Université catholique de Louvain
Département des Sciences économiques

Thèse présentée en vue de lobtention du grade de
docteur en sciences économiques

Essays on economic dynamics under heterogeneity

Anton O. Belyakov

Composition du jury:

Julio Davila (promoteur)
Raouf Boucekkine
Carmen Camacho
Vladimir Veliov
Mathieu Parenti

Acknowledgements

First of all I would like to thank my supervisors Raouf Boucekkine and Julio Davila. I'm grateful to other members of my doctoral jury: Mathieu Parenti, Carmen Camacho, and Vladimir Veliov. It has been a pleasure and a privilege to collaborate with Vladimir Veliov, Josef Haunschmied, and Tsvetomir Tsachev. Results of this collaboration are presented in the first two chapters of this thesis.

I would like to thank Alain Venditti, David de la Croix, and all participants of Workshop Overlapping Generations Days, Marseille (May 14-16, 2012) for their helpful comments on the results that constitute the first chapter. This chapter has also greatly benefited from the discussions at doctoral workshops in UCL. For that I'm thankful to the participants and especially to the discussants of my talks Hélène Latzer and Carmen Camacho. I also want to express my gratitude for fruitful discussions to Michael Kuhn, Klaus Prettner, and to all my colleagues from the Institute of Mathematical Methods in Economics in the Vienna University of Technology.

I'm grateful to Shlomo Weber and Alexey Savvateev for recommendations they gave me for a Marie Curie Fellowship at the Center for Operations Research and Econometrics. For the fellowship I'm thankful to Jean Hindriks. My gratitude goes also to Claude d'Aspremont, Céline Poilly, Michel de Vroey, Yuri Yatsenko, Vincent Vannetelbosch, and all professors whom I learned from during my stay in UCL. I'm thankful to François Maniquet and all participants of lunch seminars at the CORE for their useful comments on the results presented in the third chapter.

My thoughts also go to my colleagues and friends, that have made the time spent at the CORE especially inspiring: Paolo, Salome, Maia, Rafael, Peter, Besik, Olivier, Jean-Sébastien, Joel, Denis, Deniz, Aysegul, Asel, Kenmei, Santanu, Hiroshi, Andrea, Michal, Gilles, Ying, Christian, Bastien, Thang, Alejandro, the two Tanjas, and many more.

This research was partly financed by Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF) under grant No MA07-002 and by the Austrian Science Foundation (FWF) under grants No I476-N13 and P25979-N25.

Contents

General Introduction

Heterogeneity plays an important role in modern economic modeling. First of all, it is worth to mention the models of overlapping generations (OLG), where age distribution of the population is not homogeneous. For example, age distribution of a country population can influence the pension system and may lead to adoption of an optimal age-selective migration policy by the state, see e.g. (Simon, Belyakov, & Feichtinger, 2012) or (Belyakov, Feichtinger, & Simon, 2011). Besides age distribution, heterogeneity can be present in preferences, technologies, or/and wealth of economic agents. Heterogeneity of space, age, or size distribution of a renewable product is considered in the models of optimal fishery or forestry, see e.g. (Belyakov, Davydov, & Veliov, 2013; Belyakov & Veliov, 2014) and references therein.

This thesis consists of three essays studding the influence of heterogeneity upon the economic dynamics. In the first model of R&D-based economic growth the agents have heterogeneous preferences over a growing continuum of consumption goods. In the second model a monopolist firm or an industry distributes investments in physical capital among a growing set of technologies the payoff of which is heterogeneous. In the third model of solving territorial disputes between countries the income from the land use of a country is not equally (heterogeneously) distributed among the citizens. Each model is presented in a separate chapter.

Chapter 1 develops a general equilibrium endogenous growth model involving heterogeneous consumption by an age-structured population with uncertain but limited life span and balanced life-time budget without bequests. The heterogeneity is introduced via weights which the individuals attribute in their utility function to consumption of different goods depending on the vintage of the good. The goods are produced by monopolistically competitive firms and the variety of available goods/technologies is determined endogenously through R&D investments. The general equilibrium is characterized by a system of functional equations and is analytically or numerically determined for several particular weight functions. It is shown that the investments by agents may be insufficient to sustain growth. Therefore additional investments are provided by a competitive financial sector. The resulting imbalance between agents' assets and total value of firms can grow unboundedly in the case of homogeneous consumption. The results exhibit the qualitative difference between the dynamics of the model with heterogeneous versus homogeneous consumption. In particular heterogeneous consumption (when old goods are discounted) reduces the additional investments by the financial sector so that the values of firms become balanced by the assets of agents in the long run. These results are obtained in collaboration with Josef Haunschmied and Vladimir Veliov, (Belyakov, Haunschmied, & Veliov, 2012).

Chapter 2 deals with optimal control of heterogeneous systems, that is, families of controlled ODEs parameterized by a parameter running over a domain called *domain of heterogeneity*. The main novelty here is that the domain of heterogeneity is endogenous: it may depend on the control and on the state of the system. This extension is crucial for several economic applications and turns out to rise interesting mathematical problems. A necessary optimality condition is derived, where one of the adjoint variables satisfies a differential inclusion (instead of equation) and the maximization of the Hamiltonian takes the form of "min-max". As a consequence, a Pontryagin-type maximum principle is obtained under certain regularity conditions for the optimal control. A formula for the derivative of the objective function with respect to the control from L_∞ is presented together with a sufficient condition for its existence. A stylized economic example is investigated analytically and numerically. These results are obtained in collaboration with Tsvetomir Tsachev and Vladimir Veliov, (Belyakov, Tsachev, & Veliov, 2011).

Chapter 3 studies a mechanism governing the territory exchange between groups of people such as countries or municipalities is proposed based on land trading with approval of both sides under particular voting rule (majority rule, veto rule, etc.). Voting rules can model different forms of government such as monarchy, oligarchy, and democracy. Under these forms of government the influence of hererogeneity of wealth distribution among citizens on the territory exchange result is studied. Conquest of the territory is considered as a special case of trading, when the buyer country pays not to the seller country but to the army which conquer territory for the buyer. Conditions at which countries prefer to trade land rather than fight for it are found using a game theoretic approach. This is an extension of the author's master thesis (Belyakov, 2007a, 2007b), the results of witch were presented in (Belyakov, 2009).

Chapter 1

Heterogeneous consumption in OLG model with horizontal innovations

1.1 Introduction

In this chapter we present and investigate a general equilibrium model comprised of the following components: firms producing consumption goods, R&D firms developing new goods, bank sector, and age-structured consumers. A main new feature of the model is, that the consumption is heterogeneous and the consumers' preference may depend on the vintage of a particular good, as well as on the current technological frontier. This, together with the involvement of the bank sector, allows to reveal interesting new results about the economic growth.

In the proposed model, agents consume a continuum of perishable goods produced by monopolistically competitive firms. Each firm possesses a permanent patent which the firm buys on perfectly competitive R&D market. The needed investments are provided by a competitive bank sector, and are secured by agents' savings and future firms profits. We consider any improvement in the quality of a good as the invention of a new good by some R&D firm that increases the variety of goods, thus implying only *horizontal innovations*. The variety of goods increases with a speed proportional to the labor employed by perfectly competitive R&D firms with allowance for knowledge spillover, like in Romer (1990).

The consumers, that is, the population in the considered closed economy, have a finite (but uncertain) life span and no bequest intention. This component of the model adapts the Cass and Yaari (1967) framework and differs from the perpetual youth models (e.g., Blanchard, 1985), and from more general (possibly realistic) demographic setup in (Faruqee, 2003), since finite life span facilitates rigorous mathematical proofs. A general equilibrium model with realistic survival probability is considered in (Boucekkine, de la Croix, & Licandro, 2002).

We consider differentiated goods as final *consumption goods*, like in Judd (1985), and disentangle the concepts of productivity growth and growth of the variety of goods. In other words, scientists increase the variety of products rather than the productivity of labor. Indeed, the invention of a new car design does not mean that the sells of cars would increase in total or that new car is produced with less labor. The result of the invention can be that people would just buy cars of new design instead of old one. Although old fashioned cars may still be attractive to some customers and expensive, the production of such cars would decrease. Hence, the total production may stay the same, while the utility of a consumer increases because of the increasing variety of designs, since the consumer can find the most suitable car for her among newly invented designs. We do not specify how productivity of labor (and therefore per capita consumption) depends on R&D activity. This contrasts idea-based growth models (e.g., Romer, 1990) where the invention of a new good (technology) immediately increases productivity of labor and hence per capita consumption. That results in the scale effect prediction when the long run growth rate depends on the population of the country (scale of the economy), which is strongly at odds with 20th-century empirical evidence, see (Jones, 1995a, 1995b). We can still have a scale effect for the growth rate of agents' utilities, because the instantaneous utility depends on the increasing variety of goods. But, per capita consumption may not depend on the variety of goods in our model. Thus, we avoid determining the real growth rate of per capita consumption, included in GDP per capita, and its scale effect predictions.

We study the growth of lifetime discounted utility of generations. All agents are born with zero assets and should be insured from dying indebted. Agents can borrow money from other coexisting generations like in d'Albis and Augeraud-Véron (2011) which is a continuous time generalization of discrete pure exchange OLG models (e.g., Samuelson, 1958; Gale, 1973). In our model we introduce production without physical capital as in Sorger (2011), so that agents can invest only in patents. But, in contrast to Sorger (2011), the agents are not the only investors in the model. The reason for that is the absence of infinitely living agents' cohorts (Blanchard, 1985) or households (Romer, 1990; Sorger, 2011) to own infinitely living firms. The wish of agents to consume all their bounded lifetime income would hamper their investments in firms and stop the growth. Aggregated assets of agents are proved to be bounded and could be even negative like in pure exchange OLG models (Samuelson, 1958; Gale, 1973) depending on the income pattern over the life of an agent. So in the case of endogenous growth there could be imbalance between aggregated assets and the total value of all patents in the economy. To resolve this issue we introduce competitive banking sector which plays a role of infinitely living institution investing in new firms. Banks give loans to the startup firms for purchasing the patents under the pledge of these very patents. Because of the difference between savings and patents' value the economy needs money (liquidity) provided by the banks. Thus, the banks create additional liquidity equal to the change of the value of the patents in the economy. Hence, there are two asset markets in the economy. The one is the market of firms' loans balanced with the values of their patents. The other is the market of agents' savings balanced with their life insurances, according to which in the case of sudden death of an agent her debt is repayed or her deposit is taken by the insurance company. The banks keep deposits of agents under the same interest rate as loans for the firms because of the no-arbitrage condition. But we do not require the balance between agents' savings and firms' loans. Moreover, it will be seen that such requirement is not needed in our model for determining a general equilibrium and could make the growth impossible.

A substantial novelty of the proposed model is that it is a hybrid of continuous time OLG model and growth model with a continuum of consumption goods. It happens that optimal consumption of an agent can be obtained by a two-stage procedure similar to the one in Dixit and Stiglitz (1977). The main result is that for the agents' savings alone may be not enough for growth. Additional money is needed, which is provided by banks. One can interpret this money as a negative bubble. It is proven that aggregated assets of agents are bounded while the total value of intellectual property in the economy with homogeneous consumption can grow unboundedly so the difference between agents' savings and the value of patents can also be unlimited. The heterogeneity of consumption (discounting old goods) brings qualitative difference in the model dynamics compared to the homogenous case. The heterogeneity can make the total value of patents be bounded with zero limit, thus reducing the imbalance between savings of agents and the value of patents. The growth of the lifetime discounted utility of agents' generations could be bounded in the heterogeneous case in contrast to the case of homogeneous consumption, where the growth of utility is unbounded. It means that in the long run agents preferring new goods do not appreciate further increase of the variety of goods.

In the case of homogeneous consumption we prove that the real interest rate tends to zero which is the *biological interest rate* (the population growth rate which is zero in the long run), see Samuelson (1958). While the general equilibrium with heterogeneous consumption can be inefficient and the real interest rate can become negative as it may happen in OLG models (e.g., Blanchard, 1985; Diamond, 1965). We observed that in both heterogeneous and homogeneous cases the variety of goods can grow unboundedly.

The chapter is organized as follows. Section 1.1.1 introduces population and labor dynamics. Section 1.1.2 solves the agent's problem of finding her optimal consumption and investment profiles. Section 1.1.3 considers the monopolistically competitive production of continuum of goods and finds the price of goods in equilibrium. Section 1.1.4 introduces competitive R&D sector that increases the variety of goods. Section 1.1.5 finds the equilibrium conditions in financial sector (zero profit of banks) and in R&D sector (zero profit of R&D firms). Section 1.1.6 describes the market clearing including the full employment condition. Section 1.1.7 presents final succinct format of the general equilibrium equations. Section 1.2 studies dynamics of aggregated variables and life time aggregated utility of an agent. Section 1.3 discusses the model considering another source of heterogeneity in section 1.3.2 with introduction of heterogeneous production of labor that also allows for discussion on the economic growth in section 1.3.3. Section 1.3.1 introduces the government and shows that its fiscal policy is efficient in the heterogeneous case in contrast to the homogeneous one. Section 1.4 concludes the chapter. The proofs are located in the Appendix as well as the numerical procedure is described in Appendix 1.8.

1.1.1 Population and labor

The population is time-varying and exogenous: $n(\tau,t)$ denotes the size at time t of the population cohort born at time τ. It is assumed that there is a maximal age ω, so that $n(t,\tau)=0$ for $t-\tau \geq \omega$, but $n(\tau,t)>0$ for $0 \leq t-\tau < \omega$. That is, there are always some people in every cohort until it reaches age ω.

For the population function $n(\cdot,\cdot)$ it is assumed that it is continuous in τ for any fixed t and is non-increasing and absolutely continuous in t for every fixed τ. Then the function

$$S(\tau,t) \equiv \frac{n(\tau,t)}{n(\tau,\tau)}$$

is the survival probability at time t of an individual of cohort τ. Survival probability defines mortality rate as

$$\mu(\tau,t) \equiv -\frac{\dot{S}(\tau,t)}{S(\tau,t)} = -\frac{\dot{n}(\tau,t)}{n(\tau,t)} \tag{1.1}$$

(the dot above a symbol denotes the derivative with respect to the time t). Thus, we have the expression

$$S(\tau,t) = e^{-\int_\tau^t \mu(\tau,\theta)\,\mathrm{d}\theta} \tag{1.2}$$

for the survival probability, thus

$$n(\tau,t) = n(\tau,\tau)\, e^{-\int_\tau^t \mu(\tau,\theta)\,\mathrm{d}\theta}. \tag{1.3}$$

The total population at t is obtained by the integration of n over all currently living cohorts

$$N(t) \equiv \int_{t-\omega}^{t} n(\tau,t)\,\mathrm{d}\tau.$$

We assume that each individual is endowed with $l(\tau,t)$ units of homogeneous labor per time, so that

$$L(t) \equiv \int_{t-\omega}^{t} l(\tau,t)\, n(\tau,t)\,\mathrm{d}\tau \tag{1.4}$$

is the total amount of available labor units in the economy at time t.

1.1.2 Consumption and savings

The evolution of the assets is determined by the consumption/saving decisions of the individuals. The assets are homogeneous, while the consumption goods are heterogeneous: products labeled by the numbers $q \in [0, Q(t)]$ are available at time t. These products are ordered according to the invention time of the respective technology q, so that $Q(t)$ is the newest product at time t – the one just created at t.

It is assumed that each individual has perfect foresight for the wage $w(t)$, the real return rate on asset $r(t)$, the available consumption goods $q \in [0, Q(t)]$, and their real prices $p(t,q)$, for all t within her life-horizon. Under particular normalization of wage and productivity of labor, the prices p will be proven to be constant over both t and q. To write everything in real terms we will use labor as *numeriare*, that is, we set the wage equal to unity, $w(t) \equiv 1$. The wage, normalized to 1, is assumed to be equal for all jobs, which is reasonable in a model where qualification is not taken into account.

A representative agent born at time τ chooses her consumption $c(\tau,t,q)$ of good $q \in [0,Q(t)]$ at time t so that her expected total discounted utility

$$u(\tau) \equiv \int_{\tau}^{\tau+\omega} e^{-\rho(t-\tau)} S(\tau,t) \int_{0}^{Q(t)} m(\tau,t,q,Q(t))\, c(\tau,t,q)^{\alpha}\, \mathrm{d}q\, \mathrm{d}t \tag{1.5}$$

is maximized subject to the dynamic budget constraint

$$\dot{a}(\tau,t) = l(\tau,t) + (r(t) + \mu(\tau,t))\, a(\tau,t) - \int_{0}^{Q(t)} p(t,q)\, c(\tau,t,q)\, \mathrm{d}q, \tag{1.6}$$

with the following boundary conditions, due to the absence of bequest,

$$a(\tau,\tau) = 0, \tag{1.7}$$

$$a(\tau,\tau+\omega) = 0, \tag{1.8}$$

where $a(\tau,t)$ is the real amount of assets that the agent has at time $t \in [\tau,\tau+\omega]$, $\alpha \in (0,1)$ is an elasticity parameter, $m(\tau,t,q,Q(t)) \geq 0$ is the weight function, with which an agent aggregates her utilities from consumption of different products. The weight function $m(\tau,t,q,Q(t))$ represents agent's preferences among available goods $q \in [0,Q(t)]$ and is assumed to be strictly positive at least on some subset of $[0,Q(t)]$ with positive measure. That will result in positive demand for new goods. The weight function $m(\tau,t,q,Q(t))$ makes the goods heterogeneous if it depends on q. Notice that it can depend on the current technological frontier Q and on the age $t-\tau$ of the agent. We will show how the shape of the function m qualitatively determines the outcome of general equilibrium.

We assume that individuals are insured against the risk of dying with positive assets by a fair life-insurance company in the spirit of Yaari (1965) that redistributes wealth of individuals who died to those who are still alive in the same cohort[1]. Therefore the real rate of return $r(t)$ is augmented by the mortality rate $\mu(\tau,t)$. Thus, functional (1.5) is an extension of that in Judd (1985). The main novelty is the function $m(\tau,t,q,Q(t))$ which makes goods heterogeneous[2]. Since the insurance should ensure that each cohort ultimately consumes all its assets we add the end-point condition (1.8). The absence of bequest implies the initial condition (1.7).

[1]The equal saving/debt redistribution of deceased agent only within her cohort satisfies the balance of fare insurance, because agents from the same cohort have the same assets and probability of death. This redistribution is possible due to the assumption that there are always some people in the cohort until it becomes ω years old ($n(\tau,t) > 0$ for $0 \leq t-\tau < \omega$).

[2]The functional form $m(\tau,t,q,Q(t))$ is given exogenously, thus it differs from the quality in *quality-adjusted Dixit-Stiglits consumption index* used in some models with *vertical innovations* (e.g., Dinopoulos & Thompson, 1998).

For any fixed τ problem (1.5)–(1.8) has an optimal control $c(\tau,\cdot,\cdot)$ which can be obtained by a two-stage procedure similar to the one in Dixit and Stiglitz (1977).

In the *inner* stage we define the total real expenditures for consumption at time t as

$$E(\tau,t) \equiv \int_0^{Q(t)} p(t,q)\, c(\tau,t,q)\, \mathrm{d}q. \tag{1.9}$$

Then we fix an arbitrary non-negative function $E(\cdot)$ and determine for fixed τ and t the optimal distribution $c(\tau,t,\cdot)$ that maximizes the inner integral in (1.5) subject to (1.9). The solution is

$$c(\tau,t,q) = \left(\frac{m(\tau,t,q,Q(t))}{p(t,q)}\right)^{\frac{1}{1-\alpha}} \frac{E(\tau,t)}{G(\tau,t)}, \tag{1.10}$$

where

$$G(\tau,t) \equiv \int_0^{Q(t)} \left(\frac{m(\tau,t,q,Q(t))}{p(t,q)}\right)^{\frac{1}{1-\alpha}} p(t,q)\, \mathrm{d}q. \tag{1.11}$$

The resulting value of the instantaneous utility (the inner integral in (1.5)) can be written as $(G(\tau,t))^{1-\alpha}(E(\tau,t))^{\alpha}$. Then, inserting expression (1.2) for the survival probability and solution (1.10) of the inner problem in (1.5), we obtain the following *outer* problem (for a fixed cohort τ)

$$u(\tau) = \int_\tau^{\tau+\omega} e^{-\rho(t-\tau)-\int_\tau^t \mu(\tau,\theta)\,\mathrm{d}\theta}\, (G(\tau,t))^{1-\alpha}\, (E(\tau,t))^{\alpha}\, \mathrm{d}t \to \max_{E(\tau,\cdot)}, \tag{1.12}$$

subject to

$$\dot{a}(\tau,t) = l(\tau,t) + (r(t)+\mu(\tau,t))\, a(\tau,t) - E(\tau,t), \quad E(\tau,t) \geq 0, \tag{1.13}$$

$$a(\tau,\tau) = 0, \quad a(\tau,\tau+\omega) = 0. \tag{1.14}$$

Using the Pontryagin maximum principle we obtain the following first-order conditions for problem (1.12)–(1.14):

$$\alpha \left(\frac{E(\tau,t)}{G(\tau,t)}\right)^{\alpha-1} - \lambda(\tau,t) = 0, \tag{1.15}$$

$$-\dot{\lambda}(\tau,t) = \lambda(\tau,t)\,(r(t)-\rho), \tag{1.16}$$

where $\lambda(\tau,t)$ is the adjoint variable representing the marginal utility per unit of income. Note that like in Sorger (2011) the adjoint equation (1.16) does not depend on the mortality rate μ because the agent has fair life insurance. Adjoint equation (1.16) has the solution

$$\lambda(\tau,t) = \lambda_0(\tau)\, e^{-\int_\tau^t (r(\eta)-\rho)\,\mathrm{d}\eta}, \tag{1.17}$$

which defines along with (1.15) the solution of the outer problem

$$E(\tau,t) = G(\tau,t)\left(e^{\int_\tau^t (r(\theta)-\rho)\,\mathrm{d}\theta}\,\frac{\alpha}{\lambda_0(\tau)}\right)^{\frac{1}{1-\alpha}}, \tag{1.18}$$

where $\lambda_0(\tau)$ is the initial value (at time $t=\tau$) of the adjoint variable that has to be adjusted to ensure the end-condition $a(\tau+\omega)=0$ in (1.14). In doing this it will be notationally convenient to define the discount factors

$$R_\mu(\tau,t) \equiv \exp\left(-\int_\tau^t (r(\theta)+\mu(\tau,\theta))\,\mathrm{d}\theta\right), \tag{1.19}$$

$$R_\rho(\tau,t) \equiv \exp\left(-\int_\tau^t \frac{\rho-r(\theta)}{1-\alpha}\,\mathrm{d}\theta\right). \tag{1.20}$$

Standard calculations using the Cauchy formula for the solution of (1.13) and notations (1.19)–(1.20) give the following expression for $\lambda_0(\tau)$, for which conditions (1.14) are satisfied

$$\left(\frac{\alpha}{\lambda_0(\tau)}\right)^{\frac{1}{1-\alpha}} = \frac{\int_\tau^{\tau+\omega} R_\mu(\tau,t)\,l(\tau,t)\,\mathrm{d}t}{\int_\tau^{\tau+\omega} R_\rho(\tau,t)\,R_\mu(\tau,t)\,G(\tau,t)\,\mathrm{d}t}. \tag{1.21}$$

Then expression (1.18) with the use of (1.20) and (2.3) takes the explicit form

$$E(\tau,t) = G(\tau,t)\frac{R_\rho(\tau,t)\int_\tau^{\tau+\omega} R_\mu(\tau,s)l(\tau,s)\,\mathrm{d}s}{\int_\tau^{\tau+\omega} R_\rho(\tau,s)R_\mu(\tau,s)G(\tau,s)\,\mathrm{d}s}. \tag{1.22}$$

Using (1.10) and (1.22), one can obtain the consumption of product q at time t by an agent born at τ:

$$c(\tau,t,q) = \left(\frac{m(\tau,t,q,Q(t))}{p(t,q)}\right)^{\frac{1}{1-\alpha}} \frac{R_\rho(\tau,t)\,h(\tau)}{\int_\tau^{\tau+\omega} R_\rho(\tau,s)R_\mu(\tau,s)G(\tau,s)\,\mathrm{d}s}. \tag{1.23}$$

It is proportional to the agent's *human wealth* defined as in Blanchard (1985)

$$h(\tau) \equiv \int_\tau^{\tau+\omega} R_\mu(\tau,s)l(\tau,s)\,\mathrm{d}s, \tag{1.24}$$

which is the present value of agent's income flow from labor. The propensity to consume the product q decreases with its relative price

$$\frac{p(t,q)}{m(\tau,t,q,Q(t))}\left(\int_{\tau}^{\tau+\omega} R_{\rho}(\tau,s)R_{\mu}(\tau,s)G(\tau,s)\,\mathrm{d}s\right)^{1-\alpha}$$

and (due to the multiplier $R_{\rho}(\tau,t)$ in (1.23)) with the age of the agent $t-\tau$, provided that $\rho > r(\cdot)$ in (1.20). There is no *nonhuman wealth* in the model because other factors of production (besides labor, e.g. physical capital available for agents to invest into) are absent. Thus, agents' assets only redistribute values among generations in time.

