and across qualifications, respectively. Note that

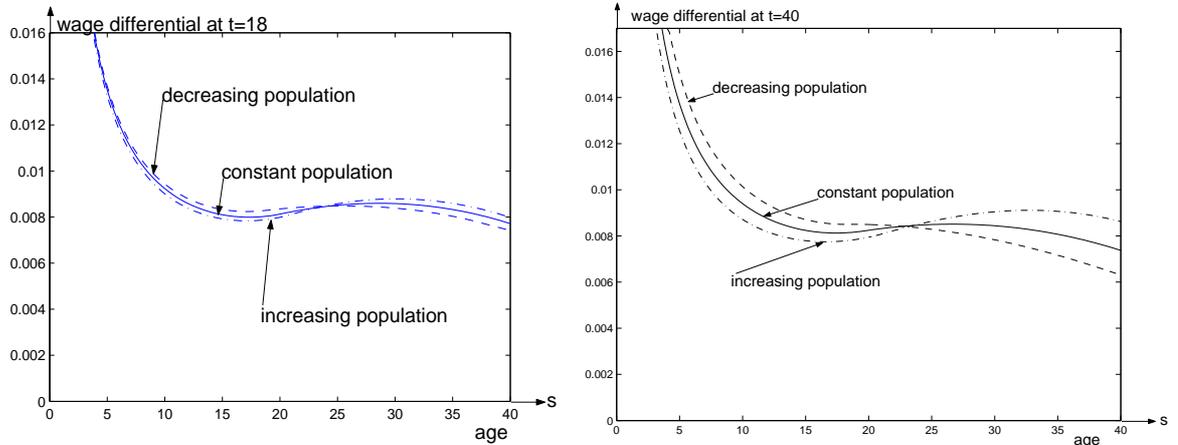


Figure 6: Productivity (wage) differentials for increasing, stationary and decreasing populations and for $\lambda_L = 0.9$, $\lambda_H = 0.1$: in the anticipation/transition phase $t = 18$ (left plot), and in the steady-state phase, $t = 40$ (right plot).

the case of perfect substitutability of workers of different ages within each educational category correspond to taking the limit $\sigma_L \rightarrow \infty$ and $\sigma_H \rightarrow \infty$. Then the wage differential will only depend on the aggregate composition of high and low-skilled workers, \tilde{L} and \tilde{H} , and not on the age-group specific supply of high and low skilled workers. In addition, if the elasticity of substitution across skill groups is also infinite, then the wage differential depends only on θ_H/θ_L and π_H/π_L , which complies with Proposition 2 in Section 4.

The two plots in Figure 6 represent the productivity (wage) differential, $w_H(t, s) - w_L(t, s)$, for increasing/stationary/decreasing populations with imperfect substitutability of labor across ages: $\lambda_L = 0.9$, $\lambda_H = 0.1$. It is remarkable that the dependence of the productivity (wage) differential on the demography in the anticipation/transition phase ($t = 18$) and in the steady-state ($t = 40$) is qualitatively the same. Namely, in both phases we observe *higher wage differentials in young ages for a decreasing population, and lower ones for an increasing population*.

As we stressed several times above, not the absolute size of the population, rather its age-distribution is responsible for the influence of the demography on the composition of labor across age and qualification. On the other hand, an increasing population has a younger age-distribution (lower average age) than a stationary or decreasing one, and the opposite is true for a decreasing population. Thus we may reformulate the above finding in the following way:

If the average age of labor is lower, the wage differential in young ages is lower, and vice versa.

Our normative model therefore provides an explanation of the evolution of the wage differentials for young male workers in the USA in the period 1960–1990 similar to the positive economic model set up as presented in Card and Lemieux 2001. The left plot on Figure 7 represents a smoothed version

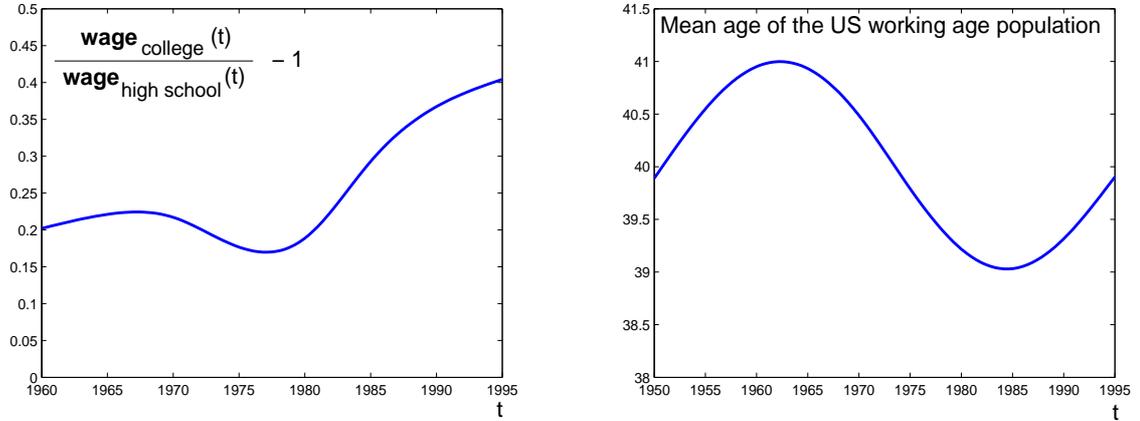


Figure 7: The (normalized) wage differentials for young male workers in the USA (left plot) and the mean age of working age (20-65 years) population (right plot).