Note that the optimal consumption profile (1.23) is completely defined by the price $p(\cdot,\cdot)$, the real interest rate $r(\cdot)$, and the frontier of the product variety $Q(\cdot)$. This is the key difference from the models with infinitely living households (like in Sorger, 2011) or OLG models with bequests, where consumption also depends on some free variable ($\lambda_0(\tau)$ in Sorger, 2011) which is to be defined in equilibrium.

1.1.3 Production sector

Further we assume that each product $q \in [0,Q(t)]$ available at time t is produced by a single firm, that is, we implement the concept of monopolistic competition. Starting from its creation, the firm producing product q may exist forever.[3] Moreover, we assume, similarly as Sorger (2011), that the production of each good involves only labor, and that the production of one unit of any good requires one unit of labor.

Here we denote by $C(t,q)$ the production of good q and use the fact that at equilibrium it is equal to the aggregated consumption of good q:

$$C(t,q) \equiv \int_{t-\omega}^{t} c(\tau,t,q)\,n(\tau,t)\,\mathrm{d}\tau. \tag{1.25}$$

The firm holding permanently the patent for product q sets the price $p = p(t,q)$. Inserting (1.23) into (1.25) one can express the demanded quantity at price p as

$$C(t,q) = \frac{F(t,q)}{p^{\frac{1}{1-\alpha}}}, \tag{1.26}$$

[3] There is no scrap time presented in classical vintage models. Nevertheless the firm may stop production of its product, when demand for it becomes zero, because of the function m.

where the function

$$F(t,q) = \int_{t-\omega}^{t} \frac{(m(\tau,t,q,Q(t)))^{\frac{1}{1-\alpha}} R_\rho(\tau,t)\,h(\tau)}{\int_\tau^{\tau+\omega} R_\rho(\tau,s) R_\mu(\tau,s) G(\tau,s)\,\mathrm{d}s}\, n(\tau,t)\,\mathrm{d}\tau$$

does not depend on the price $p(t,q)$ of any particular good q due to the assumption of monopolistic competition. It follows from the definition of $G(\tau,t)$ in (1.11) that it depends on the prices of the goods in an integrated form, so that no single firm may influence $F(t,q)$. The q-th firm produces quantity $C(t,q)$, the production cost of which in real terms is $C(t,q)$, according to the assumption that the production of one unit of any good requires one unit of labor. Then the real operating profit of the q-th firm at time t is

$$\pi(t,q) = pC(t,q) - C(t,q). \tag{1.27}$$

Due to (1.26) the firm's profit takes the form

$$\pi(t,q) = (p-1)\, p^{\frac{-1}{1-\alpha}} F(t,q),$$

which attains its maximum with respect to the price p at

$$p(t,q) = \frac{1}{\alpha}. \tag{1.28}$$

Thus, all goods have the same price (1.28), although the consumption of these goods (1.23), (1.25) can be different due to different preferences. The reason for such simple (constant) prices is the normalization of wage and productivity of labor to unity, i.e. that any product q is measured in labor units spent for its production. For the optimal operating profit of the q-th firm at time t (with $q \in [0, Q(t)]$) we obtain from (1.27) and (1.28) the expression

$$\pi(t,q) = \frac{1-\alpha}{\alpha} C(t,q), \tag{1.29}$$

relating this profit with the production $C(t,q)$, which is equal to the aggregated consumption according to (1.25).

1.1.4 R&D sector

The R&D sector produces new technologies increasing in this way the variety $[0, Q(t)]$ of available consumer's goods. The R&D industry sells patents of infinite life for new productions. The R&D sector is a perfectly competitive industry which requires only labor and shares the labor market with the production sector, so that the wage at time t is $w(t) = 1$. Similarly as Jones (1995b); Kortum (1997); Segerstrom (1998) we assume that the dynamics of the technological frontier $Q(t)$ is given by the equation

$$\dot{Q}(t) = \beta \, (Q(t))^{\varphi} \, L_Q(t), \tag{1.30}$$

where $L_Q(t)$ is the total labor employed in R&D, $\beta > 0$ is a productivity parameter, φ is the parameter determining how productivity depends on $Q(t)$, that is on the already existing technologies. Thus, we allow for knowledge spillover if $\varphi > 0$ [4]. With $\varphi > 0$, formulation (1.30) implies increasing returns to scale in R&D, when previous inventions raise the productivity of current research effort. Alternately, with $\varphi < 0$, the formulation allows for diminishing returns in R&D, as if past inventions make it more difficult to find new ideas. For models with different values of the parameter φ see Jones (1999) and the references therein. For example, with $\varphi = 1$, in Romer (1990); G. M. Grossman and Helpman (1991a); Aghion and Howitt (1992) models, each unit of research effort can produce a proportionate increase in the technological variety.

Entrepreneurs buying new patents have perfect foresight. An entrepreneur knows future demand for the product and (1.26) thus, her future operational profit (1.29). If weight function m is strictly positive in the vicinity of $q = Q(t)$, then at least during some nonzero time interval, the operational profit flow from production of new goods is nonzero. Hence, entrepreneurs have incentives to buy new patents and found firms, because people will consume new goods.

1.1.5 Financial sector

The entrepreneur, willing to establish a firm for production of a new good, takes a loan from a bank at real interest rate $r(t)$ in order to buy a patent from the R&D sector. The bank cannot make higher interest rate for the firm than that on agents' deposits because in this case the firm can attract investments from other bank (because of the free entry condition in the bank sector) or directly from people offering them the same rate $r(t)$ of return as the bank does. Banks can "create money" taking patents as security, thus giving more loans than the total deposits of people. Actually, there is no restriction in the model on how much loans a bank can give since the firms for sure will repay

[4] Different extension of equation (1.30) are possible, including the dependence on a weight average of the distribution of labor across different technologies, which is reasonable (see e.g. Aghion & Howitt, 1996, 1998; Sorger, 2011). This would cause only some technical burden, therefore below we focus only on dynamics (1.30).

their debts in the future. These debts are balanced in the bank accounting by the values of patents that firms put in pledge.

Similar to Sorger (2011) we denote the present value of the real profit flow for the firm producing product $q \in [0, Q(t)]$ over interval $[t, +\infty)$ as

$$v(t,q) = \int_t^\infty \exp\left(-\int_t^s r(\theta)\,\mathrm{d}\theta\right)\pi(s,q)\,\mathrm{d}s, \tag{1.31}$$

which, on the other hand, is the debt that the q-th firm owes to the bank.

Since the market for patents is competitive we assume that the R&D industry as a whole makes no profit, and the firm who buys a patent makes zero net profit[5] too, because of the free entry assumption. This idea of *intertemporal zero net profit constraint* (as in Romer, 1990) is taken from G. M. Grossman and Helpman (1989). Thus, the market price $v(t, Q(t))$, of the patents produced at time t is determined by the zero-profit condition

$$v(t,Q(t))\,\dot{Q}(t) = L_Q(t), \tag{1.32}$$

which results from the fact that the R&D industry creates $\dot{Q}(t)$ patents in a unit of time at t and the cost is what is paid for the involved labor $L_Q(t)$.

The time-derivative of expression (1.31) for the price of a new patent $q = Q(t)$ yields the following no-arbitrage condition

$$\underbrace{\frac{\frac{\mathrm{d}}{\mathrm{d}t}v(t,Q(t))}{v(t,Q(t))} + \frac{\pi(t,Q(t))}{v(t,Q(t))}}_{\textit{rate of return on the ownership}} = r(t) + \underbrace{\frac{\dot{Q}(t)}{v(t,Q(t))}\int_t^\infty e^{-\int_t^s r(\theta)\,\mathrm{d}\theta}\,\frac{\partial\pi}{\partial q}(s,q)\bigg|_{q=Q(t)}\mathrm{d}s}_{\textit{rate of novelty premium}}. \tag{1.33}$$

It states that the rate of return on the ownership of a firm — consisting of the rate of capital gain $\frac{\frac{\mathrm{d}}{\mathrm{d}t}v(t,Q(t))}{v(t,Q(t))}$ and the profit rate $\frac{\pi(t,Q(t))}{v(t,Q(t))}$ — has to equal the real interest rate $r(t)$ plus the *rate of novelty premium* (as we call the last term). It differs from the no-arbitrage condition in the literature (e.g. Romer, 1990; G. M. Grossman & Helpman, 1991a, Chapter 3) by the rate of novelty premium that appears because of the heterogeneity of goods.

In special cases the improper integral in (1.33) can be expressed via $v(t, Q(t))$. Then, no-arbitrage condition (1.33) yields an explicit expression for the real interest rate $r(t)$. Such expression is more convenient for calculating the general equilibrium, than the original integral equation (1.31) in zero-profit condition (1.32).

[5] The *net profit* of the firm is its *operating profit* π minus taxes (that will be introduced later) and minus interest on debt.

1.1.6 Market clearing

We have already taken into account the clearing on product market (1.25) where production equals consumption. In view of the technological assumption in Section 1.1.3 we can express the total labor in production as

$$L_P(t) = \int_0^{Q(t)} C(t,q)\,\mathrm{d}q. \tag{1.34}$$

The labor market is cleared by the full employment condition

$$L_P(t) + L_Q(t) = L(t), \tag{1.35}$$

where the total labor $L(t)$ is given exogenously by (1.4).

Financial market is cleared by intertemporal zero-profit condition (1.32) for the firms, meaning that all invented technologies are purchased by the firms whose debts to the banks are balanced with the values of their patents. The banks have zero profit too, because of the absence of entry barriers for new banks. Thus, the real interest rate $r(t)$ on firms' debts is the same as that on agents' deposits.

1.1.7 Succinct format of the general equilibrium equations

With the use of (1.31) and (1.29) the zero-profit condition (1.32) in the R&D sector takes the form

$$L_Q(t) = \beta(Q(t))^{\varphi} L_Q(t) \int_t^{\infty} e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} \frac{1-\alpha}{\alpha} C(s,Q(t))\,\mathrm{d}s. \tag{1.36}$$

Due to (1.34) the labor balance equation (1.35) becomes

$$L_Q(t) = L(t) - \int_0^{Q(t)} C(t,q)\,\mathrm{d}q. \tag{1.37}$$

The dynamics of the variety frontier Q is as in (1.30)

$$\dot{Q}(t) = \beta\ (Q(t))^{\varphi}\ L_Q(t), \qquad Q(0) = Q^0 > 0. \tag{1.38}$$

The above three equations involve the total consumption $C(t,q)$ given by (1.25), which in view of (1.23) takes the form

$$C(t,q) = \alpha \int_{t-\omega}^{t} \frac{(m(\tau,t,q,Q(t)))^{\frac{1}{1-\alpha}}}{g(\tau)} R_\rho(\tau,t)\,h(\tau)\,n(\tau,t)\,\mathrm{d}\tau. \tag{1.39}$$

Here

$$g(\tau) \equiv \int_{\tau}^{\tau+\omega} R_\rho(\tau,s) R_\mu(\tau,s) M(\tau,s)\,\mathrm{d}s \tag{1.40}$$

and $M(\tau,s)$ is a substitution for $G(\tau,s) = \alpha^{\frac{\alpha}{1-\alpha}} M(\tau,s)$, so that

$$M(\tau,s) \equiv \int_0^{Q(s)} (m(\tau,s,q,Q(s)))^{\frac{1}{1-\alpha}}\,\mathrm{d}q. \tag{1.41}$$

The integral $M(\tau,s)$ represents how much the agent in cohort τ enjoys the variety of goods at time $s \geq \tau$, and $g(\tau)$ is the present discounted value of this propensity for a newborn at τ. So we have the three equations (1.36)–(1.38) for three unknown functions, Q, L_Q, and r which determine the general equilibrium. Further we use these equations for analytical and numerical investigation of the economy.

1.2 Analysis of general equilibrium

Notice that because of the assumption that for any time t there is a subset of goods in $[0,Q(t)]$ with nonzero measure, such that $m(\tau,t,q,Q(t)) > 0$, the integrals $M(\tau,t)$ and $g(\tau)$ are strictly positive for all τ. Thus, the integral in (1.37) of consumption (1.39) is also strictly positive since functions R_ρ, h and n are strictly positive in the integration domain. Hence, there are always workers in the production sector, meaning $L_Q(t) < L(t)$.

We also mention that expression (1.22) for the consumption expenditure can be written in terms of h, g, and M defined in (1.24), (1.40), and (1.41)

$$E(\tau,t) = R_\rho(\tau,t)\,h(\tau)\,\frac{M(\tau,t)}{g(\tau)}. \tag{1.42}$$

1.2.1 Aggregated state variables

For better understanding of the model dynamics we introduce aggregated state variables and derive corresponding aggregated equations. Integration of the profit expression (1.27) over all existing products $[0,Q(t)]$ yields

$$\Pi(t) = I(t) - L_P(t), \tag{1.43}$$

where we introduce the total profit Π and income I of the firms

$$\Pi(t) \equiv \int_0^{Q(t)} \pi(t,q)\,\mathrm{d}q, \tag{1.44}$$

$$I(t) \equiv \int_0^{Q(t)} p(t,q)\,C(t,q)\,\mathrm{d}q. \tag{1.45}$$

Thus, (1.43) reads as the total profit of the firms equals their total income minus the total labor expenses in production.

At equilibrium the total income of all firms in the economy is equal to the aggregated expenditure for consumption,

$$I(t) = \int_{t-\omega}^{t} E(\tau,t)\, n(\tau,t)\, \mathrm{d}\tau, \tag{1.46}$$

where the expenditure $E(\tau,t)$ of each cohort τ is defined in (1.42).

It follows from (1.34), (1.45), and (1.28) that the labor in production $L_P(t)$ has the following simple relation with the total firms' income $I(t)$

$$L_P(t) = \alpha I(t), \tag{1.47}$$

and defines the total profit of firms

$$\Pi(t) = \frac{1-\alpha}{\alpha} L_P(t). \tag{1.48}$$

Now let us aggregate the assets of individuals at time t,

$$A(t) \equiv \int_{t-\omega}^{t} a(\tau,t)\, n(\tau,t)\, \mathrm{d}\tau, \tag{1.49}$$

that are on deposits in the bank.

Recalling that newborns have no initial assets, see (1.7), we have the following formula for the derivative of $A(t)$ obtained by time-differentiation of (1.49) with the use of expression (1.2)

$$\dot{A}(t) = \int_{t-\omega}^{t} \dot{a}(\tau,t)\, n(\tau,t)\, \mathrm{d}\tau - \int_{t-\omega}^{t} a(\tau,t)\, \mu(\tau,t)\, n(\tau,t)\, \mathrm{d}\tau. \tag{1.50}$$

The aggregation of the budget constraints (1.13) over age cohorts gives

$$\int_{t-\omega}^{t} \dot{a}(\tau,t)\, n(\tau,t)\, \mathrm{d}\tau = \int_{t-\omega}^{t} [l(\tau,t) + (r(t) + \mu(\tau,t))\, a(\tau,t) - E(\tau,t)]\, n(\tau,t)\, \mathrm{d}\tau.$$

Then equations (1.4), (1.46), and (1.50) yield the following asset balance equation:

$$\dot{A}(t) = r(t)A(t) + L(t) - I(t). \tag{1.51}$$

Thus, the aggregated assets A rise at the amount of interest rA and what is left from the labor income L after the consumption expenses (equal to the total income of all firms I).

We also introduce the aggregated value of all patents in the economy,

$$V(t) \equiv \int_0^{Q(t)} v(t,q)\,\mathrm{d}q, \tag{1.52}$$

where $v(t,q)$ is given by (1.31). Since the value $V(t)$ is equal to the firms' debt to the bank we have the following differential equation for $V(t)$:

$$\dot{V}(t) = r(t)V(t) + L_Q(t) - \Pi(t), \tag{1.53}$$

meaning that the debt (besides interest rV) raises with the loan for research L_Q and falls with the current firms' profit Π payed to the banks. Equation (1.53) can also be obtained by differentiation of (1.52) with respect to time, taking into account expression (1.31), definition (1.43), and the R&D zero-profit condition (1.32).

Proposition 1 *The total value of patents $V(t)$ and the total assets of the individuals $A(t)$ are governed by the same differential equation*

$$\dot{A}(t) = r(t)A(t) + L(t) - \frac{1}{\alpha}L_P(t), \tag{1.54}$$

and are related as follows:

$$V(t) = A(t) + \psi e^{\int_0^t r(s)\,\mathrm{d}s},$$

where ψ is a constant.

Proof See Appendix 1.5.1.

Although the dynamics of the aggregated assets and the firms' debts is described by the same differential equation (1.54) they are not equal in general ($\psi \neq 0$), as it will be shown by the next proposition.[6]

It shows that the aggregated real assets, $A(t)$, are bounded due to the constant wages, and the no-bequest scenario of the model.

Proposition 2 *If the real interest rate r and total labor endowment L are bounded (for all t, $|r(t)| < \bar{r}$ and $L(t) \leq \bar{L}$), then the aggregated assets of the individuals are also bounded: $|A(t)| < \bar{A}$.*

[6] We do not require the aggregated assets of agents to be equal to the total debts of firms like it is done in (Sorger, 2011), because in our model the optimal consumption profile and investments of a finitely living agent are completely defined via initial condition (1.7) and terminal conditions (1.8), like in (Cass & Yaari, 1967; d'Albis & Augeraud-Véron, 2011), which is not the case for infinitely living households in (Sorger, 2011), where he needs additional condition $V(t) \equiv A(t)$ to specify general equilibrium.

Proof See Appendix 1.5.2. □

Corollary 1 *If the value of all patents in the economy V is bounded ($V(t) \leq \bar{V}$ for all t) and the real interest rate converges to a constant, $r(t) \to \hat{r}$ with $t \to \infty$, then $\hat{r} \leq 0$ or $V(t) = A(t)$ for all t.*

Proof See Appendix 1.5.3. □

Bounded growth

If we assume bounded growth of the variety frontier $Q(t) \to \bar{Q}$, so that due to (1.38) we have $L_Q(t) \to 0$, then asset markets are balanced ($A(t) \equiv V(t)$) and the long run interest rate is positive ($r(t) \to \hat{r} > 0$) as shown by the following proposition. Thus, the imbalance between asset markets ($A(t) < V(t)$ for all t) implies growth.

Proposition 3 *Let there exist a general equilibrium with bounded Q, with bounded real interest rate r and bounded total labor endowment L. Let also $r(t) \to \hat{r}$ and $L(t) \to \hat{L}$ with $t \to \infty$. Then the value of patents V is balanced by aggregated assets A for all t and has the following limit*

$$A(t) = V(t) \to \frac{1-\alpha}{\alpha}\frac{\hat{L}}{\hat{r}} > 0,$$

where limit of the real interest rate is positive $\hat{r} > 0$.

Proof See Appendix 1.5.5. □

Obviously, utility in (1.56) is also bounded in this case.

Notice that the corner solution ($L_Q(t) \equiv 0$) is the special case of the bounded growth $Q(t) \to \bar{Q}$, where $L_Q(t) = 0$ for all $t \geq \bar{t}$, i.e. there is no growth of $Q(t) = \bar{Q}$ since R&D is absent.

From now on we will consider only **unbounded growth** of Q. Moreover, we assume that there is always R&D activity, $L_Q(t) > 0$ for all t. Then, from (2.2) and (1.32) we obtain the price of a new patent

$$v(t, Q(t)) = \frac{1}{\beta\,(Q(t))^{\varphi}}. \tag{1.55}$$

The newest patent becomes cheaper as $Q \to \infty$ because of the knowledge spillover, $\varphi > 0$, while without spillovers, $\varphi = 0$, its price stays constant $v(t, Q(t)) = 1/\beta$.

1.2.2 Utility of an agent

We are to analyze the growth of the expected life-time aggregated utility (1.12) of a representative individual in cohort τ (see Appendix 1.6.1)

$$u(\tau) \quad = \quad (\alpha\, h(\tau))^{\alpha}\,(g(\tau))^{1-\alpha}, \tag{1.56}$$

where h and g are defined in (1.24) and (1.40). This expression shows that $u(\tau)$ depends positively on the human wealth of the agent and on the discounted integral M defined in (1.41) representing how much the agent values the variety of goods.

1.2.3 Benchmark case of homogeneous consumption with $\varphi = 1$

Let us consider the usually studied in the literature (e.g. Dixit & Stiglitz, 1977) the case when

$$m(\tau,t,q,Q) \equiv 1,$$

meaning that all existing products are equally appreciated by the consumer. We will show that in the case of knowledge spillover with $\varphi = 1$, as in Romer (1990); G. M. Grossman and Helpman (1991a); Aghion and Howitt (1992), growth of variety frontier is exponential and a *balanced growth path* is found analytically.

Remark 1 *Because of the homogeneity ($m(\tau,t,q,Q) \equiv 1$), the consumption in (1.39)*

$$C(t,q) = \frac{L_P(t)}{Q(t)},$$

the profit in (1.29)

$$\pi(t,q) = \frac{\Pi(t)}{Q(t)},$$

and, consequently, the patent value

$$v(t,q) = \frac{V(t)}{Q(t)},$$

defined in (1.31), do not depend on q.

Hence, for an internal solution $L_Q(t) > 0$, we can calculate, using equation (1.55), the value of patents in the economy as follows

$$V(t) = Q(t)\,v(t,Q(t)) = Q(t)\,\frac{L_Q(t)}{\dot{Q}(t)} = \frac{1}{\beta}, \tag{1.57}$$

that happens to be constant. No-arbitrage condition (1.33) yields the following expression for real interest rate

$$r(t) = \beta \left(\frac{L_P(t)}{\alpha} - \hat{L} \right). \tag{1.58}$$

The next proposition finds a "balanced growth path" corresponding to the exponential growth of the variety of goods in the economy.

First we find relation between labor endowment L defined in (1.4) and human wealth h defined in (1.24) when $r(t) \equiv 0$. For the sake of simplicity we consider the *stationary population and labor endowment*.

Definition 1 *We say that an economy has stationary population and labor endowment if the functions $n(\tau,t)$ and $l(\tau,t)$ depend only on the age $t-\tau \in [0,\omega]$ of the agent.*

Equivalently we can say that the number of newborns $n(\tau,\tau) \equiv n_0$ is constant and the mortality rate $\mu(\tau,t) \equiv \hat{\mu}(t-\tau)$ and the labor endowment $l(\tau,t) \equiv \hat{l}(t-\tau)$ depend only on the age $t-\tau$.

Lemma 1 *Let the population and labor endowment be stationary (see Definition 1). If the real interest rate is zero ($r(t) \equiv 0$) in the equilibrium, then the human wealth, given by (1.24), is constant $h(t) \equiv \hat{h}$, having the expression:*

$$\hat{h} = \int_0^{\omega} e^{-\int_0^{\varsigma} \hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\, \hat{l}(\varsigma)\,\mathrm{d}\varsigma, \tag{1.59}$$

and the total labor endowment is also constant and equals $L(t) \equiv \hat{L} = n_0 \hat{h}$, where n_0 is the number of newborns.

Proof See Appendix 1.5.4. □

The following Proposition finds an explicit solution of our general equilibrium model, when Q grows exponentially.

Proposition 4 *In the economy with stationary population and labor endowment if goods are homogeneous ($m(\tau,t,q,Q) \equiv 1$) and the knowledge spillover parameter is equal to one ($\varphi = 1$), then:*

(i) the triple $L_P(t) \equiv \alpha\hat{L}$, $r(t) \equiv 0$, $Q(t) = Q_0\, e^{\beta(1-\alpha)\hat{L}t}$ is a steady state solution of the system (1.36)–(1.38),

(ii) g in (1.40) grows exponentially w.r.t. time of births τ:

$$g(\tau) = \hat{g}\, e^{\beta(1-\alpha)\hat{L}\tau}, \tag{1.60}$$

where constant $\hat{g} = Q_0 \int_0^{\omega} e^{\left(\beta(1-\alpha)\hat{L}-\frac{\rho}{1-\alpha}\right)\varsigma - \int_0^{\varsigma} \hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\,\mathrm{d}\varsigma.$

Proof See Appendix 1.5.6. □

It follows from Lemma 1 and (1.60) that for $\varphi = 1$ the growth of the lifetime aggregated utility (1.56) is unbounded as the birth time of generations $\tau \to \infty$:

$$u(\tau) = \left(\alpha\,\hat{h}\right)^{\alpha}\left(\hat{g}\right)^{1-\alpha} e^{\beta(1-\alpha)^2\hat{L}\tau} \to \infty.$$

Aggregated assets A cannot grow infinitely because of the Proposition 2, thus converge to a constant $\hat{A}$.

For smaller values of spillover parameter φ, we show that the real interest rate r and the shares of labor L_p and L_Q converge to those of the balanced growth path.

1.2.4 Homogeneous consumption with $\varphi < 1$

In this case and we can obtain the real interest rate so we have

$$r(t) = \frac{\beta}{(Q(t))^{1-\varphi}}\left(\frac{1-\alpha}{\alpha}L_P(t) - \varphi L_Q(t)\right), \tag{1.61}$$

since $L_Q(t) > 0$, see Appendix 1.7 for general proof. We can calculate the value of patents in the economy with the use of (1.32) and (1.38) as

$$\begin{aligned} V(t) &= \int_0^{Q(t)} v(t,q)\,\mathrm{d}q \\ &= Q(t)\,v(t,Q(t)) \\ &= Q(t)\,\frac{L_Q(t)}{\dot{Q}(t)} \\ &= \frac{(Q(t))^{1-\varphi}}{\beta}, \end{aligned} \tag{1.62}$$

where we take into account that, because of the homogeneity ($m(\tau,t,q,Q) \equiv 1$), the consumption $C(t,q)$ in (1.39), the profit $\pi(s,q)$ in (1.29), and, consequently, the patent value $v(t,q)$ defined in (1.31), do not depend on q, see Remark 1.

Because of the assumed unbounded growth of the variety frontier, $Q(t) \to \infty$ as $t \to \infty$, see Fig. 1.1 b), and the first equation in (1.61), we have that $r(t) \to 0$, see Fig. 1.1 a). Thus, from equation (1.54) we obtain the convergence

$$L_P(t) \to \alpha\hat{L}, \qquad L_Q(t) \to (1-\alpha)\hat{L}, \tag{1.63}$$

see Fig. 1.1 c). The product variety frontier, $Q(t)$, grows asymptotically as driven by the equation

$$\dot{Q}(t) = \beta(1-\alpha)\hat{L}(Q(t))^{\varphi},$$

which implies the unbounded growth of patents' value (1.62):

$$V(t) = \frac{(Q(t))^{1-\varphi}}{\beta} \to \infty,$$

as $t \to \infty$, see Fig. 1.1 d). Hence, $V(t) > A(t)$ for all t since the assets A are bounded due to Proposition 2 and the patents' value V cannot intersect A transversely because both V and A are governed by the same ODE (1.54), see Proposition 1.

The growth of the life-time aggregated utility $u(\tau) = (\alpha h(\tau))^{\alpha} (g(\tau))^{1-\alpha}$ is also unbounded since $g(\tau) \sim Q(\tau) \to \infty$ due to (1.40)–(1.41) and $h(\tau) \to \check{h} \in (0, \hat{h}]$ due to (1.24).

The unbounded growth of the utility $u(\tau)$ means that the future generations will live unrestrictedly better than their presently living ancestors. This happens because in equilibrium the agent enjoys equal consumption of all goods, and her instantaneous utility can be described via the common consumption level $c(t) = c(\tau, t, q)$ and variety frontier $Q(t)$ as $Q(t)\, c(t)^{\alpha}$. It is clear, that the $Q(t)$-elasticity of the instantaneous utility is greater than its $c(t)$-elasticity, since $1 > \alpha$. Thus, the gain in utility from the increase of the variety frontier $Q(t)$ overweighs the loss from the decrease of the consumption level $c(t)$ so that the utilities of agents grow proportionally to $(Q(\tau))^{1-\alpha}$.

1.2.5 Heterogeneous consumption

We assume that at equilibrium $L_Q(t) > 0$ for all t and agents have heterogeneous preferences, such that

$$m(\tau, t, q, Q) = e^{-\gamma(Q-q)},$$

where $\gamma > 0$ is the parameter of heterogeneity[7]. This corresponds to the special case, where $m_0(\tau, t, Q) = e^{-\gamma Q}$ and $m_1(q) = e^{\gamma q}$. Thus, we can calculate the real interest rate r from (1.102), obtained in Appendix 1.7, as follows

$$r(t) = \frac{\beta}{(Q(t))^{1-\varphi}} \left(Q(t) \left(\frac{(1-\alpha)}{\alpha \tilde{M}(Q(t))} L_P(t) - \frac{\gamma}{1-\alpha} L_Q(t) \right) - \varphi L_Q(t) \right), \tag{1.64}$$

[7] The case of $\gamma = 0$ corresponds to homogeneous goods studied above.

where

$$\tilde{M}(Q(t)) = \frac{1-\alpha}{\gamma}(1 - e^{\frac{-\gamma}{1-\alpha}Q(t)}). \tag{1.65}$$

The value of patents can be expressed in the following way, with the use of (1.32) and (1.38) (see Appendix 1.6.2)

$$V(t) \quad = \quad \frac{1-\alpha}{\gamma\beta}\frac{1 - e^{\frac{-\gamma}{1-\alpha}Q(t)}}{Q(t)^{\varphi}}. \tag{1.66}$$

If we assume convergence to a steady state $r(t) \to \hat{r}$, then it follows from (1.64) that

$$\frac{(1-\alpha)}{\alpha\tilde{M}(Q(t))}L_P(t) - \frac{\gamma}{1-\alpha}L_Q(t) \to 0,$$

because otherwise $r(t) \to \infty$ as $Q(t) \to \infty$. Since $\tilde{M}(Q(t)) \to \frac{1-\alpha}{\gamma}$ in (1.65) when $Q(t) \to \infty$, we have

$$L_P(t) - \frac{\alpha}{1-\alpha}L_Q(t) \to 0$$

and, taking into account labor balance (1.37), we obtain

$$L_P(t) - \alpha L(t) \to 0.$$

The value of patents in the economy (1.66) tends to zero $V(t) \to 0$ as time $t \to \infty$. According to Corollary 1, $A(t) \equiv V(t)$ or $\hat{r} \leq 0$. We study the second type of solution, $\hat{r} \leq 0$, that is confirmed by the numerical calculation depicted in Fig. 1.2.

The life-time utility of an agent $u(\tau) = (\alpha\, h(\tau))^{\alpha}\,(g(\tau))^{1-\alpha}$ in (1.56) is bounded, because $h(\tau) \to const$ and $\tilde{M}(Q(t)) = M(Q(t)) \to \frac{1-\alpha}{\gamma}$ (therefore $g(\tau) \to const$). The intuition for the boundedness of the expected life time utility u is as follows. When the variety frontier Q increases the consumption profile, roughly speaking, shifts to the newer goods, thus exponentially (in Q) decreasing the consumption of the older goods, because of the function $m(q,Q) = e^{-\gamma(Q-q)}$ in (1.39). Exponential decrease in consumption of the older goods prevails over the increase of their variety Q. As a result the instantaneous utility converges to a limit constant so, consequently, does the life time utility $u(\tau)$.

We have qualitatively different behavior of V and u in the homogeneous ($m \equiv 1$) and the heterogeneous ($m \equiv e^{-\gamma(Q-q)}$) cases. Mathematically, this difference is conditioned by the properties of the integral $\tilde{M}(Q)$. Intuitively speaking, the reason for vanishing V and bounded u in the heterogeneous case is the abandonment of older goods because of the agent's preferences.

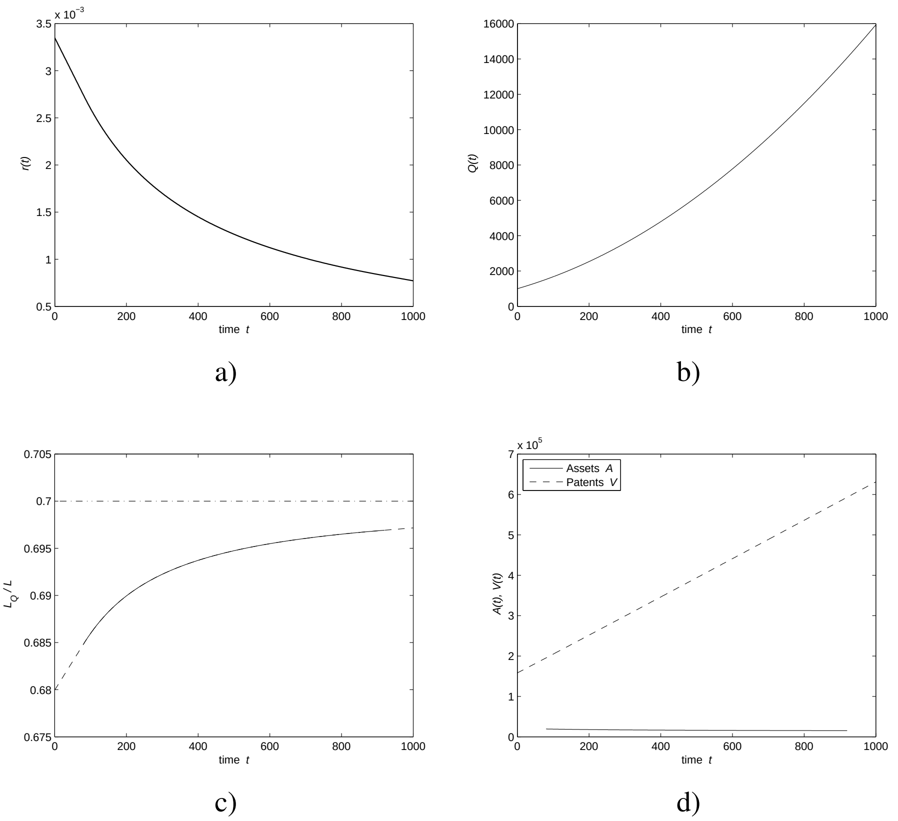


Figure 1.1: The case of homogeneous goods $m \equiv 1$. Labor endowment is $l(\tau,t) = 0.3$ for $t - \tau \leq 45$ and $l(\tau,t) = 0$ after retirement ($t - \tau > 45$), where we consider the agent's life from her adulthood. Mortality $\mu(\tau,t) = 0$ for $t - \tau < \omega$, life horizon $\omega = 80$, age concentration of people $n(\tau,t) = n_0 = 100$, initial frontier of the products' variety $Q_0 = 1000$, $\beta = 0.0002$, $\varphi = 0.5$, individual discounting $\rho = 0.01$. a) Real interest rate r. b) The variety frontier Q. c) The relative labor employed in R&D L_Q/L (solid line with dashed line depicting the linear extrapolation) and its limit value $1 - \alpha$ (dashed-dotted line). d) Dynamics of total assets A (solid line) and value of patents V (dashed line).

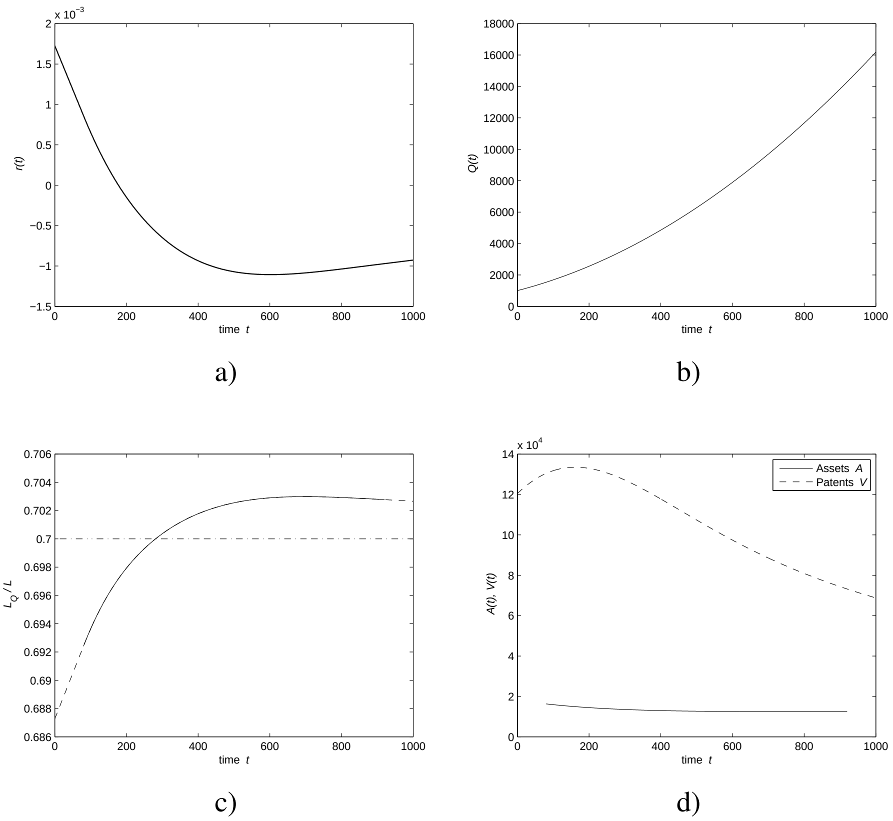


Figure 1.2: The case of heterogeneous goods $\gamma = 0.0004$. Labor endowment is $l(\tau,t) = 0.3$ for $t - \tau \leq 45$ and $l(\tau,t) = 0$ after retirement ($t - \tau > 45$), where we consider the agent's life from her adulthood. Mortality $\mu(\tau,t) = 0$ for $t - \tau < \omega$, life horizon $\omega = 80$, age concentration of people $n(\tau,t) = n_0 = 100$, initial frontier of the products' variety $Q_0 = 1000$, $\beta = 0.0002$, $\varphi = 0.5$, individual discounting $\rho = 0.01$. a) Real interest rate r. b) The variety frontier Q. c) The relative labor employed in R&D L_Q/L (solid line with dashed line depicting the linear extrapolation) and its limit value $1 - \alpha$ (dashed-dotted line). d) Dynamics of total assets A (solid line) and value of patents V (dashed line).

1.3 Discussion of the model and extensions

1.3.1 Efficiency and fiscal policy

In this subsection we investigate how the government can improve the utility of the agents in the long run by taxing the firms and paying subsidies and pensions to the agents. We assume that there is a constant tax rate $\delta < 1$ payed by each firm from its profit $\pi(t,q)$ and the sum of the collected taxes $\delta\,\Pi(t)$ are distributed among the currently living agents of different generations in the form of lump subsidies. For the sake of simplicity, we assume that the collected taxes are divided equally among the newborns and put on their deposits. The equation for government balance reads as

$$a(\tau,\tau)\,n(\tau,\tau) = \delta\,\Pi(\tau). \tag{1.67}$$

Taking the expression (1.48) for the total profit $\Pi(\tau)$ we obtain the following initial assets of the agents (instead of (1.7)):

$$a(\tau,\tau) = \delta\,\frac{1-\alpha}{\alpha}\,\frac{L_P(\tau)}{n(\tau,\tau)}. \tag{1.68}$$

The value $v(t,q)$ in (1.31) of the patent to produce product q decreases to

$$v(t,q) = (1-\delta)\int_t^\infty \exp\left(-\int_t^s r(\theta)\,\mathrm{d}\theta\right)\pi(s,q)\,\mathrm{d}s, \tag{1.69}$$

while the agent's *human wealth* $h(\tau)$ in (1.24) is augmented by *nonhuman wealth* from the subsidies $a(\tau,\tau) \geq 0$. Thus, in all the previous formulas we should replace $h(\tau)$ with the *total wealth* $a(\tau,\tau)+h(\tau)$. Taking into account these changes, the aggregated equations (1.51) and (1.53) take the form

$$\begin{aligned}\dot{A}(t) &= r(t)A(t)+L(t)-I(t)+\delta\,\Pi(t), &(1.70)\\ \dot{V}(t) &= r(t)V(t)+L_Q(t)-(1-\delta)\,\Pi(t), &(1.71)\end{aligned}$$

and similarly as in (1.54) we have

$$\dot{A}(t) = r(t)A(t)+L(t)-\frac{1-\delta\,(1-\alpha)}{\alpha}\,L_P(t). \tag{1.72}$$

Following the same arguments as before we can study the general equilibrium under simplifying assumptions, $L(t)\to\hat{L}$, etc.

Real interest rate

We can state that if $L_Q(t) > 0$ the real interest rate can be calculated as follows.

Homogeneous consumption ($m(\tau, t, q, Q) \equiv 1$):

$$r(t) = \frac{\beta}{(Q(t))^{1-\varphi}} \left(\frac{(1-\alpha)(1-\delta)}{\alpha} L_P(t) - \varphi L_Q(t) \right). \tag{1.73}$$

Heterogeneous consumption ($m(\tau, t, q, Q) \equiv e^{-\gamma(Q-q)}$ with $\gamma > 0$):

$$r(t) = \frac{\beta}{(Q(t))^{1-\varphi}} \left(Q(t) \left(\frac{(1-\alpha)(1-\delta)}{\alpha \tilde{M}(Q(t))} L_P(t) - \frac{\gamma}{1-\alpha} L_Q(t) \right) - \varphi L_Q(t) \right), \tag{1.74}$$

where $\tilde{M}(Q(t)) = \frac{1-\alpha}{\gamma}(1 - e^{\frac{-\gamma}{1-\alpha} Q(t)})$.

Labor shares

With both homogenous and heterogeneous consumptions we have

$$L_P(t) \to \frac{\alpha}{1-\delta(1-\alpha)} \hat{L}, \qquad L_Q(t) \to \frac{(1-\delta)(1-\alpha)}{1-\delta(1-\alpha)} \hat{L}, \tag{1.75}$$

extending (1.63) to the case of taxation and distribution of subsidies. Hence, the tax increases the amount of labor in production and decreases it in R&D, so that the product variety frontier $Q(t)$ grows asymptotically slower with the tax $\delta > 0$ then without it $\delta = 0$, which can be seen from its asymptotical differential equation

$$\dot{Q}(t) = \beta \frac{(1-\delta)(1-\alpha)}{1-\delta(1-\alpha)} \hat{L} (Q(t))^{\varphi}.$$

Using the explicit formula for the solution of this equation we can write

$$\frac{Q(t)|_{\delta=0}}{Q(t)|_{\delta>0}} \quad \to \quad \begin{cases} \left(1 + \frac{\delta\alpha}{1-\delta}\right)^{\frac{1}{1-\varphi}} , & \varphi < 1 \\ \exp\left(\frac{\beta \hat{L}(1-\alpha)\alpha\delta}{1-\delta(1-\alpha)} t\right), & \varphi = 1 \end{cases}, \tag{1.76}$$

where $Q(t)|_{\delta>0}$ denotes the variety frontier, when tax rate is positive, $\delta > 0$, while $Q(t)|_{\delta=0}$ is the variety frontier for economy without taxation.

Aggregated utility

Now we can ask the question: Can the fiscal policy improve the expected life-time aggregated utility (1.56) of a representative agents in the long run?

Substitution of the expression for nonhuman capital (1.68) in the utility of cohort τ brings it to the form:

$$u(\tau) = \alpha^{\alpha} \left(a(\tau,\tau) + h(\tau)\right)^{\alpha} \left(g(\tau)\right)^{1-\alpha}. \tag{1.77}$$

We see that $u(\tau)$ depends positively on the agents' human wealth and on the discounted integral M defined in (1.41).

Further we assume that population is stationary and labor is fixed, i.e. density $n(\tau,t)$ and labor endowment $l(\tau,t) \equiv \hat{l}(t-\tau)$ depend only on the age $t-\tau$, see Definition 1. Hence, the total labor endowment is constant, so we can take $L(t) = n_0 \bar{l}$, where $n_0 = n(\tau,\tau)$ is the constant number of newborns, and $\bar{l}$ is the constant effective labor of a newborn:

$$\bar{l} \equiv \int_0^{\omega} \hat{l}(\vartheta) e^{-\int_0^{\varsigma} \hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\,\mathrm{d}\varsigma, \tag{1.78}$$

which coincides with the limit of human wealth $\bar{l} = \hat{h}$ in the case of $\hat{r} = 0$, see expression (1.59) in Lemma 1. Thus, when $\tau \to \infty$ we have from (1.68) and (1.75) the initial assets converging as follows:

$$a(\tau,\tau) \to \frac{\delta\,(1-\alpha)}{1-\delta\,(1-\alpha)}\,\bar{l}. \tag{1.79}$$

In order to answer the question of the efficiency of fiscal policy we consider the asymptotics of the utility $u(\tau)$ in (1.77) separately in homogeneous and heterogeneous cases.

Long run efficiency of the fiscal policy

We will show, that the fiscal policy in hand can improve well-being ($u(\tau)$) of the future generations with heterogeneous consumption, which is not the case for homogeneous consumption. We check if, for all sufficiently high times of birth τ, life time utility in the economy with taxation, $u(\tau)|_{\delta>0}$, is grater than that without taxation, $u(\tau)|_{\delta=0}$.

Homogeneous consumption. In the homogeneous case $r(t) \to 0$ when $t \to \infty$, hence, according to Lemma 1, $h(\tau) \to \hat{h} = \bar{l}$ when $\tau \to \infty$. Thus, with the use of (1.76) and (1.79) we can write for $\varphi \in [0,1)$ the following limit for ratio of utilities with tax and without tax (see Appendix 1.6.3)

$$\frac{u(\tau)|_{\delta=0}}{u(\tau)|_{\delta>0}} \to \left(1-\delta\,(1-\alpha)\right)^{\alpha} \left(1+\frac{\delta\alpha}{1-\delta}\right)^{\frac{1-\alpha}{1-\varphi}} > \frac{1-\delta\,(1-\alpha)}{(1-\delta)^{1-\alpha}} \geq 1. \tag{1.80}$$

When $\varphi = 1$ this utility ratio tends to infinity with $t \to \infty$, due to the ratio in (1.76). Thus, in the homogeneous case the fiscal policy cannot improve agents' utilities in the long run.[8]

Heterogeneous consumption. Below we show that in contrast with the homogeneous case an appropriate fiscal policy of the government may improve the agents' utility in the long run. It is shown in Section 1.2.5 that $r(t) \to \hat{r} \le 0$.