of the data for the wage differentials for young male workers published in the above mentioned paper. The right plot represents the average age of labor (20–65 years old population). The latter figure is a smoothing of data extracted from United Nations 2004. Both plots have three intervals of monotonicity. Although there are small shifts in the times of minimal and maximal wage differentials versus average age of labor in the two figures, we point out that:

- (i) in the years before 1963 both the average age of labor and the wage differential are increasing;
- (ii) in the years 1967–1978 both the average age of labor and the wage differential are decreasing (the former, due to the baby-boom in USA some 20 years earlier);
- (iii) in the years 1984–1995 both the average age of labor and the wage differential are increasing (the former, due to the ageing of the baby-boom cohort).

Card and Lemieux (2001) have established also that there is no substantial change of the wage differential for old labor in the considered time period.

The slight differences in the qualitative behavior (increase/decrease) in the empirical data seen in Figure 7 and in our predictions may have many reasons: (i) the average age used in Figure 7 is only a rough indicator for the age structure of the population, on which the results of our model depend; (ii) additional factors such as migration, differences in the optimal solution for the central planner from the decisions of the economic agents in a market driven economy, etc.

For the qualitative replication of the evolution of the wage differential presented above it was crucial to assume that the elasticity of substitution across ages of the unskilled labor is higher than that for skilled labor, $\lambda_L > \lambda_H$. According to Stapleton and Young 1988 there is an empirical evidence for this assumption.

7 Conclusions

Skill and age heterogeneity have only rarely been integrated into formal macro-economic models. However, such models constitute the framework for a normative analysis of the optimal age- and education-specific labor force. The aim of our paper is to provide a formal model of the optimal education policy at the macro level allowing for heterogeneity of the workforce with respect to its age and qualification skills. Within this framework we study the optimal education rate in the context of changes in the labor demand (as represented by the elasticity of substitution across ages and qualification) and labor supply (as represented by the population growth rates).

We establish a number of numerical and analytical results on the optimal age-specific training rate. In case of perfect substitutability of labor across age and qualification we show that the optimal age-specific training rate is independent of labor supply. Once we allow for imperfect substitutability of workers across age and qualification the optimal training rate depends on labor demand as well as labor supply factors. The analysis clearly exhibits the importance of the relation between the elasticities of substitution across ages of skilled and unskilled labor, which may lead to different (opposite) impacts of the demographic change on the optimal educational policy. As these results indicate, the relation between the elasticities of substitution of labor across ages plays a crucial role for the way the demographic changes affect (both in the short and in the long run) the optimal educational policy. A further interesting result we obtained is the anticipation of future changes in labor supply. Already several years in advance of the time when the actual change takes place the optimal educational rate will change.

Various extensions of our analysis are promising. Our production technology depends only on labor. An obvious extension of our framework is to allow for physical capital in addition to human capital. So far we assumed a relatively simple educational process only implicitly taking into account the fact that people in education might not be active full time in the labor market. These assumptions as well as the assumption of full employment could be relaxed in further extensions of our model. So far, we assume labor supply—as represented by the demographic change—to be exogenous, but human capital composition will affect fertility and mortality. The interdependence of education, fertility and human capital (similar as in Connelly and Gottschalk 1995) could be studied. The assumption of the CES production function within each educational aggregate is restrictive as well. When workers from one age group are substituted by members of any other age group, the actual age difference does not matter. As it was recently indicated in Prskawetz et al. 2008 the production function could be extended to allow for a more flexible pattern of the substitutability of workers across ages.

Appendix 1: Data specifications

In the numerical experiments we have used the following data setting:

$\omega = 40$ – age of retirement minus initial working age $a^0 = 20$;
 $T = 140$ – end of the time interval;
 $[45, 95]$ – the time interval used for the plots;
 $\delta(t, s) \equiv 0$ – there is no decay of the human capital;
 $e(s) \equiv 0$ – there is no learning by doing;
 $l(t, s) \equiv 0.1386$ – constant in age and time efficiency of learning;
 $\theta_L(t) \equiv 0.3, \theta_H(t) \equiv 0.7$;
 $p(s, u) \equiv p(u) = \frac{9}{520}u + \frac{3}{260}u^2$ – the cost of educating one low-skilled worker if the educational effort is at level u ;
 $\rho = 1$ – the aggregate low- and aggregate high-skilled labor are perfect substitutes;
 λ_L – the different values are specified in the respective sections and figure captions;
 λ_H – the different values are specified in the respective sections and figure captions;
 $r = 0.03$ – discount rate;
 $\bar{\pi}_L(s) = c_L \exp(\frac{q_1^2}{(s-m_L)^2 - q_2^2}), \bar{\pi}_H(s) = c_H \exp(\frac{q_1^2}{(s-m_H)^2 - q_2^2})$ are age-specific productivities of low-skilled and high-skilled workers (the functional form is taken from Prskawetz and Veliov 2007, where the parameters are identified using data from France, 1998);
 $\pi_L(s) = \frac{\bar{\pi}_L(s)}{\int_0^\omega \bar{\pi}_L(\sigma) d\sigma}, \pi_H(s) = \frac{\bar{\pi}_H(s)}{\int_0^\omega \bar{\pi}_H(\sigma) d\sigma}$ are age-specific relative efficiency parameters in the CES functions for the low- and high-skilled aggregate labor;
 $c_L = 500, c_H = 1000$ – scaling factors,
 $m_L = 13, m_H = 20$ – age of maximal productivity,
 $q_1 = 100, q_2 = 60$ – parameters identified from data;
 $L_0(t) \equiv 1000, H_0(t) \equiv 10^{-6}$ – the inflow of low- and high-skilled workers at the initial working age in the constant population scenario;
 $L_0(t) = 1000 \exp(0.0072(t - 45)), H_0(t) = 10^{-6} \exp(0.0072(t - 45))$ – the inflow of low- and high-skilled workers at the initial working age in the increasing population scenario;
 $L_0(t) = 1000 \exp(-0.0072(t - 45)), H_0(t) = 10^{-6} \exp(-0.0072(t - 45))$ – the inflow of low- and high-skilled workers at the initial working age in the decreasing population scenario.

Appendix 2: The maximum principle

In this appendix we explain how the optimality condition described in Section 3 are obtained using the result in Feichtinger et al. 2003. This paper provides a Pontryagin type conditions for a class of problems of the form

$$\text{minimize } \int_0^T \int_0^\omega F(t, s, y(t, s), u(t, s), q(t)) ds dt, \quad (23)$$

subject to the equations

$$y_t(t, s) + y_a(t, s) = g(t, s, y(t, s), u(t, s)), \quad y(0, s) = y^0(s), \quad y(t, 0) = \varphi(t, q(t)), \quad (24)$$

$$q(t) = \int_0^\omega h(s, y(t, s), u(t, s)) da, \quad (25)$$

and the control constraint

$$u(t, s) \in U. \quad (26)$$

Here t is the time, running in a given interval $[0, T]$, $s \in [0, \omega]$ is a scalar variable interpreted as age, $y(t, s) = (y_1(t, s), \dots, y_m(t, s)) \in \mathbf{R}^m$, and $q(t) = (q_1(t), \dots, q_r(t)) \in \mathbf{R}^r$ are the states of the system, $u(t, s) \in U$ is the control, U is a subset of \mathbf{R}^m , F, g, h, y^0, φ are given functions.

We mention that models like this often arise in the population dynamics (like in the present paper), but also in other areas of economics, where s has the meaning of age of physical capital or technology (cf. Barucci and Gozzi 2001, Feichtinger et al. 2005, 2006).

In our model $y = (L, H)$, but in order to pass to a model in the above form we need a slight modification of equations (4), (5). Namely, we set $q = (\bar{L}, \bar{H}, P)$, where \bar{L} and \bar{H} are new variables (replacing \tilde{L} and \tilde{H}) defined by

$$\bar{L}(t) = \int_0^\omega \pi_L(s)(L(t, s))^{\lambda_L} ds, \quad (27)$$

$$\bar{H}(t) = \int_0^\omega \pi_H(s)(H(t, s))^{\lambda_H} ds. \quad (28)$$

We have to modify correspondingly the objective function (1):

$$\text{maximize } \int_0^T e^{-rt} \left[\left(\theta_L(t)(\bar{L}(t))^{\rho/\lambda_L} + \theta_H(t)(\bar{H}(t))^{\rho/\lambda_H} \right)^{1/\rho} - P(t) \right] dt. \quad (29)$$

Obviously problem (29), (2), (3), (27), (28), (6), (7) is equivalent to our original problem (1)–(7) and in the same time is in the form (23)–(26). Then Theorem 1 in Feichtinger et al. 2003 readily gives the following: the optimal control $u(t, s)$ maximizes the function

$$(-\xi(t, s) + \eta(t, s))l(t, s)L(t, s)u - p(s, u)L(t, s), \quad (30)$$

where the adjoint variables ξ and η corresponding to L and H , respectively, satisfy the equations

$$\begin{aligned} \xi_t + \xi_s &= r\xi + (\xi - \eta)(e(s) + l(t, s)u(t, s)) - \\ &\quad (Y(t))^{1-\rho}\theta_L(t)\pi_L(s)(\tilde{L}(t))^{\rho-\lambda_L}(L(t, s))^{\lambda_L-1} + p(s, u(t, s)), \\ \xi(t, \omega) &= 0, \quad \xi(T, s) = 0, \\ \eta_t + \eta_s &= r\eta + (-\xi + \eta)\delta(t, s) - \\ &\quad (Y(t))^{1-\rho}\theta_H(t)\pi_H(s)(\tilde{H}(t))^{\rho-\lambda_H}(H(t, s))^{\lambda_H-1}, \\ \eta(t, \omega) &= 0, \quad \eta(T, s) = 0. \end{aligned}$$

Setting $\Delta(t, s) = \eta(t, s) - \xi(t, s)$ and subtracting the first equation from the second we obtain equation (8). Having in mind that $L(t, s)$ is positive we obtain the claim of Proposition 1 from (30).