The limit of human wealth (1.24), as $\tau \to \infty$, can be calculated as follows

$$h(\tau) \to \int_0^{\omega} \hat{l}(\vartheta) e^{-\hat{r}\varsigma - \int_0^{\varsigma} \hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\,\mathrm{d}\varsigma \equiv \bar{h} \tag{1.81}$$

where $\bar{h}$ does not depend on δ. The integral in (1.77), in contrast to the homogeneous case, converges to a constant independent of tax rate δ:

$$g(\tau) \to \frac{1-\alpha}{\gamma} \int_0^{\omega} e^{-\frac{\rho-\alpha\hat{r}}{1-\alpha}\varsigma - \int_0^{\varsigma} \hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\,\mathrm{d}\varsigma \equiv \bar{g}. \tag{1.82}$$

Then, we can make the following approximations of the asymptotic expected utility (see Appendix 1.6.4)

$$u(\tau)|_{\delta} \to \alpha^{\alpha} \left(\frac{\delta(1-\alpha)}{1-\delta(1-\alpha)} \bar{l} + \bar{h} \right)^{\alpha} \bar{g}^{1-\alpha}. \tag{1.83}$$

where $\bar{l}$ is defined in (1.78). It is seen from (1.83), that the implementation of the fiscal policy ($\delta > 0$) increases the agents' expected life-time aggregated utilities in the long run

$$\lim_{\tau\to\infty} \frac{u(\tau)|_{\delta=0}}{u(\tau)|_{\delta>0}} \le 1. \tag{1.84}$$

Notice, that these results hold for any reasonable functions of mortality, $\hat{\mu}$, and labor endowment, $\hat{l}$.

1.3.2 Sources of heterogeneity

There are two, mathematically equivalent, sources of heterogeneity of goods. The first is that we have already described, where goods are weighted heterogeneously in the agent's utility function. The other source of heterogeneity is the potential difference in productivity of labor allocated to different firms.

[8]Note that for $\varphi < 0$ even small tax rate $\delta > 0$ improves agents' utilities in the long run, because the derivative $\frac{\mathrm{d}}{\mathrm{d}\delta} \frac{u(\tau)|_{\delta=0}}{u(\tau)|_{\delta>0}}\Big|_{\delta=0} = \alpha\varphi\frac{1-\alpha}{1-\varphi}$ becomes negative. A value $\varphi < 0$ means that past inventions make it more difficult to find new ideas, which we think to be unlikely.

That can be described by the same model after renormalization. Indeed, let $\zeta(t,q,Q(t))$ be the amount of physical units of the good q that can be produced with one unit of labor, then equation for the profit (1.27) would take the form

$$\pi(t,q) = \tilde{p}(t,q)\tilde{C}(t,q) - \frac{\tilde{C}(t,q)}{\zeta(t,q,Q(t))}, \tag{1.85}$$

where $\tilde{p}(t,q)$ is the price of one physical unit of the good q, $\tilde{C}(t,q)$ is the aggregated production (and consumption) in physical units of the good q. Let the agent value all goods equally maximizing her following expected lifetime utility

$$u(\tau) = \int_{\tau}^{\tau+\omega} e^{-\rho(t-\tau)-\int_{\tau}^{t}\mu(\tau,\theta)\,\mathrm{d}\theta} \int_0^{Q(t)} \tilde{c}(\tau,t,q)^{\alpha}\,\mathrm{d}q\,\mathrm{d}t, \tag{1.86}$$

subject to the dynamic budget constraint

$$\dot{a}(\tau,t) = l(\tau,t) + (r(t)+\mu(\tau,t))\,a(\tau,t) - \int_0^{Q(t)} \tilde{p}(t,q)\,\tilde{c}(\tau,t,q)\,\mathrm{d}q, \tag{1.87}$$

with boundary conditions

$$a(\tau,\tau) = 0, \qquad a(\tau,\tau+\omega) = 0, \tag{1.88}$$

where $\tilde{c}(\tau,t,q)$ is her consumption in physical units. Then, if we change the variables as follows

$$\tilde{p}(t,q) = \frac{p(t,q)}{\zeta(t,q,Q(t))}, \quad \tilde{c}(\tau,t,q) = c(\tau,t,q)\,\zeta(t,q,Q(t)),$$

so that

$$\tilde{C}(t,q) = C(t,q)\,\zeta(t,q,Q(t)),$$

then equation (1.85) for the profit will coincide with (1.27) and problem (1.86)–(1.88) will take the form (1.5)–(1.8), where

$$m \equiv (\zeta(t,q,Q(t)))^{\alpha}.$$

Thus, the performed analysis is also applicable to the problem with heterogeneous productivity of labor, but with a different meanings of the function m.

1.3.3 Economic growth

So far we have studied only the growth of agents' utilities as indicator of prosperity. However, measuring goods in physical units as we did in Section 1.3.2 we can also discuss economic growth in terms of production and consumption. If we assume limited labor endowment L the economy can grow only due to increase in productivity of labor $\zeta(t,q,Q(t))$, introduced in Section 1.3.2. The growth can be endogenous if the productivity of labor depends on the state variable $Q(t)$.

We can consider a function m in problem (1.5) in the form

$$m(\tau,t,q,Q(t)) \equiv (\zeta(t,q,Q(t)))^{\alpha}\,\tilde{m}(\tau,t,q,Q(t)), \tag{1.89}$$

where the function $\tilde{m}$, inducing the actual agent's preferences, is bounded, while the function $\zeta(q,t,Q(t))$, describing the dependence of the productivity upon the variety frontier $Q(t)$, may grow unboundedly when $Q(t) \to \infty$. Thus, we would have an unbounded growth of per capita consumption (in physical units) i.e. an infinite economic growth.

The functional form of $\zeta(t,q,Q(t))$ is supposed to be chosen to fit an observed path of the per capita consumption (Jones, 1995b), but this is beyond the scope of the present study.

1.4 Conclusions and prospects

We suggest an endogenous growth model of an economy, where technological growth is promoted by the entrepreneurial activity of new firms. The liquidity for these new enterprises is provided by the banking sector that *de jure* owns all intellectual property. We interpret the technological growth as a growth of the variety of goods. The question of how productivity of labor depends on the variety of available technologies (goods) still needs to be answered. One can try to determine this dependence empirically, but we believe that the productivity of labor should be related with qualification, hence education should be explicitly included in the model as a new decision variable of the individuals.

Heterogeneity of consumption ($m(\tau,t,q,Q) = e^{-\gamma(Q-q)}$, where $\gamma > 0$) brings the following qualitative effects. Firstly, the total value of patents V become bounded with zero limit $V(t) \to 0$ when $t \to \infty$ (in case of knowledge spillovers, $\varphi > 0$), thus balancing savings of agents and values of patents in the long run ($V(t) - A(t) \to 0$). Secondly, the growth of the expected life-time aggregated utilities of agents becomes bounded. Finally, the general equilibrium with heterogeneous consumption can loose its dynamic efficiency. So that it becomes possible to introduce a fiscal policy that improves all agents' utilities in the long run. Dynamic inefficiency is observed in OLG models quite often, but in our model general equilibrium is inefficient only when consumption is heterogeneous.

In the case of constant labor endowment and stationary population consuming homogeneous goods we have found the simple steady steady state solution with zero real interest rate and ($r \equiv 0$) and constant share of labor in production ($L_P/L \equiv \alpha$) with knowledge spillover parameter $\varphi = 1$. This solution sustain exponential growth of goods' variety. The same interest rate and labor distribution happen to be an attractor for the solutions in homogeneous cases with $\varphi < 1$ and heterogeneous cases with $\varphi < 0$. Thus, in all cases the variety of goods can grow unboundedly.

Since the general equilibrium can be numerically calculated without any steady state assumptions and with arbitrary exogenous population dynamics the model allows to investigate the effect of different shocks including demographical changes. The proposed model allows for introduction of functional dependence of labor productivity on the goods' variety frontier, thus describing the endogenous growth of production. Finding such dependence that would mitigate the scale effect could be the topic for further research.

Appendix

1.5 Proofs

1.5.1 Proposition 1

Proof Let us substitute profit expression (1.43) into differential equation (1.53) for V and use labor balance (1.35). Thus, we obtain exactly the same equation (1.51) as for the aggregated assets A. One can check that the substitution of the relation

$$V(t) = A(t) + \psi e^{\int_0^t r(s)\,\mathrm{d}s}$$

into equation (1.53) gives equation (1.51). Moreover, due to (1.47) equation (1.51) can be written in terms of labor, (1.54). □

1.5.2 Proposition 2

Proof Expression (1.49), with $a(\tau,t)$ taken from the solution of personal budget constraints (1.13) same as in (Cass & Yaari, 1967), has the following form

$$\begin{aligned} A(t) &= \int_{t-\omega}^{t} \frac{n(\tau,t)}{R_\mu(\tau,t)} \int_{\tau}^{t} R_\mu(\tau,s)\,(l(\tau,s)-E(\tau,s))\,\mathrm{d}s\,\mathrm{d}\tau \\ &= \int_{t-\omega}^{t} n(\tau,t) \int_{\tau}^{t} e^{-\int_t^s (r(\theta)+\mu(\tau,\theta))\,\mathrm{d}\theta}\,(l(\tau,s)-E(\tau,s))\,\mathrm{d}s\,\mathrm{d}\tau \\ &= \int_{t-\omega}^{t} n(\tau,t) \int_{\tau}^{t} e^{\int_s^t r(\theta)\,\mathrm{d}\theta} \frac{n(\tau,s)}{n(\tau,t)}\,(l(\tau,s)-E(\tau,s))\,\mathrm{d}s\,\mathrm{d}\tau \\ &= \int_{t-\omega}^{t} \int_{\tau}^{t} e^{\int_s^t r(\theta)\,\mathrm{d}\theta} n(\tau,s)\,(l(\tau,s)-E(\tau,s))\,\mathrm{d}s\,\mathrm{d}\tau \\ &= \int_{t-\omega}^{t} e^{\int_s^t r(\theta)\,\mathrm{d}\theta} \int_{t-\omega}^{s} n(\tau,s)\,(l(\tau,s)-E(\tau,s))\,\mathrm{d}\tau\,\mathrm{d}s \end{aligned} \tag{1.90}$$

Since $l(\tau,s) \geq 0$ and $E(\tau,s) \geq 0$ we have the following chain of inequalities

$$\begin{aligned} |A(t)| &\leq e^{\omega\bar{r}} \int_{t-\omega}^{t} \int_{t-\omega}^{s} n(\tau,s)\,(l(\tau,s)+E(\tau,s))\,\mathrm{d}\tau\,\mathrm{d}s \\ &\leq e^{\omega\bar{r}} \int_{t-\omega}^{t} \int_{t-\omega}^{t} n(\tau,s)\,(l(\tau,s)+E(\tau,s))\,\mathrm{d}\tau\,\mathrm{d}s \\ &= e^{\omega\bar{r}} \int_{t-\omega}^{t} (L(s)+I(s))\,\mathrm{d}s \\ &= e^{\omega\bar{r}} \int_{t-\omega}^{t} \left(L(s)+\frac{L_P(s)}{\alpha}\right)\mathrm{d}s \\ &\leq e^{\omega\bar{r}} \int_{t-\omega}^{t} \left(\bar{L}+\frac{\bar{L}}{\alpha}\right)\mathrm{d}s \\ &\leq \omega e^{\omega\bar{r}} \frac{1+\alpha}{\alpha}\bar{L}, \end{aligned}$$

where we use expression (1.4) for total labor, balance equation (1.46) and relation (1.47). □

1.5.3 Corollary 1

Proof Proposition 2 claims that the aggregated assets A in the economy are always bounded. Since both A and V are bounded, the term with exponent in the relation

$$V(t) = A(t) + \psi e^{\int_0^t r(s)\,\mathrm{d}s}$$

from Proposition 1 is also bounded. This happens only if constant $\psi = 0$ (then $V(t) = A(t)$ for all t) or when the exponent $\exp\left(\int_0^t r(s)\,\mathrm{d}s\right)$ is bounded. But we have that $r(t) \to \hat{r}$, hence boundedness of $\int_0^t r(s)\,\mathrm{d}s$ occurs only if $\hat{r} \leq 0$. □

1.5.4 Lemma 1

Proof First we show that the human wealth h defined in (1.24) is constant:

$$\begin{aligned} h(\tau) &\equiv \int_\tau^{\tau+\omega} R_\mu(\tau,s)\, l(\tau,s)\,\mathrm{d}s \\ &= \int_\tau^{\tau+\omega} e^{-\int_\tau^s \hat{\mu}(\theta-\tau)\,\mathrm{d}\theta}\, \hat{l}(s-\tau)\,\mathrm{d}s \\ &= \int_\tau^{\tau+\omega} e^{-\int_0^{s-\tau} \hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\, \hat{l}(s-\tau)\,\mathrm{d}s \\ &= \int_0^{\omega} e^{-\int_0^{\varsigma} \hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\, \hat{l}(\varsigma)\,\mathrm{d}\varsigma \equiv \hat{h}, \end{aligned} \tag{1.91}$$

where we used the definition of $R_\mu(\tau,s)$ in (1.19) along with $r(t) \equiv 0$. Then, from the definition of L in (1.4) and expression (1.3) we have

$$\begin{aligned} L(t) &\equiv \int_{t-\omega}^{t} l(\tau,t)\, n(\tau,t)\,\mathrm{d}\tau \\ &= \int_{t-\omega}^{t} \hat{l}(t-\tau)\, n(\tau,\tau)\, e^{-\int_\tau^t \hat{\mu}(\theta-\tau)\,\mathrm{d}\theta}\,\mathrm{d}\tau \\ &= n_0 \int_{t-\omega}^{t} \hat{l}(t-\tau)\, e^{-\int_0^{t-\tau} \hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\,\mathrm{d}\tau \\ &= n_0 \int_0^{\omega} \hat{l}(\varsigma)\, e^{-\int_0^{\varsigma} \hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\,\mathrm{d}\varsigma = n_0 \hat{h}. \end{aligned} \tag{1.92}$$

□

1.5.5 Proposition 3

Proof We consider two cases: inner and corner solutions.
Inner solution. If $Q(t) \to \bar{Q}$ and $Q(t) < \bar{Q}$ for all t, then due to $L_Q(t) \to 0$ equation (1.102) yields a bounded real interest rate r with strictly positive limit:

$$\lim_{t\to\infty} r(t) = \hat{r} = \beta \bar{Q}^{\varphi} \frac{1-\alpha}{\alpha} \frac{m_1(\bar{Q})^{\frac{1}{1-\alpha}}}{M_1(\bar{Q})} \hat{L} > 0.$$

From expression (1.31) and equation (1.29) we have bounded V due to the boundedness of r and L

$$\begin{aligned}
V(t) &= \int_0^{Q(t)} v(t,q)\,\mathrm{d}q \\
&= \int_0^{Q(t)} \int_t^\infty e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} \pi(s,q)\,\mathrm{d}s\,\mathrm{d}q \\
&= \int_t^\infty e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} \frac{1-\alpha}{\alpha} \int_0^{Q(t)} C(s,q)\,\mathrm{d}q\,\mathrm{d}s \\
&\le \int_t^\infty e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} \frac{1-\alpha}{\alpha} \int_0^{Q(s)} C(s,q)\,\mathrm{d}q\,\mathrm{d}s \\
&= \int_t^\infty e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} \frac{1-\alpha}{\alpha} L_P(s)\,\mathrm{d}s \\
&\le \int_t^\infty e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} \frac{1-\alpha}{\alpha} L(s)\,\mathrm{d}s = \frac{1-\alpha}{\alpha} \int_t^\infty e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} L(s)\,\mathrm{d}s < \infty.
\end{aligned}$$

Since $\bar{r} > 0$ the upper bound of $V(t)$ is finite and has the following limit with $t \to \infty$

$$\frac{1-\alpha}{\alpha} \int_t^\infty e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} L(s)\,\mathrm{d} \quad \to \quad \frac{1-\alpha}{\alpha} \hat{L} \int_t^\infty e^{-\hat{r}(s-t)\,\mathrm{d}\theta}\,\mathrm{d}s = \frac{1-\alpha}{\alpha} \frac{\hat{L}}{\hat{r}}.$$

Corner solution. If there exist a time $\bar{t}$ such that $L_Q(t) = 0$ and $Q(t) = \bar{Q}$ for all $t \ge \bar{t}$. Convergence of $r(t) \to \hat{r}$ and $L(t) \to \hat{L} > 0$ results in bounded A, according to Proposition 2. Hence, the total amount of assets A also converges to a constant, $A(t) \to \hat{A}$, due to equation (1.54) from which we have the following

$$0 = \hat{r}\hat{A} - \frac{1-\alpha}{\alpha} \hat{L}, \tag{1.93}$$

so that $\hat{r} \neq 0$ and $\hat{A} \neq 0$. If $\hat{r} > 0$ then for $t \ge \bar{t}$

$$\begin{aligned}
V(t) &= \int_0^{Q(t)} v(t,q)\,\mathrm{d}q = \int_0^{\bar{Q}} v(t,q)\,\mathrm{d}q = \int_0^{\bar{Q}} \int_t^\infty e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} \pi(s,q)\,\mathrm{d}s\,\mathrm{d}q \\
&= \frac{1-\alpha}{\alpha} \int_t^\infty e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} \int_0^{\bar{Q}} C(s,q)\,\mathrm{d}q\,\mathrm{d}s \\
&= \frac{1-\alpha}{\alpha} \int_t^\infty e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} L_P(s)\,\mathrm{d}s \\
&\to \frac{1-\alpha}{\alpha} \hat{L} \int_t^\infty e^{-\hat{r}(s-t)\,\mathrm{d}\theta}\,\mathrm{d}s = \frac{1-\alpha}{\alpha} \frac{\hat{L}}{\hat{r}} = \hat{A} > 0.
\end{aligned}$$

The case of $\hat{r} < 0$ contradicts Proposition 1. It follows from equation (1.93) that $\hat{A} < 0$ when $\hat{r} < 0$. But according to the relation

$$V(t) = A(t) + \psi e^{\int_0^t r(s)\,\mathrm{d}s}$$

in Proposition 1 we have $V(t) \to \hat{A}$ so that $\hat{A} > 0$ because $V(t) > 0$ by definition.

Then, in both inner and corner solutions we have $\hat{r} > 0$ and bounded V. Thus, the balance $A(t) \equiv V(t)$ follows from Corollary 1. □

1.5.6 Proposition 4

Proof We need to check if $L_P(t) \equiv \alpha\hat{L}$, $r(t) \equiv 0$, and $Q(t) = Q_0\, e^{\beta(1-\alpha)\hat{L}t}$ satisfy equations (1.36)–(1.38). It is easy to check that (1.36) is satisfied due to the no-arbitrage condition in (1.61). The solution of (1.38) is obviously

$$Q(t) = Q_0\, e^{\beta(1-\alpha)\hat{L}t}.$$

Let us check (1.37). From (1.41) it follows that

$$M(\tau,s) = Q(s) = Q_0\, e^{\beta(1-\alpha)\hat{L}s}.$$

Then from (1.40) we obtain expression (1.60) for $g(\tau)$ as follows

$$\begin{aligned} g(\tau) &\equiv \int_\tau^{\tau+\omega} R_\rho(\tau,s) R_\mu(\tau,s) M(\tau,s)\,\mathrm{d}s \\ &= Q_0\, e^{\beta(1-\alpha)\hat{L}\tau} \int_\tau^{\tau+\omega} e^{\left(\beta(1-\alpha)\hat{L} - \frac{\rho}{1-\alpha}\right)(s-\tau) - \int_0^{s-\tau} \hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\,\mathrm{d}s \\ &= Q_0\, e^{\beta(1-\alpha)\hat{L}\tau} \int_0^{\omega} e^{\left(\beta(1-\alpha)\hat{L} - \frac{\rho}{1-\alpha}\right)\varsigma - \int_0^{\varsigma} \hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\,\mathrm{d}\varsigma \\ &= Q_0\, e^{\beta(1-\alpha)\hat{L}\tau}\, \hat{g}, \end{aligned}$$

where the last integral is a constant denoted by $\hat{g}$. Thus, from (1.39) we have

$$\begin{aligned} C(t,q) &= \alpha \int_{t-\omega}^{t} \frac{h(\tau)}{g(\tau)} R_\rho(\tau,t)\, n(\tau,t)\,\mathrm{d}\tau \\ &= \alpha\hat{h} \int_{t-\omega}^{t} \frac{1}{g(\tau)}\, e^{-\frac{\rho}{1-\alpha}(t-\tau)}\, n(\tau,\tau)\, e^{-\int_0^{t-\tau}\hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\,\mathrm{d}\tau \\ &= \alpha\hat{h}n_0 \int_{t-\omega}^{t} \frac{1}{g(\tau)}\, e^{-\frac{\rho}{1-\alpha}(t-\tau)}\, e^{-\int_0^{t-\tau}\hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\,\mathrm{d}\tau. \end{aligned} \tag{1.94}$$

After substitution of $g(\tau)$ from (1.60) and canceling the constant we obtain the following expression with the use of the relation $\hat{L} = \hat{h}n_0$ from Lemma 1

$$\begin{aligned}
C(t,q) &= \frac{\alpha \hat{h} n_0}{Q_0\,\hat{g}} \int_{t-\omega}^{t} e^{-\beta(1-\alpha)\hat{L}\tau}\, e^{-\frac{\rho}{1-\alpha}(t-\tau) - \int_0^{t-\tau}\hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\,\mathrm{d}\tau \\
&= \frac{\alpha \hat{h} n_0}{Q_0\,\hat{g}} e^{-\beta(1-\alpha)\hat{L}t} \int_{t-\omega}^{t} e^{\beta(1-\alpha)\hat{L}(t-\tau) - \frac{\rho}{1-\alpha}(t-\tau) - \int_0^{t-\tau}\hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\,\mathrm{d}\tau \\
&= \frac{\alpha \hat{h} n_0}{Q_0\, e^{\beta(1-\alpha)\hat{L}t}\,\hat{g}} \int_0^{\omega} e^{\left(\beta(1-\alpha)\hat{L} - \frac{\rho}{1-\alpha}\right)\varsigma - \int_0^{\varsigma}\hat{\mu}(\vartheta)\,\mathrm{d}\vartheta}\,\mathrm{d}\varsigma \\
&= \frac{\alpha \hat{h} n_0}{Q_0\, e^{\beta(1-\alpha)\hat{L}t}} \\
&= \frac{\alpha \hat{h} n_0}{Q(t)} = \frac{\alpha \hat{L}}{Q(t)} = \frac{L_P(t)}{Q(t)} \\
&= \frac{\hat{L} - L_Q(t)}{Q(t)}. \qquad (1.95)
\end{aligned}$$

Thus (1.37) is also satisfied. □

1.6 Calculations

1.6.1 Expression (1.56)

With the use of consumption (1.23) and relation $G(\tau,s) = \alpha^{\frac{\alpha}{1-\alpha}} M(\tau,s)$

$$\begin{aligned}
u(\tau) &= \int_{\tau}^{\tau+\omega} e^{-\rho(t-\tau) - \int_{\tau}^{t}\mu(\tau,\theta)\,\mathrm{d}\theta} \int_0^{Q(t)} m(\tau,t,q,Q(t))\; c(\tau,t,q)^{\alpha}\,\mathrm{d}q\,\mathrm{d}t \\
&= \int_{\tau}^{\tau+\omega} e^{-\rho(t-\tau) - \int_{\tau}^{t}\mu(\tau,\theta)\,\mathrm{d}\theta} (G(\tau,t))^{1-\alpha} (E(\tau,t))^{\alpha}\,\mathrm{d}t \\
&= \int_{\tau}^{\tau+\omega} e^{-\rho(t-\tau) - \int_{\tau}^{t}\mu(\tau,\theta)\,\mathrm{d}\theta} (\alpha^{\frac{\alpha}{1-\alpha}} M(\tau,s))^{1-\alpha} \left(R_{\rho}(\tau,t)\, h(\tau)\, \frac{M(\tau,t)}{g(\tau)} \right)^{\alpha} \mathrm{d}t \\
&= \left(\alpha \frac{h(\tau)}{g(\tau)} \right)^{\alpha} \int_{\tau}^{\tau+\omega} R_{\rho}(\tau,t) R_{\mu}(\tau,t) M(\tau,t)\mathrm{d}t \\
&= (\alpha\, h(\tau))^{\alpha} (g(\tau))^{1-\alpha},
\end{aligned}$$

we have (1.56).