In the analysis of the optimal solution we implicitly assume also that the solution to our optimal control problem is unique, and that the necessary optimality condition (maximum principle) presented in Section 3 is also a sufficient condition. This assumption is automatically fulfilled if the mapping “control $u \rightarrow$ objective value” is strongly convex, for which there is a strong evidence. However, this purely mathematical issues go beyond the scope of the present paper.

Appendix 3: Proofs

1. Long run effects

We continue with the mathematical analysis from Subsection 5.1. We consider the adjoint equation (8) corresponding to the problem (16)–(22), in which we make the substitution $\Delta^\gamma(t, s) = e^{\gamma s} \tilde{\Delta}(t, s)$. Then the equation takes the form

$$\Delta_t^\gamma + \Delta_s^\gamma = (r + e(s) + \delta(s) + l(s)u(t, s))\Delta^\gamma - p(s, u(t, s)) - f^\gamma(t, s), \quad (31)$$

where

$$u(t, s) = p_{u+}^{-1}(s, l(s)\Delta^\gamma(t, s)), \quad (32)$$

and

$$f^\gamma(t, s) = \theta_H \pi_H(s) \left(\frac{e^{\gamma s} \tilde{H}^\gamma(t)}{H^\gamma(t, s)} \right)^{1-\lambda_H} - \theta_L \pi_L(s) \left(\frac{e^{\gamma s} \tilde{L}^\gamma(t)}{L^\gamma(t, s)} \right)^{1-\lambda_L}.$$

Using (19) and (20), the last expression can be rewritten as

$$\begin{aligned} f^\gamma(t, s) = & \theta_H \pi_H(s) \left(\int_0^\omega e^{\lambda_H \gamma (s-\sigma)} \pi_H(\sigma) \left(\frac{H^\gamma(t, \sigma)}{H^\gamma(t, s)} \right)^{\lambda_H} d\sigma \right)^{\frac{1-\lambda_H}{\lambda_H}} \\ & - \theta_L \pi_L(s) \left(\int_0^\omega e^{\lambda_L \gamma (s-\sigma)} \pi_L(\sigma) \left(\frac{L^\gamma(t, \sigma)}{L^\gamma(t, s)} \right)^{\lambda_L} d\sigma \right)^{\frac{1-\lambda_L}{\lambda_L}}. \end{aligned}$$

Since in this section we are interested in the long run behavior, we pass to the optimal steady-state system, which is obtained by assuming all functions in equations (17)–(32) independent of t . We keep the same notations L^γ , H^γ , etc., in which the argument t is skipped. To show in a more transparent way how one could make use of the above formula for the productivity differential in the analysis of the long run behavior of the optimal educational policy, we take first the simpler specifications⁶: Then we have

$$L^\gamma(s) = e^{-\int_0^s l(\theta)u^\gamma(\theta) d\theta} L_0, \quad u^\gamma(s) = p_{u+}^{-1}(s, l(s)\Delta^\gamma(s)) = \frac{l(s)}{c} \Delta^\gamma(s),$$

⁶We note here that in the illustrative calculations we have used data as specified in Appendix 1, especially that $p(s, u) \equiv p(u)$ has a linear as well as a quadratic term in u .

and after obvious transformations the formula for $f^\gamma(s)$ takes the form

$$f^\gamma(s) = \theta_H \pi_H(s) - \theta_L \pi_L(s) \left(\int_0^\omega \pi_L(\sigma) e^{\lambda_L \int_\sigma^s (\gamma + l(\theta) u^\gamma(\theta)) d\theta} d\sigma \right)^{\frac{1-\lambda_L}{\lambda_L}}.$$

Since $u^\gamma(s) = \frac{l(s)}{c} \Delta^\gamma(s)$, we obtain the adjoint equation in the feed-back form

$$\dot{\Delta}^\gamma(s) = r \Delta^\gamma(s) + \frac{(l(s))^2}{2c} (\Delta^\gamma(s))^2 - \bar{f}^\gamma(s, \Delta^\gamma(\cdot)), \quad (33)$$

where

$$\bar{f}^\gamma(s, \Delta^\gamma(\cdot)) = \theta_H \pi_H(s) - \theta_L \pi_L(s) \left(\int_0^\omega \pi_L(\sigma) e^{\lambda_L \int_\sigma^s \left(\gamma + \frac{(l(\theta))^2}{c} \Delta^\gamma(\theta) \right) d\theta} d\sigma \right)^{\frac{1-\lambda_L}{\lambda_L}}. \quad (34)$$

Due to the form of the functional \bar{f}^γ , where Δ^γ appears integrated, equation (33) is still a complicated integral-differential equation which is difficult to analyze. However, we are only interested in the qualitative dependence of Δ^γ on γ , in comparison to the case $\gamma = 0$. Therefore we pass to its first order approximation (with respect to γ , which is presumably a “small” number). Namely, we denote the marginal change of the difference in the shadow price of high- and low- skilled workers w.r.t. γ as follows

$$\Gamma(s) = \frac{d\Delta^\gamma}{d\gamma}(s),$$

where the derivative is evaluated at $\gamma = 0$. Differentiating (33) with respect to γ we obtain the following linear integral-differential equation for Γ :

$$\Gamma'(s) = r\Gamma(s) + \frac{(l(s))^2}{c} \Delta^0(s) \Gamma(s) - \frac{d\bar{f}^\gamma(s, \Delta^\gamma(\cdot))}{d\gamma}, \quad \Gamma(\omega) = 0. \quad (35)$$