1.6.2 Expression (1.66)

$$
\begin{aligned}
V(t) &= \int_0^{Q(t)} v(t,q)\,\mathrm{d}q \\
&= \int_0^{Q(t)} \int_t^{\infty} e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} \pi(s,q)\,\mathrm{d}s\,\mathrm{d}q \\
&= \int_0^{Q(t)} \int_t^{\infty} e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} \left(\frac{m(q,Q(s))}{m(Q(t),Q(s))}\right)^{\frac{1}{1-\alpha}} \pi(s,Q(t))\,\mathrm{d}s\,\mathrm{d}q \\
&= \int_0^{Q(t)} \int_t^{\infty} e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} e^{-\frac{\gamma}{1-\alpha}(Q(t)-q)} \pi(s,Q(t))\,\mathrm{d}s\,\mathrm{d}q \\
&= \int_t^{\infty} e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} \pi(s,Q(t)) \int_0^{Q(t)} e^{-\frac{\gamma}{1-\alpha}(Q(t)-q)}\,\mathrm{d}q\,\mathrm{d}s \\
&= \tilde{M}(Q(t)) \int_t^{\infty} e^{-\int_t^s r(\theta)\,\mathrm{d}\theta} \pi(s,Q(t))\,\mathrm{d}s = \tilde{M}(Q(t))\, v(t,Q(t)) \\
&= \tilde{M}(Q(t)) \frac{L_Q(t)}{\dot{Q}(t)} = \frac{\tilde{M}(Q(t))}{\beta Q(t)^{\varphi}} = \frac{1-\alpha}{\gamma\beta} \frac{1-e^{\frac{-\gamma}{1-\alpha}Q(t)}}{Q(t)^{\varphi}}.
\end{aligned}
$$

1.6.3 Expression (1.84)

$$
\begin{aligned}
\frac{u(\tau)|_{\delta=0}}{u(\tau)|_{\delta>0}} &\to \left(\frac{\bar{l}}{\frac{\delta(1-\alpha)}{1-\delta(1-\alpha)}\bar{l}+\bar{l}}\right)^{\alpha} \left(1+\frac{\delta\alpha}{1-\delta}\right)^{\frac{1-\alpha}{1-\varphi}} \\
&= (1-\delta(1-\alpha))^{\alpha} \left(1+\frac{\delta\alpha}{1-\delta}\right)^{\frac{1-\alpha}{1-\varphi}} \\
&> (1-\delta(1-\alpha))^{\alpha} \left(1+\frac{\delta\alpha}{1-\delta}\right)^{1-\alpha} \\
&= \frac{1-\delta(1-\alpha)}{(1-\delta)^{1-\alpha}} \geq 1.
\end{aligned}
$$

1.6.4 Expression (1.83)

From expression (1.68) and limits (1.79), (1.81), and (1.82) we have the limit of aggregated utility

$$
u(\tau) = \alpha^{\alpha}(a(\tau,\tau)+h(\tau))^{\alpha}(g(\tau))^{1-\alpha}
$$

$$
\begin{aligned}
&= \alpha^{\alpha}\left(\delta\,\frac{1-\alpha}{\alpha}\,\frac{L_P(\tau)}{n(\tau,\tau)}+h(\tau)\right)^{\alpha}(g(\tau))^{1-\alpha}\\
&= \to \alpha^{\alpha}\left(\frac{\delta\,(1-\alpha)}{1-\delta\,(1-\alpha)}\bar{l}+\bar{h}\right)^{\alpha}\bar{g}^{1-\alpha}.
\end{aligned}
$$

1.7 No-arbitrage condition for $m = m_0(\tau,t,Q)\,m_1(q)$

Let us rewrite expression (1.39) using the assumption

$$m(\tau,t,q,Q) = m_0(\tau,t,Q)\,m_1(q).$$

Then we have

$$
\begin{aligned}
C(t,q) &= \alpha\,(m_1(q))^{\frac{1}{1-\alpha}}\int_{t-\omega}^{t}\frac{(m_0(\tau,t,Q(t)))^{\frac{1}{1-\alpha}}}{g(\tau)}R_{\rho}(\tau,t)\,h(\tau)\,n(\tau,t)\,\mathrm{d}\tau\\
&= (m_1(q))^{\frac{1}{1-\alpha}}\,f(t), \qquad (1.96)
\end{aligned}
$$

where function f denotes the integral and function $m_1(q) > 0$ is differentiable for all $q \geq Q_0$. Then from (1.34) and (1.96) we have

$$L_P(t) = \int_0^{Q(t)} C(t,q)\,\mathrm{d}q = f(t)\int_0^{Q(t)}(m_1(q))^{\frac{1}{1-\alpha}}\,\mathrm{d}q = f(t)\,M_1(Q(t)). \quad (1.97)$$

We can derive from (1.96) and (1.97) the following relations for $C(s,q)$ and, due to (1.29), for $\pi(s,q)$:

$$
\begin{aligned}
C(s,q) &= \frac{(m_1(q))^{\frac{1}{1-\alpha}}}{M_1(Q(s))}L_P(s), \qquad (1.98)\\
\pi(s,q) &= \frac{1-\alpha}{\alpha}C(s,q) = \frac{1-\alpha}{\alpha}\,\frac{(m_1(q))^{\frac{1}{1-\alpha}}}{M_1(Q(s))}L_P(s), \qquad (1.99)
\end{aligned}
$$

where we introduce notation

$$M_1(Q) = \int_0^{Q} m_1(q)^{\frac{1}{1-\alpha}}\,\mathrm{d}q, \qquad (1.100)$$

so that the integral defined in (1.41) takes the form

$$M(\tau,t) = m_0(\tau,t,Q(t))^{\frac{1}{1-\alpha}}\,M_1(Q(t)).$$

Thus, we have the expression for the derivative of $C(s,q)$ with respect to q

$$\left.\frac{\partial C}{\partial q}(s,q)\right|_{q=Q(t)} = \frac{(m_1(Q(t)))^{\frac{\alpha}{1-\alpha}}\,\frac{\partial m_1}{\partial q}(Q(t))}{(1-\alpha)\,M_1(Q(s))}L_P(s) = \frac{C(s,Q(t))}{1-\alpha}\,\frac{\frac{\partial m_1}{\partial q}(Q(t))}{m_1(Q(t))},$$

and, due to (1.29), the derivative of $\pi(s,q)$:

$$\frac{\partial \pi}{\partial q}(s,q)\bigg|_{q=Q(t)} = \frac{1-\alpha}{\alpha}\,\frac{\partial C}{\partial q}(s,q)\bigg|_{q=Q(t)} = \frac{\pi(s,Q(t))}{1-\alpha}\,\frac{\frac{\partial m_1}{\partial q}(Q(t))}{m_1(Q(t))}. \quad (1.101)$$

It is assumed that there is always R&D activity $L_Q(t) > 0$. Due to (1.101), (1.99), and (1.55) no-arbitrage condition (1.33) takes the form

$$r(t) = \frac{\beta}{(Q(t))^{1-\varphi}}\left(Q(t)\left(\frac{1-\alpha}{\alpha}\,\frac{L_P(t)}{\tilde{M}(Q(t)))} - \frac{L_Q(t)}{1-\alpha}\,\frac{\frac{\partial m_1}{\partial q}(Q(t))}{m_1(Q(t))}\right) - \varphi L_Q(t)\right), \quad (1.102)$$

where

$$\tilde{M}(Q(t)) = \frac{M_1(Q(t))}{m_1(Q(t))^{\frac{1}{1-\alpha}}}.$$

This condition is easier to use instead of zero-profit condition (1.36) in the R&D sector. Although condition (1.102) does not depend explicitly on the function $m_0(\tau,t,Q)$, it is still affected by m_0 through expressions (1.39), (1.40), and (1.41). However, if the function m_0 depends only on τ, then it does not influence the solution $(L_Q(t), r(t), Q(t))$ of system ((1.37), (1.38), (1.102)) and matters only for the utilities of the agents. Indeed, consumption profile (over goods q and times t) of an agent would remain the same if she have changed her weight function m proportionately for all available goods and times.

1.8 Iterative procedure to calculate general equilibrium

The expression for the real interest rate (1.102) allows to construct a simple iterative procedure for numerical calculation of a general equilibrium in the case of interior solution ($L_Q(t) > 0$ for all t).

In order to solve system (1.37), (1.38), (1.102) first, we give the initial labor in research $L_Q[1](t)$ in the time interval $[0,T]$, where $T \gg \omega$. With $L_Q[1](t)$ we calculate the product variety domain $Q[1](t)$ in $[0,T]$ from (1.38). Then, we calculate the real interest rate $r[1](t)$ in $[0,T]$ with the use of (1.102). The functions $Q[1](t)$ and $r[1](t)$ determine the aggregated consumption of agents $C[1](t,q)$ in $[\omega, T-\omega]$ by (1.39). Knowing $C[1](t,q)$ we obtain the labor in R&D for the next iteration, $L_Q[2](t)$, $t \in [\omega, T-\omega]$ with the formula (1.37) and extrapolate it linearly to the whole interval $[0,T]$. Then we continue in the same way.

If the sequence of so defined iterations converges, then as a result we have a numerical approximation of the general equilibrium with two allowances:

finiteness of time horizon T and linear extrapolation to the ends of the time interval.

There are two cases to be considered separately. The first case is when consumers evaluate all goods homogenously ($\frac{\partial m_1}{\partial q}(q) \equiv 0$) and the second case is when consumers value newest goods more ($\frac{\partial m_1}{\partial q}(q) > 0$ for all $q \in [0, Q(t)]$).

Chapter 2

Optimal Control of Heterogeneous Systems with Endogenous Domain of Heterogeneity

2.1 Introduction

Generally speaking, in our terminology *heterogeneous control systems* are represented by a family of controlled ODEs parameterized by a parameter q varying in a measurable space $\mathscr{Q}$, called *domain of heterogeneity*. The family is coupled by "aggregated states" (involving integration over $\mathscr{Q}$) that enter in all ODEs and in the respective initial conditions. For each $q \in \mathscr{Q}$ the respective ODE has its own time-interval $[\theta(q), \bar{\theta}(q)]$ in which it matters, at the beginning of which an initial condition is posed. A basic theory of optimal control problems for such heterogeneous systems is presented in (Veliov, 2008). Applications in many areas are indicated in this study, including problems in population dynamics, epidemiology, air pollution control, and several problems in economics. In (Veliov, 2008) the initial time $\theta(q)$ for the q-th ODE is exogenously given.

In order to explain what are the motivations for the present research, let us consider a particular economic context originating from (Jones, 1976; Dixit & Stiglitz, 1977) in which the parameter q is interpreted as an available technology (or consumption good) and $\theta(q)$ is the time at which the q-th technology emerges. Then, in the spirit of the endogenous economic growth literature , e.g. (Romer, 1990; G. M. Grossman & Helpman, 1991b; Jones, 1995a), $\theta(q)$ depends on the rate of technological advancement which depends, in its turn, on the investments in R&D. That is, the time, $\theta(q)$, of appearance of the q-th technology is endogenous. Such a situation, where the domain of heterogeneity at each time is endogenously determined, is not covered by the results in (Veliov, 2008) and by the existing literature. In the endogenous growth literature involving a variety of technologies/products it is assumed that the products are indistinguishable, therefore the amount of physical capital allocated to production of each of them is equal (Barro & Martin, 2004). This reduces the originally distributed optimal control problem to a lumped one.

In the present research we address the issue of endogenous domain of heterogeneity. The class of problems that we consider is relatively simple and does not contain many of the economic applications. Such will be presented in follow-up publications in economics-oriented journals. This is because here we want to stress the mathematical challenges that the endogenous domain of heterogeneity brings in the optimal control context. The main trouble is caused by the fact that the objective value considered as a function of the control is, in general, non-differentiable (in any reasonable space setting). This effect does not arise in standard optimal control problems with smooth data if the set of admissible controls is a subset of L_∞. As a result of this intrinsic non-differentiability, the necessary optimality condition of Pontryagin's type takes a non-standard form in which the adjoint systems is represented by a differential inclusion (rather than equation) although the data are assumed smooth with respect to the state variables. However, this situation may happen only if the optimal control is "irregular" with respect to the parameter of heterogeneity q, which is hard to exclude a priori.

The "irregularity" of the optimal control that requires the abovementioned non-standard form of the maximum principle does not happen in the economic application we have in mind, thus it is perhaps mainly of "academic interest". If the optimal control is "regular" enough, the optimality conditions take a form corresponding to the heuristic application of the Lagrange principle. For example, it is possible to prove that the optimal control in the more specific problem, considered in (Skritek, Tsachev, & Veliov, 2014), satisfies the regularity condition discussed above. As a consequence of this the optimality conditions of Pontryagins type, which are good enough for analytic and numerical investigation can be obtained.

However, the trouble with non-differentiability of the objective function still remains. It creates certain difficulties in the derivation of the optimality condition in the form of global maximum principle, therefore we present it below in detail. The possible non-differentiability also requires a special care about the numerical approaches to the problem based on gradient-type methods.

The chapter is organized as follows. In Section 2.2 we formulate the problem and the assumptions. In Section 2.3 we derive a necessary optimality condition without any *a priori* assumptions for the optimal control. In this optimality condition the adjoint variable satisfies a differential inclusion and the maximization of the Hamiltonian takes the form of “min-max”. From here, under certain regularity of the optimal control we derive also a Pontryagin-type maximum principle. Section 2.4 presents a formula for the derivative of the objective function with respect to the control from L_∞ and a sufficient condition for its existence. Moreover, a version of the gradient projection method in the control space is briefly described, which at each iteration involves only controls with respect to which the objective function is differentiable. In Section 2.5 we give two examples of non-differentiability which justify the special treatment in the preceding two sections. Section 2.6 gives a stylized economic example. In a simple case we obtain the solution analytically and show that the existence issue is complicated: an optimal solution exists for some configurations of the parameters and fails to exist for others. Also, numerical results are presented and interpreted. One longer proof is shifted to Appendix.

2.2 Formulation of the problem

Let $[0,T]$ be a fixed time-interval and let $[0,\bar{Q}]$ be an interval where the parameter of heterogeneity, q, will take values (here $\bar{Q} > 0$ could be $+\infty$, in which case the interval should be interpreted as $[0,\infty)$). Denote $D = [0,T] \times [0,\bar{Q}]$. The state variables in the model below will be the functions[1]

$$x : D \mapsto \mathbf{R}^n, \quad Q : [0,T] \mapsto [0,\bar{Q}], \quad y : [0,T] \mapsto \mathbf{R}^m,$$

while $u : D \mapsto U \subset \mathbf{R}^r$ will be a control function. The optimal control problem we consider reads as follows:

$$\max_u \int_0^T \int_0^{Q(t)} L(t,q,x(t,q),Q(t),y(t),u(t,q)) \,\mathrm{d}q\,\mathrm{d}t, \tag{2.1}$$

subject to the equations

$$\dot{Q}(t) \quad = \quad g(t,Q(t),y(t)), \quad Q(0) = Q^0 \geq 0, \;\; t \in [0,T], \tag{2.2}$$

[1] In economics q, for instance, can denote technology, Q the technological frontier, and y can be some additional externality.

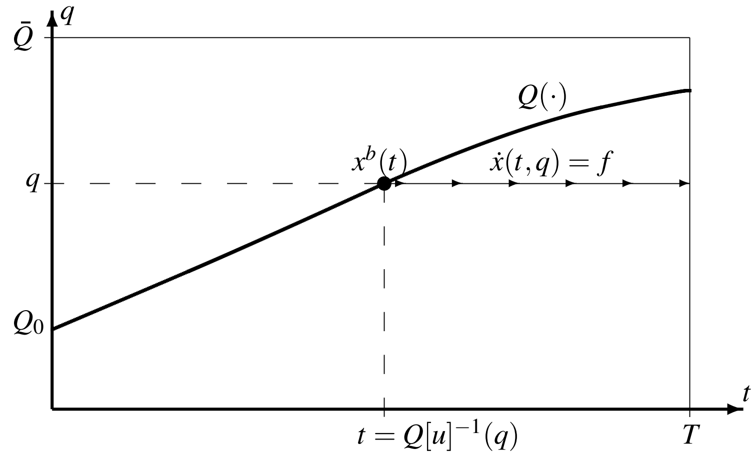


Figure 2.1: Increasing domain of heterogeneity

$$y(t) = \int_0^{Q(t)} h(t,q,u(t,q))\,\mathrm{d}q, \tag{2.3}$$

$$\dot{x}(t,q) = f(t,q,x(t,q),Q(t),y(t),u(t,q)), \tag{2.4}$$

$$x(0,q) = x^0(q), \;\; q \in [0,Q^0],$$

$$x(t,Q(t)) = x^b(t), \;\; t \in (0,T],$$

$$u(t,q) \in U. \tag{2.5}$$

Here

$$L : D \times \mathbf{R}^n \times [0,\bar{Q}] \times \mathbf{R}^m \times U \mapsto \mathbf{R},$$

$$f : D \times \mathbf{R}^n \times [0,\bar{Q}] \times \mathbf{R}^m \times U \mapsto \mathbf{R}^n, \quad g : D \times \mathbf{R}^m \mapsto \mathbf{R}, \quad h : D \times U \mapsto \mathbf{R}^m,$$

$\dot{x}(t,q)$ is the derivative with respect to t.

The informal meaning is as follows, see Figure 2.1. Given a control function u with values in U, equations (2.2) and (2.3) define the interval $[0,Q(t)]$ in which the parameter q takes values at time t. The state $y(t)$ represents an aggregated (over the domain of heterogeneity $[0,Q(t)]$) quantity. Equation (2.4) with the respective side conditions defines the distributed state x. Then the objective functional (2.1) is to be maximized with respect to the control u.

Before giving the formal definition of the problem we enlist *Standing Assumptions* (i) – (vi) which will hold throughout the chapter:

(i) The set $U \subset \mathbf{R}^r$ is compact.

(ii) The functions L, f, g, h are measurable in (t,q) and continuous in the rest

of the variables, locally essentially bounded, differentiable in (x,Q,y), with locally Lipschitz partial derivatives, uniformly with respect to $u \in U$ and $(t,q) \in D$. The function h is locally Lipschitz continuous in u uniformly with respect to $(t,q) \in D$.
(iii) $g(t,Q,y) \geq \alpha_0 > 0$ for every $(t,Q) \in [0,T] \times [Q^0, \bar{Q}]$ and every $y = \int_0^Q h(t,q,u)\,\mathrm{d}q$ for some $u \in U$.
(iv) $x^b : [0,T] \mapsto \mathbf{R}^n$ is continuously differentiable, $x^0 : [0,Q^0] \mapsto \mathbf{R}^n$ is measurable and bounded.

Denote $\mathscr{U} = \{u \in L_\infty(D) : u(t,q) \in U\}$. Since for any given $u \in \mathscr{U}$ one can represent

$$g(t,Q,y(t)) = g\left(t,Q,\int_0^Q h(t,q,u(t,q))\,\mathrm{d}q\right)$$

and the function in the right-hand side is Lipschitz in Q, equation (2.2) has locally a solution $Q = Q[u]$ and it is unique on its maximal interval of existence in $[0,T]$.

Standing Assumption (v): For every $u \in \mathscr{U}$ the solution $Q[u]$ exists in $[0,\bar{Q}]$ on the whole interval $[0,T]$.

Given $u \in \mathscr{U}$, we define for $q \in [0,\bar{Q}]$

$$\theta[u](q) = \begin{cases} 0 & \text{if } q \in [0,Q^0], \\ Q[u]^{-1}(q) & \text{if } q \in (Q^0, Q[u](T)), \\ T & \text{if } q \in [Q[u](T), \bar{Q}]. \end{cases} \tag{2.6}$$

The definition is correct, since $Q[u]$ is invertible according to Assumption (iii) and its image is $[Q^0, Q[u](T)]$, see Figure 2.1. Notice that $\theta[u]$ is Lipschitz continuous with constant $1/\alpha_0$ due to Assumption (iii). We extend the definition of x^b by setting

$$x^b(t,q) = \begin{cases} x^0(q) & \text{if } q \in [0,Q^0], \\ x^b(t) & \text{if } q \in (Q^0, \bar{Q}]. \end{cases} \tag{2.7}$$

Then we may view (2.4) as a family of ODEs (one for each $q \in [0,\bar{Q}]$), where the functions $y = y[u]$ and $Q = Q[u]$ are already defined from (2.2), (2.3) as described above. For each such q the solution $x[u]$ of (2.4) with the additional condition $x(\theta[u](q),q) = x^b(\theta[u](q),q)$ locally exists and is unique on its maximal interval of existence in $[\theta[u](q),T]$. We extend $x[u]$ for $q \in (Q^0,\bar{Q}]$ and $t \in [0,\theta[u](q))$ as $x[u](t,q) = x^b(t)$.

Standing Assumption (vi): For every $u \in \mathscr{U}$ and for almost every $q \in [0,\bar{Q}]$ the solution $x[u](\cdot,q)$ exists on $[0,T]$.

Due to the continuous dependence of the solution of the ODE (2.4) on the parameter q and on the initial data, and due to the measurability of a superposition of a measurable and a continuous function, $x[u]$ is measurable with respect to q. Then the meaning of the optimization problem (2.1) is clear.

We mention that several of the Standing Assumptions could be relaxed, however at a certain price. In some extensions this price is just a technical complication, as for example considering a function h in (2.3) depending also on x, or requiring only continuity (instead on Lipschitz continuity) with respect to u in Assumption (ii). Other extensions require a substantial additional analysis. For example, dependence of the side condition $x^b(t)$ on y or on an additional non-distributed control $v(t)$. A third class of extensions require conceptual clarification. This concerns mainly Assumption (iii). How the solution should be defined if $Q(t)$ is not strictly monotone increasing? Apparently the "right" definition depends on the particular application. All these extensions have clear interpretations in several economic contexts and is investigated for a particular economic models in forthcoming papers, see e.g. (Skritek et al., 2014). In this study, however, we focus on the mathematical complication that the endogenous heterogeneity brings already in the simplest case which is general enough to cover some applications (see Section 2.6 for an example).

2.3 The optimality conditions

In this section we derive necessary optimality conditions of Pontryagin's type for the problem (2.1)–(2.5). What makes this derivation not a routine work, is that the objective functional (2.1) is, in general, non-differentiable with respect to the control function, as it will be demonstrated in Section 2.5. Moreover, the form of the maximum principle a la Pontryagin is non-standard, in general, although under an additional (non-restrictive for the typical applications) condition it takes a form that could be heuristically derived by an appropriate application of the Lagrange multiplier rule. The problem of existence of an optimal solution is not systematically investigated in this chapter, although it is also challenging, as we show in Section 2.6.

To make the expressions below more compact we interpret x, y and u as columns, in contrast to the adjoint variables λ and ν that will be involved later, which are viewed as row-vectors with corresponding dimensions.

We start with a variational analysis the result of which will be summarized in Proposition 1 below. It will be used in the proofs of the three theorems to follow in this and in the next section.

Let $u \in \mathscr{U}$ be fixed and let $u^\sigma \in \mathscr{U}$ be a sequence of controls parameterized by positive $\sigma \to 0$. We shall denote by x, Q and y the values of the state variables corresponding to the control u and by x^σ, Q^σ and y^σ – the values of the state variables corresponding to the control u^σ. Further, we denote $\Delta u = u^\sigma - u$, $\Delta Q = Q^\sigma - Q$, etc.

For the sequence of controls u^σ we require that there exists a constant c such that for all sufficiently small σ

$$\|\Delta u\|_{L_1(D)} + \|\Delta y\|_{L_1(0,T)} + \|\Delta Q\|_{C(0,T)} + \max_{t\in[0,T]} \|\Delta x(t,\cdot)\|_{L_1(0,\bar{Q})} \le c\sigma, \quad (2.8)$$

$$\int_0^T \|\Delta u(t,\cdot)\|_{L_\infty(0,\bar{Q})} \, \mathrm{d}t + \|\Delta y\|_{L_\infty(0,T)} + \max_{t\in[0,T]} \|\Delta x(t,\cdot)\|_{L_\infty(0,\bar{Q})} \le c\sqrt{\sigma}, \quad (2.9)$$

where $\|\cdot\|_{C(0,T)}$ denotes the *max*-norm of the space of the continuous functions on $[0,T]$ and $\|\cdot\|_{L_p}$ is the usual L_p-norm for $p = 1$ or $p = \infty$.

There are two essential cases of sequences u^σ for which the above requirements are satisfied:

Case 1: L_1(-simple needle)-variation, where

$$u^\sigma(t,q) = \begin{cases} v & \text{if } (t,q) \in [\tau, \tau + \sqrt{\sigma}] \times [\kappa, \kappa + \sqrt{\sigma}], \\ u(t,q) & \text{elsewhere} \end{cases}$$

and $\tau \in [0,T)$, $\kappa \in [0,\bar{Q})$ and $v \in U$ are arbitrarily fixed.

Case 2: L_∞-variation, where

$$u^\sigma = u + \sigma v$$

and $v \in L_\infty(D)$ is such that $u^\sigma(t,q) \in \mathscr{U}$ for all sufficiently small σ.

Using assumptions (i), (ii), (iii) and (v) it can be easily verified that in both cases requirements (2.8) and (2.9) are fulfilled.

Let us denote by $J(v)$ the value of the objective function (2.1) corresponding to $v \in \mathscr{U}$. By a similar analysis as in the usual proof of the Pontryagin maximum principle for ODE control systems with unconstrained state we obtain the following result. For every function $\lambda : D \mapsto \mathbf{R}^n$ which is absolutely continuous in t for a.e. $q \in [0,\bar{Q}]$, with $\dot{\lambda} \in L_\infty(D)$ and $\lambda(T,q) = 0$, for every absolutely continuous function $\mu : [0,T] \mapsto \mathbf{R}$ satisfying $\mu(T) = 0$, and for every $v \in L_\infty(0,T)$ the following variational representation holds (the proof is not straightforward and will be presented in Appendix):

$$\begin{aligned} \Delta J &= \int_0^T \int_0^{Q(t)} [L_x + \dot{\lambda} + \lambda f_x] \Delta x \, \mathrm{d}q \, \mathrm{d}t \\ &+ \int_0^T \Bigg[\dot{\mu} + \mu g_Q + \int_0^{Q(t)} (L_Q + \lambda f_Q) \, \mathrm{d}q \end{aligned}$$

$$
\begin{aligned}
& + \frac{1}{\Delta Q(t)} \int_{Q(t)}^{Q^\sigma(t)} [L + \lambda(f - \dot{x}^b) + \nu h] \, \mathrm{d}q \Bigg] \Delta Q(t) \, \mathrm{d}t \\
+ & \int_0^T \left[-\nu + \mu g_y + \int_0^{Q(t)} (L_y + \lambda f_y) \, \mathrm{d}q \right] \Delta y \, \mathrm{d}t \\
+ & \int_0^T \int_0^{Q(t)} [\Delta_u L + \lambda \Delta_u f + \nu \Delta_u h] \, \mathrm{d}q \, \mathrm{d}t + o(\sigma). \qquad (2.10)
\end{aligned}
$$

In the case $\Delta Q(t) = 0$ the term $\frac{1}{\Delta Q(t)} \int_{Q(t)}^{Q^\sigma(t)}$ may be defined as any real number. Here the arguments $(t,q,x(t,q),Q(t),y(t),u(t,q))$ are skipped, $\Delta_u L = L(u^\sigma(t,q)) - L$ (similarly for the other terms $\Delta_u \ldots$), and as usual $o(\sigma)/\sigma \to 0$. This notational convention will be systematically used below: the arguments of the functions will appear only if they are different from those mentioned above, or in order to avoid confusion, or for more clarity.

Let λ be defined as the solution of the *adjoint equation*

$$
-\dot{\lambda}(t,q) = L_x(t,q) + \lambda(t,q) f_x(t,q), \quad \lambda(T,q) = 0. \qquad (2.11)
$$

Then the first term in the right-hand side of (2.10) disappears. This equation has a unique solution on $[0,T]$ in a similar sense as (2.4): for every $q \in [0,\bar{Q}]$ it is a linear ODE with end condition $\lambda(T,q) = 0$. Clearly, $\dot{\lambda} \in L_\infty(D)$, as required.

Given the control $u \in \mathscr{U}$ and the corresponding solution of (2.2)–(2.4) and adjoint function λ, let us define for $t \in [0,T]$ and $\mu \in \mathbf{R}$ the set

$$
\begin{aligned}
\Gamma(t,\mu) \quad = \quad & \operatorname*{Limsup}_{\alpha \to 0, \alpha \neq 0} \frac{1}{\alpha} \int_{Q(t)}^{Q(t)+\alpha} \Big[L(t,q) + \lambda(t,q)(f(t,q) - \dot{x}^b(t)) \\
+ \quad & (\mu g_y(t) + \eta(t)) h(t,q) \Big] \, \mathrm{d}q, \qquad (2.12)
\end{aligned}
$$

where

$$
\eta(t) = \int_0^{Q(t)} (L_y(t,q) + \lambda(t,q) f_y(t,q)) \, \mathrm{d}q \qquad (2.13)
$$

and $\operatorname*{Limsup}_{\alpha \to 0, \alpha \neq 0} G(\alpha)$ is the Kuratowski upper limit of a function G at $\alpha = 0$, consisting of all condensation points of sequences $G(\alpha_k)$ with $\alpha_k \to 0$, $\alpha_k \neq 0$. Thanks to the continuity of the right-hand side in (2.12) with respect to $\alpha > 0$ it is easy to prove that $\Gamma(t,\mu)$ is a compact interval. For the same reason it is easy to prove that $\Gamma : [0,T] \times \mathbf{R} \Rightarrow \mathbf{R}$ is measurable in t and Lipschitz in μ (theorems 8.2.8 and 8.2.5 in (Aubin & Frankowska, 1990) are used in the proof of the measurability). The Lipschitz continuity holds due to $\|g_y h\|_{L_\infty(D)} < \infty$, which is a Lipschitz constant of Γ with respect to μ. Then the differential inclusion

$$-\dot{\mu}(t) \in \mu(t)g_Q(t) + \xi(t) + \Gamma(t,\mu(t)), \quad \mu(T) = 0 \tag{2.14}$$

with

$$\xi(t) := \int_0^{Q(t)} [L_Q(t,q) + \lambda(t,q) f_Q(t,q)]\,\mathrm{d}q \tag{2.15}$$

has at least one solution (see e.g. (Aubin, 1991, Theorem 11.7.1)), therefore its reachable set $R(t)$ is nonempty for every $t \in [0,T]$.

Let us define the measurable functions

$$s^\sigma(t) = \begin{cases} 1 & \text{if } Q^\sigma(t) = Q(t), \\ 0 & \text{if } Q^\sigma(t) \neq Q(t), \end{cases}$$

and

$$d^\sigma(t) = \Delta Q(t) + \sigma^2 s^\sigma(t).$$

Now consider the equation

$$\begin{aligned} -\dot{\tilde{\mu}}^\sigma(t) &= \tilde{\mu}^\sigma(t) g_Q(t) + \xi(t) \\ &+ \frac{1}{d^\sigma(t)} \int_{Q(t)}^{Q(t)+d^\sigma(t)} [L + \lambda(f - \dot{x}^b(t)) + (\tilde{\mu}^\sigma(t) g_y(t) + \eta(t))h]\,\mathrm{d}q, \\ \tilde{\mu}^\sigma(T) &= 0. \end{aligned} \tag{2.16}$$

Notice that $d^\sigma(t) \neq 0$ for any $\sigma > 0$ and t. Since the right-hand side of (2.16) is linear in $\tilde{\mu}^\sigma$ uniformly in t and measurable in t, it has a solution $\tilde{\mu}^\sigma(t)$. Then denoting

$$\tilde{v}^\sigma(t) = \tilde{\mu}^\sigma(t) g_y(t) + \eta(t) \tag{2.17}$$

and inserting $\lambda, \tilde{\mu}^\sigma$, $\tilde{v}^\sigma$ in (2.10) we obtain

$$\Delta J = \int_0^T \int_0^{Q(t)} [\Delta_u L + \lambda \Delta_u f + \tilde{v}^\sigma \Delta_u h]\,\mathrm{d}q\,\mathrm{d}t + o(\sigma). \tag{2.18}$$

We have

$$\operatorname{dist}\left(\frac{1}{d^\sigma(t)} \int_{Q(t)}^{Q(t)+d^\sigma(t)} [L + \lambda(f - \dot{x}^b(t)) + (\tilde{\mu}^\sigma(t) g_y(t) + \eta(t))h]\,\mathrm{d}q, \right.$$
$$\left. \Gamma(t, \tilde{\mu}^\sigma(t)) \right) := \beta^\sigma(t) \underset{\sigma\to 0}{\longrightarrow} 0$$

for every t, where β^σ is bounded in σ and t. This easily follows from the continuity with respect to μ of the mappings involved, the uniform boundedness of $\tilde{\mu}^\sigma$ and the definition of Γ. By the Filippov theorem (Aubin & Frankowska, 1990, Theorem 10.4.1) we obtain that there exists a solution $\mu^\sigma(t)$ of (2.14) such that

$$|\mu^\sigma(t) - \tilde{\mu}^\sigma(t)| \le C \int_0^T \beta^\sigma(t)\,\mathrm{d}t \underset{\sigma \to 0}{\to} 0.$$

Then using (2.8), (2.17), (2.18), and the Lipschitz continuity of h with respect to u (cf. *Standing Assumption* (ii)) we obtain that

$$\Delta J = \int_0^T \int_0^{Q(t)} [\Delta_u L + \lambda \Delta_u f + \nu^\sigma \Delta_u h]\,\mathrm{d}q\,\mathrm{d}t + o(\sigma), \tag{2.19}$$

where

$$\nu^\sigma(t) = \mu^\sigma(t) g_y(t) + \eta(t). \tag{2.20}$$

Let us summarize what we have obtained so far.

Proposition 1 *Let $u \in \mathscr{U}$ be arbitrarily fixed and let $u^\sigma \in \mathscr{U}$ be a sequence of controls satisfying (2.8), (2.9). Then equation (2.19) holds true with the adjoint function $\lambda \in L_\infty(D)$ defined in (2.11) and some solution μ^σ of (2.14) and the corresponding ν^σ defined by (2.20). Here λ and μ^σ are absolutely continuous in t with $\dot{\lambda} \in L_\infty(D)$, $\dot{\mu}^\sigma \in L_\infty[0,T]$, η and ξ are defined by (2.13) and (2.15), respectively.*

We shall make use of the above finding in two ways. First, in this section, we shall consider the needle variation defined in Case 1 in order to obtain a specific global maximum principle for the problem in consideration, then in Section 2.4 we shall use variations as in Case 2 to express the derivative of the objective functional with respect to the control under additional conditions that ensure its existence.

Define $\tilde{H} : D \times \mathbf{R}^{n+1+m+r+n+m} \mapsto \mathbf{R}$ as

$$\tilde{H}(t,q,x,Q,y,u,\lambda,\nu) = L(t,q,x,Q,y,u) + \lambda\, f(t,q,x,Q,y,u) + \nu\, h(t,q,u).$$

Theorem 1 *Let $u \in L_\infty(D)$ be an optimal control in the problem (2.1)–(2.5) and let $z := (x,Q,y)$ be the corresponding trajectory. Let λ be the solution of the adjoint equation (2.11) (it exists and is unique on D) and let η, ξ and Γ be defined by (2.13), (2.15) and (2.12), respectively. Then the reachable set $R(t)$ of the differential inclusion (2.14) is nonempty, and for almost every $(t,q) \in D(u) := \{(t,q) : t \in [0,T), q \in [0,Q(t))\}$*

$$\max_v \min_\nu \left(\tilde{H}(t,q,z(t,q),v,\lambda(t,q),\nu) - \tilde{H}(t,q,z(t,q),u(t,q),\lambda(t,q),\nu)\right) \le 0, \tag{2.21}$$

subject to $v \in U$ and $\nu \in R(t) g_y(t) + \eta(t)$.

Proof We shall sketch the proof emphasizing only the points that are unusual. Assume that the claim of the theorem is not true. Denote by $\varphi(t,q,v)$ the function "min…" in (2.21). This function is measurable in (t,q), according to (Aubin & Frankowska, 1990, Theorem 8.2.11). In a routine way we obtain that there exists a set $\Omega \subset D(u)$ with positive measure, $\varepsilon > 0$ and $v \in U$ such that $\varphi(t,q,v) \geq \varepsilon$ for $(t,q) \in \Omega$. Let (τ,κ) be a Lebesgue point of Ω, that is,

$$\text{meas}(\Omega \cap B^\sigma) = \sigma + o(\sigma), \tag{2.22}$$

where B^σ is the box $[\tau, \tau + \sqrt{\sigma}] \times [\kappa, \kappa + \sqrt{\sigma}]$. Let u^σ be the simple needle control variation defined in Case 1, and let $(\lambda, \mu^\sigma, \nu^\sigma)$ be the functions from Proposition 1. Then according to (2.19)

$$\Delta J = \int_{B^\sigma} [\tilde{H}(t,q,v,\nu^\sigma(t)) - \tilde{H}(t,q,\nu^\sigma(t))]\, \mathrm{d}q\, \mathrm{d}t + o(\sigma),$$

where, consistently with our notational convention:

$$\begin{aligned} \tilde{H}(t,q,v,\nu) &= \tilde{H}(t,q,z(t,q),v,\lambda(t,q),\nu), \\ \tilde{H}(t,q,\nu) &= \tilde{H}(t,q,z(t,q),u(t,q),\lambda(t,q),\nu). \end{aligned}$$

Then taking into account that $\nu^\sigma(t) \in R(t)g_y(t) + \eta(t)$ we obtain from the inequality $\varphi(t,q,v) \geq \varepsilon$ and (2.22) that

$$\Delta J \geq \int_{B^\sigma \cap \Omega} \varepsilon\, \mathrm{d}(q,t) + \int_{B^\sigma \setminus \Omega} \ldots\, \mathrm{d}(q,t) + o(\sigma) = \varepsilon\sigma + o(\sigma).$$

Since for sufficiently small $\sigma > 0$ the right-hand side is strictly positive we come to a contradiction. □

The above theorem gives information about the optimal control only for $q \in [0, Q(t)]$. Obviously the values of u for $q > Q(t)$ are irrelevant for the objective value.

In the applications we have in mind the optimal control is regular enough to reduce the inclusion in (2.14) to an equation. Below we elaborate on this case, starting with

Assumption (vii): The functions L, L_x, f, f_x and h are continuous with respect to q (uniformly in the rest of the variables); the optimal control u is (equivalent to) a function which is continuous from the left with respect to q at $q = Q(t)$ for a.e. $t \in [0,T]$.

Since, as just mentioned, the values of u for $q > Q(t)$ are irrelevant for the objective function (2.1), we may redefine the optimal u as $u(t,q) = u(t,Q(t))$ for $q > Q(t)$. Moreover, because of assumption (vii), the functions $x(t,\cdot)$ and

$\lambda(t,\cdot)$ are continuous in q at $q = Q(t)$, thus $x(t,Q(t))$ and $\lambda(t,Q(t))$ are well defined. Then the set $\Gamma(t,\mu)$ is a singleton:

$$\Gamma(t,\mu) = L(t,Q(t)) + \lambda(t,Q(t))(f(t,Q(t)) - \dot{x}^b(t)) + (\mu g_y(t) + \eta(t))h(t,Q(t))$$

and (2.14) becomes an equation. Let $\mu(\cdot)$ be its solution and $\nu(\cdot)$ be obtained from $\mu(\cdot)$ by (2.20). Then $\mu(\cdot)$ and $\nu(\cdot)$ solve the equations

$$\begin{aligned} -\dot{\mu}(t) &= \mu(t)g_Q(t) + L(t,Q(t)) \\ &\quad + \lambda(t,Q(t))(f(t,Q(t)) - \dot{x}^b(t)) + \nu(t)h(t,Q(t)) \\ &\quad + \int_0^{Q(t)} [L_Q(t,q) + \lambda(t,q)f_Q(t,q)]\,\mathrm{d}q, \quad \mu(T) = 0, \quad (2.23) \\ \nu(t) &= \mu(t)g_y(t) + \int_0^{Q(t)} [L_y(t,q) + \lambda(t,q)f_y(t,q)]\,\mathrm{d}q. \quad (2.24) \end{aligned}$$

Notice that according to the notational convention and the side condition $x(t,Q(t)) = x^b(t)$ we have $L(t,Q(t)) = L(t,Q(t),x^b(t),Q(t),y(t),u(t,Q(t)))$ and similarly for the other functions above.

Now we introduce the function (having in many respects the traditional meaning of "hamiltonian") $H : D \times \mathbf{R}^{n+1+m+r+n+1+m} \mapsto \mathbf{R}$ as

$$\begin{aligned} H(t,q,x,Q,y,u,\lambda,\mu,\nu) &= L(t,q,x,Q,y,u) + \lambda\, f(t,q,x,Q,y,u) \\ &\quad + \mu\, g(t,Q,y) + \nu\, h(t,q,u) \end{aligned}$$

where the arguments in the left-hand side should be inserted in the right-hand side wherever appropriate. Then Theorem 1 implies that there exists a unique solution of adjoint equations, as states in the following theorem.

Theorem 2 *Under assumptions (i) – (vii) denote by $u \in L_\infty(D)$ an optimal control in the problem (2.1)–(2.5) and by $z := (x,Q,y)$ the corresponding trajectory. Then the adjoint system (2.11), (2.23)–(2.24) has a unique solution $\pi := (\lambda,\mu,\nu)$ and for a.e. $t \in [0,T]$ and a.e. $q \in [0,Q(t)]$*

$$H(t,q,z(t,q),u(t,q),\pi(t,q)) = \max_{u \in U} H(t,q,z(t,q),u,\pi(t,q)).$$

In the next section, we elaborate on numerical procedure of finding an optimal control.

2.4 A numerical approach

We include this section in the chapter for the following reason. It turns out (as shown in the next section by two examples) that the objective functional $J(u)$ is not differentiable in $u \in L_\infty(D)$, in general. Even more intriguing, although the value of $J(u)$ does not depend on the values of $u(t,q)$ for $q > Q[u](t)$, the differentiability property may depend on how $u(t,q)$ is defined above the graph of Q. This fact requires a special analysis which is the main point in this section.

For solving numerically problem (2.1)–(2.5) we apply a version of the gradient projection method in the control space $L_\infty(D)$. Denoting as before by $J(u)$ the objective value corresponding to control $u \in \mathscr{U}$, the gradient projection method consists of the following. Denote by $J'(u)$ the derivative of J with respect to u in $L_\infty(D)$ (if it exists). Given the current "approximation" $u_k \in \mathscr{U}$ of the optimal control, assume that $J'(u_k)$ can be represented by a function from $L_\infty(D)$ (this will be proved in Theorem 3 below, see equation (2.25)). We find a next approximation u_{k+1} as

$$u_{k+1} = \mathscr{P}_{\mathscr{U}}(u_k + \alpha_k J'(u_k)).$$

where $\mathscr{P}_X$ means the metric projection on X. Clearly

$$u_{k+1}(t,q) = \mathscr{P}_U\left(u_k(t,q) + \alpha_k J'(u_k)(t,q)\right).$$

The choice of the parameter $\alpha_k \geq 0$ is a subject of many publications on gradient methods and it is beyond the scope of current discussion.

Denote by $\mathscr{U}_{\mathrm{lc}}$ the set of those $u \in \mathscr{U}$ that are continuous from the left at $q = Q[u](t)$ for a.e. $t \in [0,T]$ and by $\mathscr{U}_{\mathrm{c}}$ the set of those $u \in \mathscr{U}$ that are continuous at $Q[u](t)$ for a.e. $t \in [0,T]$. First of all, in addition to assumptions (i)–(vii) we introduce the following one.

Assumption (viii). The functions L, f, h are continuously differentiable with respect to u, uniformly with respect to the rest of the variables.

Theorem 3 *Under assumptions (i)–(viii) the objective function $J(u)$ is Gateaux-differentiable at every $u \in \mathscr{U}_{\mathrm{c}}$ and the derivative $J'(u)$ can be represented by the function*

$$J'(u) = \begin{cases} H_u(t,q), & \text{if} \quad q \leq Q(t), \\ 0 \qquad , & \text{if} \quad q > Q(t), \end{cases} \tag{2.25}$$

where (λ, μ, ν) is the solution of the adjoint system (2.11), (2.23)–(2.24).

Proof Let u^σ be defined as in Case 2: $u^\sigma = u + \sigma v$, in Section 2.3, where $v \in L_\infty(D)$. Then from (2.19) and the fact that now v^σ does not depend on σ we obtain

$$\begin{aligned} \Delta J &= \int_0^T \int_0^{Q(t)} [H(t,q,u(t,q)+\sigma v(t,q)) - H(t,q,u(t,q))] \,\mathrm{d}q\,\mathrm{d}t + o(\sigma) \\ &= \sigma \int_0^T \int_0^{Q(t)} H_u(t,q) v(t,q) \,\mathrm{d}q\,\mathrm{d}t + o(\sigma), \end{aligned}$$

due to the uniform continuous differentiability of H in u, which means that the functional J is Gateaux differentiable at u in the space $L_\infty(D)$ and its derivative is (represented by) (2.25). □

In order to ensure that $J'(u)$ stays continuous from the left, we shall modify the gradient projection procedure in such a way that $u_k \in \mathscr{U}_\mathrm{c}$. Indeed, assume that $u_k \in \mathscr{U}_\mathrm{c}$. Then it is easy to verify (using the continuous dependence of the solution of an ODE on parameters) that H_u is also continuous in q at $q = Q(t)$ for a.e. t. Then $J'(u)$ is continuous from the left at $q = Q(t)$ due to (2.25). Since $\mathscr{P}_U$ is a continuous mapping, we obtain that $\tilde{u}_{k+1}(t,q) := \mathscr{P}_U\,(u_k(t,q) + \alpha_k J'(u_k)(t,q))$ has the same property. Define the operator $\mathscr{I} : \mathscr{U}_\mathrm{lc} \to \mathscr{U}_\mathrm{c}$ as

$$\mathscr{I}(u)(t,q) = \begin{cases} u(t,q) & , \quad \text{for} \quad q \le Q[u](t), \\ u(t,Q[u](t)), & \quad \text{for} \quad q > Q[u](t). \end{cases}$$

Notice that $J(\mathscr{I}(u)) = J(u)$, but J is differentiable at $\mathscr{I}(u)$, while it need not be differentiable at u. Then we define the next iteration as

$$u_{k+1}(t,q) = \mathscr{I}\left(\mathscr{P}_U\left(u_k(t,q) + \alpha_k J'(u_k)(t,q)\right)\right).$$

The numerical implementation involves discretization with respect to t and q which contains delicate technical points due to the changing domain $[0, Q(t)]$. The author wrote in MATLAB a program tackling these points, using adaptive discretization with respect to q. The detailed description of the program is outside the scope of the present study. Some results of calculations by the program are presented in Section 2.6 for an economic example.

2.5 Two examples of non-differentiability of the objective function

As mentioned in the previous section the objective value $J(u)$ in (2.1) considered as a function of the control could be non-differentiable in the space $L_\infty(D)$.

We distinguish two different cases of non-differentiability: one is harmless, while the other is not and requires the non-standard form of the maximum principle considered in the previous section.

1. Let $u \in \mathscr{U}$ and let $Q(t) = Q[u](t)$ be the corresponding solution of (2.2), (2.3). As argued in Section 2.2, the respective solution $(x[u], Q[u], y[u])$ is independent of the values of u for $q > Q(t)$, and if u is continuous from the left with respect to q at $q = Q(t)$ for a.e. t, then u can be redefined to become continuous in q for $q = Q(t)$, e.g. as $u(t,q) = u(t, Q(t))$ for $q > Q(t)$ (we use the notation $u^\#$ for the redefined control). Then J is Gateaux differentiable at $u^\#$. This is the "harmless" case of possible non-differentiability of $J(u)$, which can be avoided by the redefinition of u for $q > Q(t)$.