From (34) we easily obtain the following expression for the last term for $\gamma = 0$:

$$\frac{d\bar{f}^\gamma(s, \Delta^\gamma(\cdot))}{d\gamma} = -\zeta(s) \int_0^\omega \varphi(\sigma) \int_\sigma^s \left(1 + \frac{(l(\theta))^2}{c} \Gamma(\theta) \right) d\theta d\sigma,$$

where

$$\varphi(\sigma) = \pi_L(\sigma) \lambda_L e^{\lambda_L \int_\sigma^s \frac{(l(\theta))^2}{c} \Delta^0(\theta) d\theta},$$

and we skip the long expression for ζ , from which it is only important that $\zeta(s) > 0$. Splitting the integral $\int_0^\omega = \int_0^s + \int_s^\omega$ in the second last formula and changing the order of integration in the two integrals we obtain

$$\begin{aligned} \frac{d\bar{f}^\gamma(s, \Delta^\gamma(\cdot))}{d\gamma} = & -\zeta(s) \left[\int_0^s \left(1 + \frac{(l(\theta))^2}{c} \Gamma(\theta) \right) \left(\int_0^\theta \varphi(\sigma) d\sigma \right) d\theta \right. \\ & \left. - \int_s^\omega \left(1 + \frac{(l(\theta))^2}{c} \Gamma(\theta) \right) \left(\int_\theta^\omega \varphi(\sigma) d\sigma \right) d\theta \right]. \end{aligned}$$

Plugging the obtained expressions in (35) and using the notation

$$\psi(s, \theta) = \begin{cases} \int_0^\theta \varphi(\sigma) d\sigma & \text{if } \theta \leq s \\ -\int_\theta^\omega \varphi(\sigma) d\sigma & \text{if } \theta > s, \end{cases}$$

we obtain the following form of the equation for Γ :

$$\Gamma'(s) = r\Gamma(s) + \frac{(l(s))^2}{c} \Delta^0(s)\Gamma(s) + \zeta(s) \int_0^\omega \psi(s, \theta) \left(1 + \frac{(l(\theta))^2}{c} \Gamma(\theta)\right) d\theta, \quad \Gamma(\omega) = 0. \quad (36)$$

This is still a complicated integral-differential equation, but it allows us to prove the following result.

Proposition 6 *The solution Γ of equation (36) has the following form: there exists $\bar{s} < \omega$ such that $\Gamma(s) \geq 0$ on $[0, \bar{s}]$ and $\Gamma(s) < 0$ on (\bar{s}, ω) . It may happen that $\bar{s} < 0$, that is, $\Gamma(s) < 0$ for all $s \in (0, \omega)$.*

Proof. Let $s^* \in [0, \omega]$ be a point such that $\Gamma(s^*) \geq 0$. There exists a maximal interval $[s_1, s_2] \subset [0, \omega]$ containing s^* such that $\Gamma(s) \geq 0$ on $[s_1, s_2]$. We consider two cases.

(i) Assume that $0 < s_1 < s_2$. Since $\Gamma(\omega) = 0$, clearly we have $\Gamma(s_1) = \Gamma(s_2) = 0$. Then

$$\Gamma'(s_1) \geq 0, \quad \text{and} \quad \Gamma'(s_2) \leq 0.$$

We have $Q(\theta) = 1 + \frac{(l(\theta))^2}{c} \Gamma(\theta) > 0$ for $\theta \in [s_1, s_2]$. Moreover

$$0 \leq \Gamma'(s_1) = \zeta(s_1) \int_0^\omega \psi(s_1, \theta) Q(\theta) d\theta,$$

which implies

$$\int_0^\omega \psi(s_1, \theta) Q(\theta) d\theta \geq 0,$$

since $\zeta(s) > 0$. Then

$$\begin{aligned} \frac{\Gamma'(s_2)}{\zeta(s_2)} &= \int_0^\omega \psi(s_2, \theta) Q(\theta) d\theta \\ &= \int_0^{s_2} \int_0^\theta \varphi(\sigma) d\sigma Q(\theta) d\theta - \int_{s_2}^\omega \int_\theta^\omega \varphi(\sigma) d\sigma Q(\theta) d\theta \\ &= \left(\int_0^{s_1} + \int_{s_1}^{s_2} \right) \int_0^\theta \varphi(\sigma) d\sigma Q(\theta) d\theta - \left(\int_{s_1}^\omega - \int_{s_1}^{s_2} \right) \int_\theta^\omega \varphi(\sigma) d\sigma Q(\theta) d\theta \\ &= \int_0^\omega \psi(s_1, \theta) Q(\theta) d\theta + \int_{s_1}^{s_2} \int_0^\omega \varphi(\sigma) d\sigma Q(\theta) d\theta \\ &> \int_0^\omega \psi(s_1, \theta) Q(\theta) d\theta \geq 0, \end{aligned}$$

where we use that $Q(\theta) > 0$ on $[s_1, s_2]$ and $\int_0^\omega \varphi(\sigma) d\sigma > 0$. Thus $\Gamma'(s_2) > 0$. This is a contradiction, which proves that case (i) is impossible. This implies, in particular, that there is no maximal open interval $(s_1, s_2) \subset [0, \omega]$ in which $\Gamma(s) > 0$ with $s_1 > 0$.

(ii) $0 < s_1 = s_2 = s^*$. In this case $\Gamma(s^*) = 0$ and $\Gamma'(s^*) = 0$. Since $\Gamma(s)$ is close to zero in a neighborhood (s', s'') of s^* , we have $Q(s) > 0$ in (s', s'') . Due to the conclusion in the case (i) we have $\Gamma(s) \leq 0$ in (s', s'') . Then there exist points $s_1 < s^* < s_2$ in (s', s'') for which $\Gamma'(s_1) \geq 0$, $\Gamma'(s_2) \leq 0$. This leads to a contradiction exactly in the same way as in case (i).

Thus we obtain that $[s_1, s_2] = [0, s_2]$. To complete the proof of the claim we observe that the case $s_2 = \omega$ is not possible, since in this case $\Gamma'(\omega) > 0$ due to (36) and this implies $\Gamma(s) < 0$ for s close to ω . *Q.E.D.*

Proposition 4 obviously follows from the above one in view of the particular form of the cost function p .

2. Short run and anticipation effects

Here we formulate and prove a proposition for the anticipation effect discussed in Subsection 5.2. Let us recall that $u(t, s)$ is the optimal control for the scenario with stationary demography and $u^*(t, s)$ is the optimal control for the $N_0^*(t)$ scenario (introduced in Subsection 5.2) for the demography.

Proposition 7 *Assume the following: (i) $\bar{t} > \omega$; (ii) there exists $s_0 \in (0, \omega)$ such that $u(t, s) > 0$ for $s \in [0, s_0]$ and $t < \bar{t}$. Then there exist (t, s) with $t < \bar{t}$ such that $u^*(t, s) > u(t, s)$.*

At first glance the conclusion of the proposition may seem rather weak: it claims that the “positive” anticipation of the future growth of the working population may happen to be exhibited just at a single moment of time $t < \bar{t}$ and a single age s . The lemma below implies that this happens, in fact, in a “solid” set $[t', t''] \times [s', s''] \in (0, \bar{t}) \times [0, \omega]$ of times and ages. This completes the proof of Proposition 5. On the other hand, it is certainly not true that the anticipation is positive (that is, $u^*(t, s) > u(t, s)$) for all $t < \bar{t}$ and $s \in [0, \omega]$ since clearly it may happen that (i) $u^*(t, s) = u(t, s) = 0$ for old ages s .

Lemma 1 *The optimal training rates u and u^* are continuous functions. In addition both u and u^* are bounded functions.*

Proof of Lemma 1. For $\lambda_L = \lambda_H = 1$ equation (8) for the adjoint variable Δ takes the form

$$\begin{aligned} \Delta_t + \Delta_s &= (d(t, s) + l(t, s)u(t, s))\Delta - p(s, u(t, s)) - f(t, s), \\ \Delta(t, \omega) &= 0, \quad \Delta(T, s) = 0, \end{aligned} \tag{37}$$

where

$$u(t, s) = p_{u+}^{-1}(s, l(t, s)\Delta), \quad d(t, s) = r + e(s) + \delta(t, s) \quad (38)$$

and

$$f(t, s) = \theta_H(t)\pi_H(s) \left(\frac{Y(t)}{\tilde{H}(t)} \right)^{1-\rho} - \theta_L(t)\pi_L(s) \left(\frac{Y(t)}{\tilde{L}(t)} \right)^{1-\rho}$$

Due to the standing assumptions the aggregated states $Y(t)$, $\tilde{L}(t)$ and $\tilde{H}(t)$ are continuous, as it could be easily obtained from the absolute continuity of the solution of equations (2), (3) along the characteristic lines (although u might be discontinuous, a priori). The function $\Delta \rightarrow p_{u+}^{-1}(s, l(t, s)\Delta)$ is Lipschitz continuous, due to the uniform strong convexity of $p(s, \cdot)$. Then equation (37) with u substituted from (38) has a Lipschitz right-hand side with respect to Δ . Moreover, the side conditions for this equation are also continuous (identically zero). Then continuity of Δ follows from the classical result for continuous dependence of the solution of continuously parameterized family of ODEs on the parameter. The argument is, that (37) is such a family, parameterized by the part of the boundary where the side conditions are given. Then the continuity of u follows from (38). *Q.E.D.*

Proof of Proposition 7. We only sketch the proof leaving many details to the diligent reader. We assume that $u^*(t, s) \leq u(t, s)$ for every $t \leq \bar{t}$ and $s \in [0, \omega]$, which will bring us to a contradiction. Then from equations (2), (3) it easily follows that

$$L^*(t, s) \geq L(t, s), \quad H^*(t, s) \leq H(t, s) \quad \forall t \leq \bar{t}, \quad \forall s \in [0, \omega], \quad (39)$$

hence, in particular,

$$\tilde{L}^*(\bar{t}) \geq \tilde{L}(\bar{t}), \quad \tilde{H}^*(\bar{t}) \leq \tilde{H}(\bar{t}).$$