Although the objective value $J(u)$ does not depend on the values of u for $q > Q(t)$, its differentiability does, and requires that u be continuous in q also from the right at $q = Q(t)$ (which means that the redefinition of u as $u^\#$ is essential). The example below shows this.

Example 1. Consider the system

$$\begin{aligned} \dot{Q}(t) &= y(t), \qquad Q(0) = 1. \\ y(t) &= \int_0^{Q(t)} u(t,q)\,\mathrm{d}q \end{aligned}$$

and the functional

$$J(u) = \int_0^1 \int_0^{Q(t)} 1\,\mathrm{d}q\,\mathrm{d}t = \int_0^1 Q(t)\,\mathrm{d}t.$$

This is not interpreted as an optimal control problem. We just study the differentiability of the functional J, defined for $u \in L_\infty([0,1] \times [0,\infty))$.

Fix the control

$$u(t,q) = \begin{cases} 1, & \text{if} \quad q \in [0, e^t], \\ a, & \text{if} \quad q > e^t, \end{cases}$$

where a is a real number. Then consider the controls $u_1^h(t,q) = u(t,q) - h$ and $u_2^h(t,q) = u(t,q) + h$, $h > 0$, and the corresponding solutions (Q_i^h, y_i^h), and compare with the solution for u, which is obviously $Q(t) = e^t$. It can also be directly checked that for h close to zero we have

$$Q_1^h(t) = e^{(1-h)t} = Q(t) - hte^t + O(h^2).$$

The function $F(t,Q) = \int_0^Q u(t,q)\,\mathrm{d}q$ is Lipschitz in Q and measurable in t, and $Q(t)$ and $Q_2^h(t)$ satisfy the equations

$$\dot{Q}(t) = F(t, Q(t)), \qquad \dot{Q}_2^h(t) = F(t, Q_2^h(t)) + hQ_2^h(t), \quad Q(0) = Q_2^h(0) = 1.$$

By a standard comparison argument (or as a consequence of the viability theory, (Aubin, 1991, Chapter 11)) this implies $Q_2^h(t) \geq Q(t)$. Hence

$$\begin{aligned} y_2^h(t) &= \int_0^{Q(t)} u_2^h(t,q)\,\mathrm{d}q + \int_{Q(t)}^{Q_2^h(t)} u_2^h(t,q)\,\mathrm{d}q \\ &= (1+h)Q(t) + (a+h)(Q_2^h(t) - Q(t)) \\ &= (a+h)Q_2^h(t) + (1-a)Q(t). \end{aligned}$$

Then the equation for Q_2^h becomes

$$\dot{Q}_2^h(t) = (a+h)Q_2^h(t) + (1-a)Q(t), \qquad Q_2^h(0) = 1.$$

Using the Cauchy formula we obtain after routine calculations that

$$\begin{aligned} Q_2^h(t) &= e^t\left(1 + \frac{h}{1-a-h}\right) - \frac{h}{1-a-h}e^{(a+h)t} \\ &= Q(t) + \frac{h}{1-a-h}\left(e^t - e^{at}\right) - \frac{h}{1-a-h}hte^{at} + O(h^2). \end{aligned}$$

Now we consider the case $a \neq 1$ (hence u is discontinuous from the right). Then the second last term in the above equality is also $O(h^2)$, thus for h close to zero we have

$$Q_2^h(t) = Q(t) + \frac{h}{1-a}\left(e^t - e^{at}\right) + O(h^2).$$

Then

$$J(u-h) - J(u) = -h\int_0^1 te^t\,\mathrm{d}t + O(h^2),$$

while

$$J(u+h) - J(u) = \frac{h}{1-a}\int_0^1 \left(e^t - e^{at}\right)\mathrm{d}t + O(h^2).$$

The last two equalities show that J is not differentiable if $a \neq 1$.

In contrast, if $a = 1$ the objective J is differentiable in the direction ± 1 (as it was proved in the previous section) since in this case

$$Q_2^h(t) = Q(t) + hte^t + O(h^2).$$

The above example shows, in particular, that the special attention that we attribute to the definition of u above $Q[u](t)$ in the previous section is essential.

2. Below we give another example, which shows that the "remedy" of redefinition of u above $Q(t)$ in order to get differentiability of $J(u)$ is not applicable if u is too "bad" below $Q(t)$. This example justifies the non-standard form of the maximum principle in the case of an optimal control that is not continuous from the left at $Q(t)$.

Example 2. Consider the system

$$\begin{aligned} \dot{Q}(t) &= y(t), \qquad Q(0)=1, \qquad t\in[0,1] && (2.26)\\ y(t) &= \int_0^{Q(t)} u(t,q)\,\mathrm{d}q, && (2.27) \end{aligned}$$

and the functional

$$J(u) = \int_0^1 \int_0^{Q(t)} u(t,q)\,\mathrm{d}q\,\mathrm{d}t.$$

As in Example 1, this is not interpreted as an optimal control problem. We just study the differentiability of the functional J, defined for $u\in L_\infty([0,1]\times[0,\infty))$. We shall prove that for some u the functional J is not even directionally differentiable at u in the direction of

$$w(t,q) = \begin{cases} 0 & \text{for } (t,q)\in[0,\bar{t}]\times[0,\infty)\\ -1 & \text{for } (t,q)\in(\bar{t},1]\times[0,\infty), \end{cases}$$

where $\bar{t}\in(0,1)$ is to be chosen later sufficiently close to 1. Namely, we define

$$u(t,q)\equiv u(q) = \begin{cases} v(1-q) & \text{for } \; q\in[0,1)\\ 0 & \text{for } \; q\geq 1 \end{cases}$$

where $v(q)$ is defined on $(0,1]$ as $v(1)=-1$ and

$$v(q) = \begin{cases} 1 & \text{for } q\in\left[\frac{1}{3^k},\frac{2}{3^k}\right) \quad k=1,2,\ldots,\\ -1 & \text{for } q\in\left[\frac{2}{3^k},\frac{3}{3^k}\right) \quad k=1,2,\ldots. \end{cases}$$

The solution of (2.26), (2.27) for u is $Q(t)=1$, $y(t)=0$, since

$$y(t) = \int_0^1 v(1-q)\,\mathrm{d}q = \int_0^1 v(q)\,\mathrm{d}q = 0.$$

Now, let $u^\sigma = u+\sigma w$, where σ is a "small" positive real number. Denote by Q^σ and y^σ the corresponding solution of (2.26), (2.27). It is clear that $u^\sigma(t,q)\equiv u(t,q)$ on $[0,\bar{t}]\times[0,\infty)$ and, hence, $Q^\sigma(t)\equiv Q(t)\equiv 1$ on $[0,\bar{t}]$. Is is also clear that $Q^\sigma(t)\leq Q(t)$ for $t\in[\bar{t},1]$.

Denoting $V(q)=\displaystyle\int_0^q v(s)\,\mathrm{d}s$ for $q\in(-\infty,\infty)$, we have $0\leq V(q)\leq \frac{1}{2}q$ for $q\in[0,1]$ and also $V\left(\frac{3}{3^k}\right)=0$ and $V\left(\frac{2}{3^k}\right)=\frac{1}{2}\cdot\frac{2}{3^k}$ for $k=1,2,3,\ldots$.

For $t \in [\bar{t}, 1]$ we have $\dot{Q}^\sigma(t) = \int_0^{Q^\sigma(t)} [u(q) - \sigma] \,\mathrm{d}q = -\sigma + \int_1^{Q^\sigma(t)} [u(q) - \sigma] \,\mathrm{d}q$.
Defining $\varphi_\sigma(t) := Q^\sigma(t) - (1 - \sigma(t - \bar{t}))$ for $t \in [\bar{t}, 1]$, we have $\varphi_\sigma(\bar{t}) = 0$ and

$$\begin{aligned} \dot{\varphi}_\sigma(t) &= \int_1^{1-\sigma(t-\bar{t})+\varphi_\sigma(t)} [u(q) - \sigma] \,\mathrm{d}q \\ &= -\int_0^{\sigma(t-\bar{t})-\varphi_\sigma(t)} [v(s) - \sigma] \,\mathrm{d}s \\ &= -V[\sigma(t - \bar{t}) - \varphi_\sigma(t)] + \sigma[\sigma(t - \bar{t}) - \varphi_\sigma(t)]. \end{aligned}$$

<u>Case 1.</u> Assume $\varphi_\sigma(\tau) > 0$ for $\tau \in [t_0, t] \subset [\bar{t}, 1]$ where $\varphi_\sigma(t_0) = 0$. Then

$$0 < \varphi_\sigma(\tau) \leq \int_{t_0}^{\tau} \sigma^2(\mu - \bar{t}) \,\mathrm{d}\mu \leq \frac{1}{2}\sigma^2(\tau - \bar{t})^2 = O(\sigma^2) \quad \text{for} \quad \tau \in [t_0, t], \tag{2.28}$$

i.e.

$$Q^\sigma(\tau) = 1 - \sigma(\tau - \bar{t}) + O(\sigma^2) \quad \text{for} \quad \tau \in [t_0, t]. \tag{2.29}$$

<u>Case 2.</u> Assume $\varphi_\sigma(\tau) < 0$ for $\tau \in [t_0, t] \subset [\bar{t}, 1]$ where $\varphi_\sigma(t_0) = 0$. Then

$$V(q) \leq \frac{1}{2} q$$

and

$$\varphi_\sigma(\tau) = \int_{t_0}^{\tau} [-V(\sigma(\mu - \bar{t}) - \varphi_\sigma(\mu)) + \sigma(\sigma(\mu - \bar{t}) - \varphi_\sigma(\mu))] \,\mathrm{d}\mu$$

yield

$$0 < -\varphi_\sigma(\tau) \leq \left(\frac{1}{2}\sigma - \sigma^2\right) \frac{1}{2}(\tau - \bar{t})^2 + \int_{t_0}^{\tau} (-\varphi_\sigma(\mu)) \,\mathrm{d}\mu.$$

Applying the Gronwall inequality (cf., e.g., (Corduneanu, 1971), p. 14) we obtain

$$0 < -\varphi_\sigma(\tau) \leq \left(\frac{1}{2}\sigma - \sigma^2\right) \frac{1}{2}(\tau - \bar{t})^2 e^{t - t_0} \leq \frac{1}{4}\sigma(\tau - \bar{t})^2 e^{1-\bar{t}} \quad \text{for} \quad \tau \in [t_0, t].$$

This implies

$$Q^\sigma(\tau) \geq 1 - \sigma(\tau - \bar{t}) - \frac{1}{4}\sigma(\tau - \bar{t})^2 e^{1-\bar{t}} \quad \text{for} \quad \tau \in [t_0, t]. \tag{2.30}$$

We have

$$\begin{aligned}
J(u^\sigma) - J(u) &= \int_0^1 \int_0^{Q^\sigma(t)} [u(q) + \sigma w(t,q)]\,\mathrm{d}q\,\mathrm{d}t - \int_0^1 \int_0^1 u(q)\,\mathrm{d}q\,\mathrm{d}t \\
&= -\sigma(1-\bar{t}) - \int_{\bar{t}}^1 \int_{1-\sigma(t-\bar{t})}^1 u(q)\,\mathrm{d}q\,\mathrm{d}t \\
&\quad + \int_{\bar{t}}^1 \int_{1-\sigma(t-\bar{t})}^{Q^\sigma(t)} u(q)\,\mathrm{d}q\,\mathrm{d}t + \int_{\bar{t}}^1 \int_{Q^\sigma(t)}^1 \sigma\,\mathrm{d}q\,\mathrm{d}t.
\end{aligned}$$

From (2.29) and (2.30) we obtain that the last term above is $O(\sigma^2)$. We next estimate the third term in the last row above. From (2.28) and (2.30) we obtain that if the positive σ is small enough,

$$|Q^\sigma(t) - [1-\sigma(t-\bar{t})]| \le \frac{e^{1-\bar{t}}}{4}(t-\bar{t})^2\,\sigma \quad \text{holds true for every} \quad t \in [1-\bar{t}, 1].$$

Hence,

$$\left| \int_{\bar{t}}^1 \int_{1-\sigma(t-\bar{t})}^{Q^\sigma(t)} u(q)\,\mathrm{d}q\,\mathrm{d}t \right| \le \sigma \frac{e^{1-\bar{t}}}{4} \int_{\bar{t}}^1 (t-\bar{t})^2\,\mathrm{d}\tau = \sigma \frac{e^{1-\bar{t}}}{12}(1-\bar{t})^3. \qquad (2.31)$$

We thus obtained that

$$J(u^\sigma) - J(u) = -\sigma(1-\bar{t}) - Z^\sigma + Y^\sigma + O(\sigma^2).$$

where

$$Z^\sigma = \int_{\bar{t}}^{1} \int_{1-\sigma(t-\bar{t})}^{1} u(q)\,\mathrm{d}q\,\mathrm{d}t \quad \text{and} \quad Y^\sigma = \int_{\bar{t}}^{1} \int_{1-\sigma(t-\bar{t})}^{Q^\sigma(t)} u(q)\,\mathrm{d}q\,\mathrm{d}t.$$

In order to prove non-differentiability of J at u in the direction of w it is enough to show that $(Z^\sigma - Y^\sigma)/\sigma$ does not converge with $\sigma \to 0$. We have

$$\begin{aligned}
\frac{Z^\sigma}{\sigma} &= \frac{1}{\sigma} \int_{\bar{t}}^{1} \int_{1-\sigma(t-\bar{t})}^{1} v(1-q)\,\mathrm{d}q\,\mathrm{d}t = \frac{1}{\sigma} \int_{\bar{t}}^{1} \int_{0}^{\sigma(t-\bar{t})} v(q)\,\mathrm{d}q\,\mathrm{d}t \\
&= \frac{(1-\bar{t})^2}{\sigma(1-\bar{t})} \int_0^{\sigma(1-\bar{t})} v(q)\,\mathrm{d}q - \frac{(1-\bar{t})^2}{(\sigma(1-\bar{t}))^2} \int_0^{\sigma(1-\bar{t})} q\,v(q)\,\mathrm{d}q. \qquad (2.32)
\end{aligned}$$

We shall consider the last two integrals separately, for values $\sigma_k' = \frac{2/3^k}{1-\bar{t}}$ and $\sigma_k'' = \frac{3/3^k}{1-\bar{t}}$ of $k = 1, 2, \ldots$. Easy calculations give for the first integral in (2.32)

$$\frac{1}{\sigma_k'(1-\bar{t})} \int\limits_0^{\sigma_k'(1-\bar{t})} v(q)\,\mathrm{d}q - \frac{1}{\sigma_k''(1-\bar{t})} \int\limits_0^{\sigma_k''(1-\bar{t})} v(q)\,\mathrm{d}q = \frac{1}{2} - 0 = \frac{1}{2}.$$

More cumbersome calculations, which we skip, give for the second integral in (2.32)

$$\frac{1}{(\sigma_k'(1-\bar{t}))^2} \int\limits_0^{\sigma_k'(1-\bar{t})} q\,v(q)\,\mathrm{d}q - \frac{1}{(\sigma_k''(1-\bar{t}))^2} \int\limits_0^{\sigma_k''(1-\bar{t})} q\,v(q)\,\mathrm{d}q = \frac{11}{32} + \frac{1}{8} = \frac{15}{32}.$$

Comparing the expressions for the two integrals in (2.32) we conclude that the variation of $\frac{Z^\sigma}{\sigma}$ remains strictly positive (at least $(1-\bar{t})^2/32$) for arbitrarily small σ. On the other hand, from (2.31) we obtain that if $\bar{t} \in (0,1)$ is close enough to 1, $\frac{|Y^\sigma|}{\sigma} \leq (1-\bar{t})^2/64$ holds true. Hence $(Z^\sigma - Y^\sigma)/\sigma$ does not converge with $\sigma \to 0$. This completes the proof of the non-differentiability of J at u in the direction w.

The example, discussed above, can be easily modified to include the case of strictly increasing $Q(\cdot)$ by replacing the equation in (2.26) by $\dot{Q}(t) = 1 + y(t)$. Since the proof of the non-differentiability of J in this case requires longer and more cumbersome calculations, we omit it.

2.6 An economic example

In this section we present a stylized economic model of endogenous economic growth to which the above results can be applied. This economic model is a particular case of a model considered in (Skritek et al., 2014), for which regularity condition for the optimal control is fulfilled.

We consider a finite time horizon $[0,T]$ (presumably rather large, so that T is an "approximation" of the infinity) and a large corporation producing at time t diverse goods labeled by the real number $q \in [0, Q(t)]$. Here $Q(t)$ is the newest good (technology) available at time t. Each of the goods q is produced by a separate firm that at time t has physical capital stock $x(t,q)$. The q-th firm ($q \in [0, Q(t)]$) invests at time t an amount $u(t,q)$ that is split in two parts: $\alpha u(t,q)$, $0 \leq \alpha \leq 1$, is allocated to increase the capital stock, while $(1-\alpha)u(t,q)$ is the contribution of the q-th firm to the R&D activity of the corporation which results in development of new technologies (goods) and hence in increase of $Q(t)$.

The model reads as follows:

$$\begin{aligned}
\dot{x}(t,q) &= -\delta x(t,q)+\alpha u(t,q), \quad x(0,q)=x^0(q) \ \text{ for } \ q\in[0,Q^0], \\
&\qquad\qquad\qquad\qquad\qquad\quad x(t,Q(t))=0 \ \text{ for } \ t>0, \\
\dot{Q}(t) &= (1-\alpha)y(t), \quad Q(0)=Q^0, \\
y(t) &= \int_0^{Q(t)} u(t,q)\,\mathrm{d}q.
\end{aligned}$$

Here $y(t)$ is the total investment in R&D, $\delta \geq 0$ is the depreciation rate of the physical capital, $x^0(q)$ is the initially available capital stock for producing goods $q\in[0,Q^0]$, $Q^0>0$ is the newest technology available at time $t=0$. For all technologies obtained at $t>0$, initial capital stock is assumed to be zero, $x(t,Q(t))=0$. The objective function to be maximized is

$$\int_0^T e^{-rt}\int_0^{Q(t)} \left[p(q,Q(t))x(t,q)-bu(t,q)-cu^2(t,q)\right]\,\mathrm{d}q\,\mathrm{d}t, \tag{2.33}$$

subject to the control constraints $u(t,q)\geq 0$ and $u(t,q)\leq 1$. In the model $r\geq 0$ is the discount rate, $p(q,Q)$ is the market price of the good $q\in[0,Q]$, given that goods up to level Q are available, $bu+cu^2$, $b\geq 0$, $c>0$, is the quadratic cost of investments u. The dependence of the price p on q and Q reflects the fact that the market price of any available good decreases when newer products emerge (that is, when Q increases). For the present illustrative purpose we chose the specification

$$p(q,Q)=e^{-\gamma(Q-q)}$$

with $\gamma\geq 0$. So the price of the newest product is normalized to one (which is supported by the data for personal computers and mobile telephones, where the price of the new products does not substantially change with time, while the quality increases).

Because the integrand of the objective functional (2.33) is strictly concave in u, one may expect that optimal control would be always bounded even without the upper constraint $u(t,q)\leq 1$. We will give an example, where this is not the case. The reason is that the upper limit $Q(t)$ of the integral also depends on u, and at particular parameters the unique optimal solution hits the upper bound for sufficiently high time horizon T, see Figure 2.3.

Let us write optimality conditions. As proved in (Skritek et al., 2014), there exists a unique optimal control $u(t,q)$ that is continuous from the left at

$q = Q[u](t)$. Adjoint equations (2.11), (2.23)-(2.24) for this example, written in the terms of the "current value" adjoint variables $\tilde{\lambda} = e^{rt}\lambda$, $\tilde{\mu} = e^{rt}\mu$, $\tilde{\nu} = e^{rt}\nu$ read as follows

$$\begin{aligned} -\dot{\tilde{\lambda}}(t,q) &= -(\delta+r)\tilde{\lambda}(t,q) + e^{-\gamma(Q(t)-q)}, \quad \tilde{\lambda}(T,q) = 0, & (2.34) \\ -\dot{\tilde{\mu}}(t) &= -r\tilde{\mu}(t) - [bu(t,Q(t)) + cu^2(t,Q(t))] \\ &\quad +\alpha\tilde{\lambda}(t,Q(t))u(t,Q(t)) + \tilde{\nu}(t)u(t,Q(t)) \\ &\quad -\gamma\int_0^{Q(t)} e^{-\gamma(Q(t)-q)}x(t,q)\,\mathrm{d}q, \quad \tilde{\mu}(T) = 0, & (2.35) \\ \tilde{\nu}(t) &= \tilde{\mu}(t)(1-\alpha), & (2.36) \end{aligned}$$

and Theorem 2 implies that

$$u(t,q) = \min\left\{1, \max\left\{0, \frac{\alpha\tilde{\lambda}(t,q) + \tilde{\nu}(t) - b}{2c}\right\}\right\}. \qquad (2.37)$$

In the following we solve the optimal control problem for different parameter sets. In simple cases we obtain exact analytic solutions, in more complicated cases we derive asymptotic solutions and resort to numerics, as described in Section 2.4.

1. Let us consider first the case $\gamma = 0$, where the solution can be studied analytically. It follows from (2.37) and (2.34) that the optimal control does not depend on q, so $\tilde{\lambda}(t,q) = \tilde{\lambda}(t)$ and $u(t,q) = u(t)$. We will solve equations (2.34)–(2.37) backwards starting from time T. It follows from $\tilde{\mu}(T) = 0$, $\tilde{\lambda}(T) = 0$, and (2.37) that $u(T) = 0$. Let $[t_0, T]$ be the maximal interval which ends at T and in which $u(t) = 0$. To find t_0 we substitute $u(t) = 0$ in (2.35) and obtain (using that $\gamma = 0$) the solution $\tilde{\mu}(t) = \tilde{\nu}(t) = 0$ for $t \in [t_0, T]$. Now we get from (2.37) that $\tilde{\lambda}(t_0) = b/\alpha$ and substitute here the solution

$$\tilde{\lambda}(t) = \frac{1 - e^{-(r+\delta)(T-t)}}{r+\delta}$$

of equation (2.34). Thus, we obtain

$$t_0 = T + \frac{1}{\delta+r}\ln\left(1 - b\,\frac{\delta+r}{\alpha}\right), \quad \text{when} \quad b\,\frac{\delta+r}{\alpha} < 1.$$

It the case $b(\delta+r)/\alpha \geq 1$ the optimal control is identically zero.

In a time interval, where the control constraint is not active ($u(t) > 0$) the derivative of the hamiltonian must be equal to zero:

$$\tilde{H}_u(t,q) = -b - 2cu(t,q) + \alpha\tilde{\lambda}(t,q) + \tilde{\nu}(t) = 0.$$

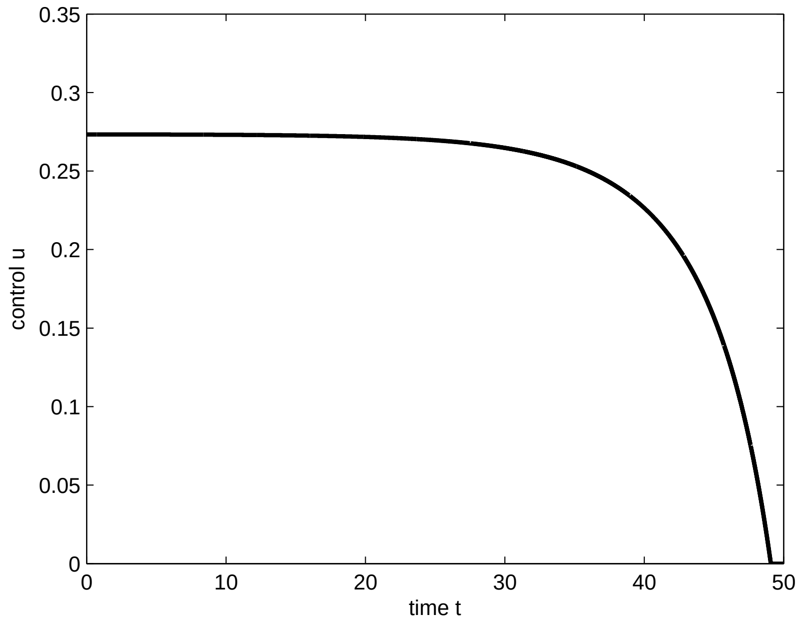


Figure 2.2: Optimal investments for $\gamma = 0$, $\delta = 0$, $r = 0.3$, $T = 50$, $b = 0.4$, $c = 3$, $\alpha = 0.5$, $A > 0$.