From the continuity of u and u^* one easily obtains that (due to the jump of $L^*(t, 0)$ at $t = \bar{t}$) for all sufficiently small $h \geq 0$

$$\tilde{L}^*(\bar{t} + h) \geq \tilde{L}(\bar{t} + h) + \alpha h, \quad \tilde{H}^*(\bar{t} + h) \leq \tilde{H}(\bar{t} + h) + o(h), \quad (40)$$

where α is a positive constant (depending on the jump $N_0^* - N_0$), and $o(h)/h \rightarrow 0$ when $h \rightarrow 0$.

Now we compare the solutions Δ and Δ^* of the adjoint equations (37) corresponding to the two scenarios (with (38) plugged into the respective equation). We have

$$\left(\frac{Y}{\tilde{L}} \right)^{1-\rho} = \left(\theta_L + \theta_H \left(\frac{\tilde{H}}{\tilde{L}} \right)^\rho \right)^{\frac{1-\rho}{\rho}}, \quad \left(\frac{Y}{\tilde{H}} \right)^{1-\rho} = \left(\theta_H + \theta_L \left(\frac{\tilde{L}}{\tilde{H}} \right)^\rho \right)^{\frac{1-\rho}{\rho}}.$$

From (40) we obtain

$$\frac{\tilde{H}(\bar{t} + h)}{\tilde{L}(\bar{t} + h)} \geq \frac{\tilde{H}^*(\bar{t} + h)}{\tilde{L}^*(\bar{t} + h)} + \alpha_1 h, \quad \frac{\tilde{L}(\bar{t} + h)}{\tilde{H}(\bar{t} + h)} \leq \frac{\tilde{L}^*(\bar{t} + h)}{\tilde{H}^*(\bar{t} + h)} - \alpha_1 h,$$

where $\alpha_1 > 0$ is an appropriate constant (depending on α) and h is sufficiently small. The above two formulas and the assumption that θ_L and θ_H are strictly positive imply that

$$\left(\frac{Y}{\tilde{L}}\right)^{1-\rho} \geq \left(\frac{Y^*}{\tilde{L}^*}\right)^{1-\rho} + \alpha_2 h, \quad \left(\frac{Y}{\tilde{H}}\right)^{1-\rho} \leq \left(\frac{Y^*}{\tilde{H}^*}\right)^{1-\rho} - \alpha_2 h,$$

where the arguments $\bar{t} + h$ are skipped and $\alpha_2 > 0$ is another constant as before. The above inequalities imply that

$$f^*(\bar{t} + h, s) \geq f(\bar{t} + h, s) + \alpha_3 h.$$

Using (39) we obtain in the same way that

$$f^*(t, s) \geq f(t, s) \quad \text{for } t < \bar{t}, \quad s \in [0, \omega]. \quad (41)$$

Then an elementary comparison (viability) argument implies (considering equation (37) along the characteristic lines starting at $(\bar{t} + h, \omega)$) that

$$\Delta^*(\bar{t}, \omega - h) > \Delta(\bar{t}, \omega - h) \quad (42)$$

for all sufficiently small $h > 0$. Having in mind (38) we obtain that if $u(\bar{t}, \omega - h) > 0$ then $u^*(\bar{t}, \omega - h) > u(\bar{t}, \omega - h)$ and the proposition is proved due to the continuity of u and u^* . In the alternative case we fix h so that (42) holds and consider $u(\bar{t} - \tau, \omega - h - \tau)$ for $\tau \in [0, \omega - h - s_0]$. Due to assumptions (i) and (ii) of the proposition there exists a last τ_0 such that for $\tau \leq \tau_0$ we have $u(\bar{t} - \tau, \omega - h - \tau) = 0$, which implies $u^*(\bar{t} - \tau, \omega - h - \tau) = 0$ according to our assumption at the beginning of the proof. Then using (41) we obtain by the same comparison (viability) argument as above that $\Delta^*(\bar{t} - \tau_0, \omega - h - \tau_0) > \Delta(\bar{t} - \tau_0, \omega - h - \tau_0)$. Due to (38) this implies $u(\bar{t} - \tau_0, \omega - h - \tau_0) = 0 < u^*(\bar{t} - \tau_0, \omega - h - \tau_0)$ and we come again to a contradiction with our initial assumption. This completes the proof. *Q.E.D.*

Acknowledgments

This research was partly financed by the Austrian Science Foundation (FWF) under grant No P18161-N13 and by the Austrian National Bank (ÖNB) under grant No 11621.

References

- [1] Barucci, E. and F. Gozzi (2001) Technology adoption and accumulation in a vintage-capital model. *Journal of Economics* 74, 1–38.
- [2] Boucekkine, R., D. de la Croix and O. Licandro (2002) Vintage human capital, demographic trends, and endogenous growth. *Journal of Economic Theory* 104, 340–375.

- [3] Boucekkine, R. and R. Ruiz-Tamarit (2008) Special functions for the study of economic dynamics: the case of the Lucas-Uzawa model. *Journal of Mathematical Economics*, 44, 1 33–54.
- [4] Brokate, M. (1985) Pontryagin’s principle for control problems in age-dependent population dynamics. *Journal of Mathematical Biology* 23, 75–101.
- [5] Card, D. and T. Lemieux (2001) Can falling supply explain the rising return to college for younger men? A cohort-based analysis. *The Quarterly Journal of Economics* May 2001, 705–746.
- [6] Chari, V.V., and H. Hopenhayn (1991). Vintage human capital, growth, and the diffusion of new technology. *Journal of Political Economy*, 99, 1142–1165.
- [7] Christiaans, T. (2003) Aging in a neoclassical theory of labor demand. *Discussion Paper No. 112-03, University of Siegen*.
- [8] Connelly, R. and P. Gottschalk (1995) The effect of cohort composition on human capital accumulation across generations. *Journal of Labor Economics* 13, 155–176.
- [9] Feichtinger, G., R.F. Hartl, P.M. Kort and V.M. Veliov (2006) Anticipation effects of technological progress on capital accumulation: a vintage capital approach. *Journal of Economic Theory* 126, 143–164.
- [10] Feichtinger, G., R.F. Hartl, P.M. Kort and V.M. Veliov (2005) Environmental policy, the Porter Hypothesis and the composition of capital. *Journal of Environmental Economics and Management* 50, 434–446.
- [11] Feichtinger, G., G. Tragler and V.M. Veliov (2003) Optimality conditions for age-structured control systems. *Journal of Mathematical Analysis and Applications* 288, 47–68.
- [12] Feichtinger, G., Ts. Tsachev and V.M. Veliov (2004) Maximum principle for age and duration structured systems: a tool for optimal prevention and treatment of HIV. *Mathematical Population Studies* 11, 3–28.
- [13] Freeman, R.B. (1979) The effect of demographic factors on age-earnings profiles. *The Journal of Human Resources* 14, 289–318.
- [14] Heijdra, Ben J. and Ward E. Romp (2009) Human capital formation and macroeconomic performance in an ageing small open economy. *Journal of Economic Dynamics and Control* 33, 725–744.
- [15] Katz, L.F. and D.H. Autor (1999) Changes in the wage structure and earnings inequality. In O. Ashenfelter and David Card (Eds.), *Handbook of Labor Economics*, vol. 3, pp. 1463–1555. Elsevier Science.
- [16] Katz, L.F. and K.M. Murphy (1992) Changes in relative wages, 1963–1987: Supply and Demand Factors. *Quarterly Journal of Economics* 107, 35–78.

- [17] Kremer, M. and J. Thomson (1998) Why isn't convergence instantaneous? Young workers, old workers, and gradual adjustment. *Journal of Economic Growth* 3, 5-28.
- [18] Lam, D. (1989) Population growth, age structure, age age-specific productivity. Does a uniform age distribution minimize lifetime wages? *Journal of Population Economics* 2, 189-210.
- [19] Lucas, R.E. (1988) On the mechanics of economic development. *Journal of Monetary Economics* 22, 3-42.
- [20] Murphy, K.M. and F. Welch (1992) The structure of wages. *The Quarterly Journal of Economics* 107, 285-326.
- [21] Pfeiffer, F. and K. Reuß (2007) Age-dependent skill formation and returns to education: Simulation based evidence. *IZA Discussion Paper No. 2882*.
- [22] Prskawetz, A., and V.M. Veliov (2007) Age-specific dynamic labor demand and human capital investment. *Journal of Economic Dynamics and Control* 31, 3741-3777.
- [23] Prskawetz, A., T. Fent and R. Guest (2008) Workforce Aging and Labor Productivity. The role of supply and demand for labor in the G7. In Alexia Prskawetz, ... Bloom and Wolfgang Lutz (eds.), *Population Aging, Human Capital Accumulation, and Productivity Growth, A supplement to Population and Development Review*, vol. 34, pp. 298-323. New York: Population Council.
- [24] Roger, M. and M. Wasmer (2009) Heterogeneity matters: labour productivity differentiated by age and skills. Paris School of Economics, Working paper No. 2009-51.
- [25] Rojas, J.A. (2005) Life-cycle earnings, cohort size effects and social security: a quantitative exploration. *Journal of Public Economics* 89, 465-485.
- [26] Stapleton, D.C. and D.J. Young (1988) Educational attainment and cohort size. *Journal of Labor Economics* 6, 330-361.
- [27] Veliov, V.M. (1997) On the time-discretization of control systems. *SIAM J. Control Optim.*, 35(5), 1470-1486.
- [28] Webb, G.F. (1985) *Theory of nonlinear age-dependent population dynamics*. New York: Marcel Dekker.
- [29] Weiss, Y. (1986) The determination of life cycle earnings: a survey. In O. Ashenfelter and R. Layard (eds.), *Handbook of Labour Economics*, vol. 1, pp. 603-640. Amsterdam: Elsevier.
- [30] Welch, F. (1979) Effects of cohort size on earnings: the baby boom babies' financial bust. *Journal of Political Economy* 87, 65-97.
- [31] United Nations (2004) *World population prospects: The 2004 revision*. New York: United Nations.