Using this condition and equations (2.34)–(2.36) we can derive the following Riccati equation for the optimal control for $t \in [0, t_0]$

$$\begin{aligned} \dot{u}(t) &= ru(t) - \frac{1-\alpha}{2}u(t)^2 - \frac{r}{2c}\left(\frac{\alpha}{\delta + r} - b\right) - \frac{\alpha\delta}{2c(\delta + r)}e^{-(\delta+r)(T-t)}, \\ u(t_0) &= 0, \end{aligned} \tag{2.38}$$

which turns out to be positive on $[0, t_0)$. Notice, that if solution $u(t)$ reaches the upper bound, $u(t_1) = 1$, at first time instance $t_1 > 0$ to the left from t_0, then optimal control $u(t) = 1$ for all $t \in [0, t_1)$.

1.1. The solution of equation (2.38) can be obtained in special functions. Under the additional condition $\delta = 0$ it takes the simpler form

$$u(t) = \frac{r}{1-\alpha} - \frac{\sqrt{A}}{c(1-\alpha)} \tanh\left(\frac{\sqrt{A}}{2c}(t_0 - t) + \operatorname{arctanh}\left(\frac{rc}{\sqrt{A}}\right)\right),$$

see Figure 2.2, if

$$A := c\left(r^2 c - (1-\alpha)(\alpha - br)\right) > 0.$$

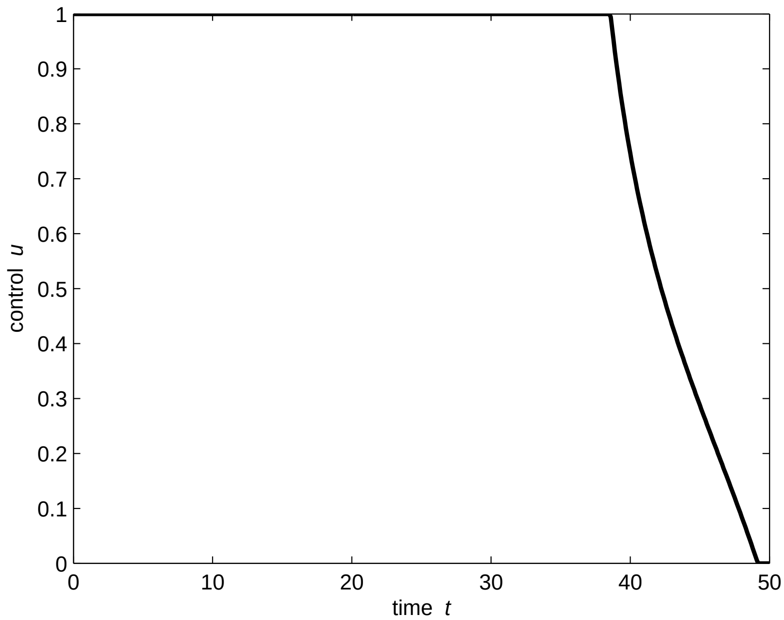


Figure 2.3: Optimal investments hit its upper bound, for $\gamma = 0$, $\delta = 0$, $r = 0.1$, $T = 50$, $b = 0.4$, $c = 3$, $\alpha = 0.5$, $A < 0$.

If $A < 0$ the solution is

$$u(t) = \frac{r}{1-\alpha} + \frac{\sqrt{-A}}{c(1-\alpha)} \tan\left(\frac{\sqrt{-A}}{2c}(t_0 - t) - \arctan\left(\frac{rc}{\sqrt{-A}}\right)\right),$$

which tends to infinity, if $t_0 - t$ may take sufficiently large values when $t \in [0, t_0]$. This is always the case if T is sufficiently large, as it follows from the expression for t_0. Then the optimal control certainly hits its upper bound, $u(t) = 1$ for $t \in [0, t_1)$, see Figure 2.3.

Thus, in order to ensure boundedness of the solution on any time horizon even without upper control constraint, like in Figure 2.2, the discount rate should be big enough:

$$r > \frac{1}{2c}\left(\sqrt{b^2(1-\alpha)^2 + 4c\alpha(1-\alpha)} - b(1-\alpha)\right). \tag{2.39}$$

1.2. In the case $\delta > 0$ we apply an asymptotic analysis assuming that T is large. Namely, we can find the asymptotic solution of (2.38) considering times

t such that $t_0 - t$ is big enough for neglecting the last term in (2.38). Then the two steady state solutions are

$$\left.\begin{array}{rcl} u_1 & = & \dfrac{\sqrt{r}}{1-\alpha}\left[\sqrt{r}+\sqrt{r-\dfrac{1-\alpha}{c}\left(\dfrac{\alpha}{\delta+r}-b\right)}\right], \\ u_2 & = & \dfrac{\sqrt{r}}{1-\alpha}\left[\sqrt{r}-\sqrt{r-\dfrac{1-\alpha}{c}\left(\dfrac{\alpha}{\delta+r}-b\right)}\right] \end{array}\right. . \tag{2.40}$$

The first of which is an attractor while the second is a repeller. In inverse time $t' = -t$ point u_2 becomes an attractor with the basin of attraction $(u_1, -\infty)$. Thus, u_2 is the horizontal asymptote of the exact solution for $t \to -\infty$ if it is still in the basin $u(t) \in (u_1, -\infty)$ when the last term in (2.38) is already insignificant, see Figure 2.4 (top). For a large time horizon $[0, T]$ if the optimal control $u(t)$ exists then most of the time $t \in [0, t_0]$ it is close to u_2 when $u_2 > 0$, or $u(t) \equiv 0$ when $u_2 \leq 0$. Nonnegativity of the expression under the square root in (2.40) gives us condition for existence of asymptote u_2

$$r > \frac{1}{2c}\left(\sqrt{(b(1-\alpha)-\delta c)^2 + 4c\alpha(1-\alpha)} - b(1-\alpha) - \delta c\right) \tag{2.41}$$

that generalizes condition (2.39) to all $\delta \geq 0$. It follows from (2.40) and (2.41) that when $\delta > \alpha/b$ even $r = 0$ allows for bounded solution but this solution is trivial $u(t, q) \equiv 0$.

The essence of the above analysis is that even for a very simple economic problem the issue of existence of a solution is not simple. To ensure existence we introduce here upper constraint $u(t) \leq 1$, that has not much of economic sense. While inequality (2.41) provides a necessary condition under which a solution exists for any time horizon $[0, T]$ without upper constraint.

2. The above considerations concern the "degenerate" case $\gamma = 0$, where all data of the problem are independent of q. In the general case $\gamma \geq 0$ we have obtained only numerical results.[2]

Figure 2.5 presents the optimal investments (top) and the corresponding capital stock (bottom) of the firm with $\gamma = 0.7$. The time horizon is $T = 50$ and the initial product variety is $[0, Q^0] = [0, 1]$. Notice that due to the technological development ($Q(t)$ increases from 1 to about 2) the firm completely abandons investments in some older technologies. At time $t = 30$, for example, the firm invests only in technologies $q \in [0.6, 1.63]$. As a result, the physical capital

[2] The numerical results are obtained by our own MATLAB solver in which the modification of the gradient projection method described in Section 2.4 is implemented.

of technologies q much smaller than the technological frontier $Q(t)$ is close to zero.[3] Every section $x(\cdot,q)$ has the meaning of *diffusion curve* of technology q. It starts at the time $\theta(q)$ (see (2.6)) and reaches its maximum at some later time, after which it begins decreasing.

3. In the above numerical example the dependence of the optimal control on the technology q results from the dependence of the price function $p(q,Q)$ on q. However, a minor nonlinearity inserted in the problem may lead to technology-dependent investments even if all data of the problem are technology-independent. Let, for example, the revenue function $p(q,Q)x$ in (2.33) be replaced with $p\,(x+x_m)^\sigma$ with $\sigma \in (0,1)$ and constant $p>0$ and $x_m>0$. This corresponds to the revenue of a firm with market power or scarce non-capital production factors. The numerically obtained optimal control is plotted on Figure 2.6. Clearly, the investments in the newest technologies are larger than those of the older ones due to the higher marginal productivity of the technologies with lower capital stock. But the nature of the higher marginal productivity of new technologies is the nonlinearity of the production function rather than heterogeneity of market prises.

[3] A more profound investigation of the issue of obsolescence in the spirit of (Boucekkine, Río, & Licandro, 2005) are yet to be done elsewhere.

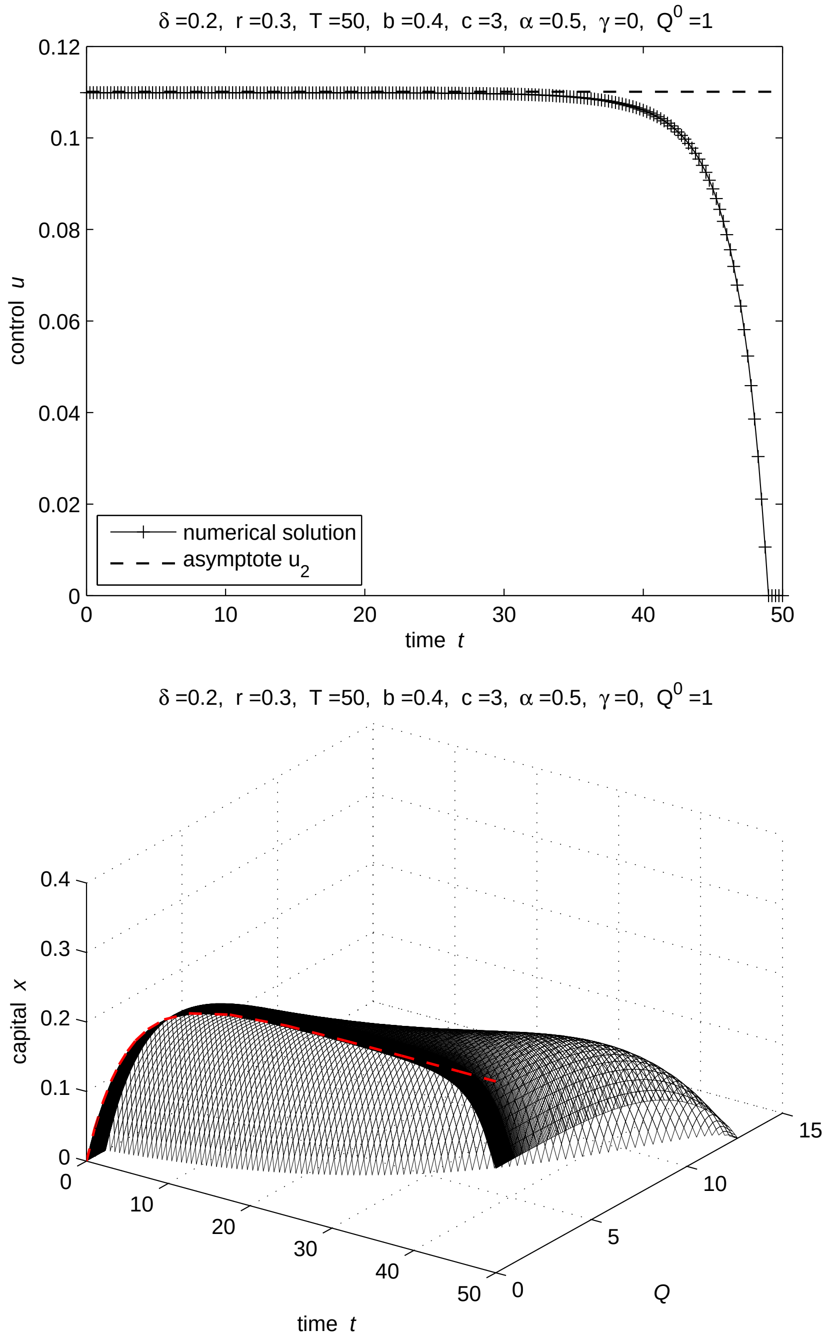


Figure 2.4: Optimal investments (top) and the corresponding capital stock (bottom) for $\gamma = 0$. Numerical results are compared with asymptotic solution (dashed lines).

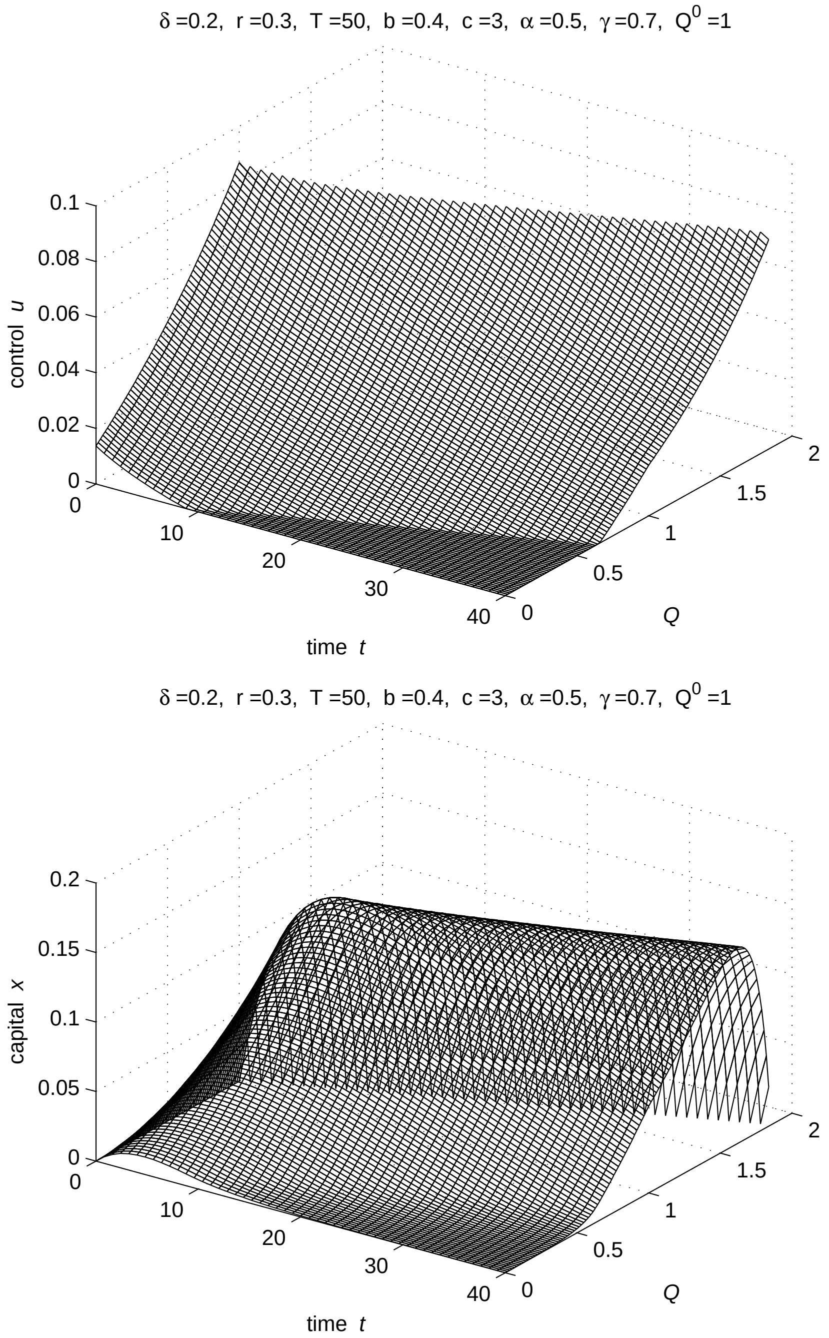
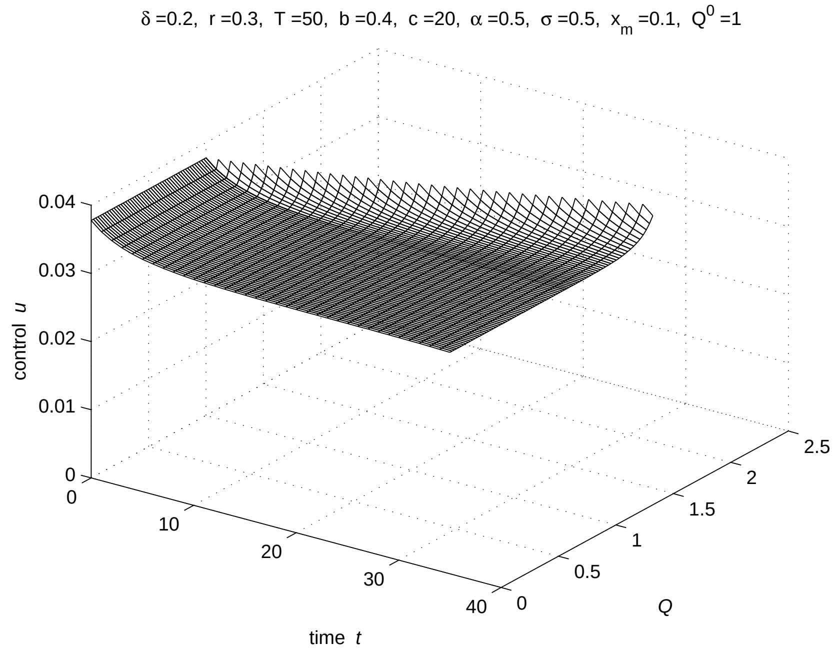


Figure 2.5: Optimal investments (top) and the corresponding capital stock (bottom) for $\gamma = 0.7$



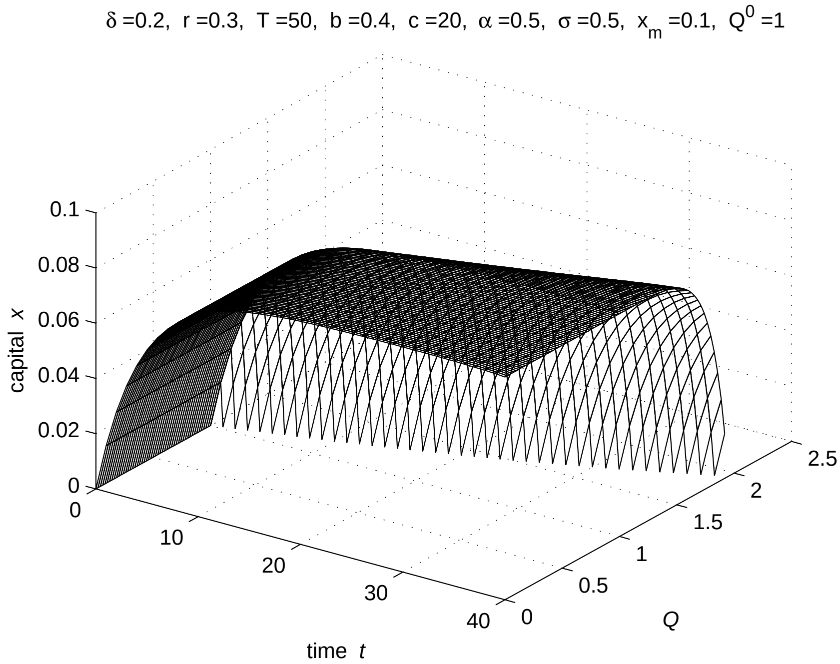


Figure 2.6: Optimal investments (top) and the corresponding capital stock (bottom) for nonlinear production function $\sqrt{x+0.1}$

2.7 Conclusion

Optimal control of heterogeneous systems, that is, families of controlled ODEs parameterized by a parameter running over a growing domain of heterogeneity is considered. The main novelty, added in this system, is that the domain of heterogeneity is endogenous: it may depend on the control and on the state of the system. This extension is crucial for several economic applications and turns out to rise interesting mathematical problems. A necessary optimality condition is derived, where one of the adjoint variables satisfies a differential inclusion (instead of equation) and the maximization of the Hamiltonian takes the form of "min-max". As a consequence, a Pontryagin-type maximum principle is obtained under certain regularity conditions for the optimal control. A formula for the derivative of the objective function with respect to the control from L_∞ is presented together with a sufficient condition for its existence. A stylized economic example of R&D driven economic growth of a corporation consisting of continuum of firms is given. The economic model is investigated both analytically and numerically. Sufficient conditions for the optimal solution being bounded even without upper constraint were derived for homogeneous prices. The effect of obsolescence of old technologies is demonstrated on diffusion curves in the case of heterogeneous prices, when goods of new design cost more than those of old ones. We also show that a minor nonlinearity inserted in the problem may lead to technology-dependent (heterogeneous) investments even if all data of the problem are technology-independent.

Appendix

Here we prove the variational representation (2.10) under the conditions of Section 2.3 and with the notational convention made there. Consider

$$\begin{aligned}
\Delta J \quad := \quad & J(u^\sigma) - J(u) = \int_0^T \int_0^{Q^\sigma(t)} L(t,q,x^\sigma(t,q),Q^\sigma(t),y^\sigma(t),u^\sigma(t,q))\,\mathrm{d}q\,\mathrm{d}t \\
& - \int_0^T \int_0^{Q(t)} L(t,q,x(t,q),Q(t),y(t),u(t,q))\,\mathrm{d}q\,\mathrm{d}t \\
= \quad & \int_0^T \int_0^{Q^\sigma(t)} [L + L_x \Delta x + L_Q \Delta Q + L_y \Delta y + \Delta_u L]\,\mathrm{d}q\,\mathrm{d}t \\
& - \int_0^T \int_0^{Q(t)} L\,\mathrm{d}q\,\mathrm{d}t + o(\sigma).
\end{aligned}$$

The same convention is systematically used below. The above equality follows in a standard way from Assumption (ii) and (2.8), (2.9)[4]. Using again (2.8) and (2.9), we obtain

$$\Delta J = \int_0^T \int_{Q(t)}^{Q^\sigma(t)} L \,\mathrm{d}q\,\mathrm{d}t + \int_0^T \int_0^{Q(t)} [L_x \Delta x + L_Q \Delta Q + L_y \Delta y + \Delta_u L] \,\mathrm{d}q\,\mathrm{d}t + o(\sigma). \tag{2.42}$$

Let $\lambda : D \mapsto \mathbf{R}^n$ be absolutely continuous in t for a.e. $q \in [0, \bar{Q}]$, $\dot{\lambda} \in L_\infty(D)$ and $\lambda(T,q) = 0$, $q \in [0, \bar{Q}]$. We remind that λ is considered as a row-vector.

Consider the value

$$\begin{aligned}
&\int_0^T \int_0^{Q^\sigma(t)} \lambda(t,q)[\dot{x}^\sigma(t,q) - \dot{x}^b(t)] \,\mathrm{d}q\,\mathrm{d}t \\
&- \int_0^T \int_0^{Q(t)} \lambda(t,q)[\dot{x}(t,q) - \dot{x}^b(t)] \,\mathrm{d}q\,\mathrm{d}t \\
= &\int_0^T \int_0^{Q^\sigma(t)} \lambda(t,q)[\dot{x}^\sigma(t,q) - \dot{x}^b(t,q)] \,\mathrm{d}q\,\mathrm{d}t \\
&- \int_0^T \int_0^{Q(t)} \lambda(t,q)[\dot{x}(t,q) - \dot{x}^b(t,q)] \,\mathrm{d}q\,\mathrm{d}t \\
&= \int_0^{Q^\sigma(T)} \int_{\theta^\sigma(q)}^T \lambda[\dot{x}^\sigma - \dot{x}^b] \,\mathrm{d}t\,\mathrm{d}q \\
&- \int_0^{Q(T)} \int_{\theta(q)}^T \lambda[\dot{x} - \dot{x}^b] \,\mathrm{d}t\,\mathrm{d}q \\
= &- \int_0^{Q^\sigma(T)} \int_{\theta^\sigma(q)}^T \dot{\lambda}[x^\sigma - x^b] \,\mathrm{d}t\,\mathrm{d}q \\
&+ \int_0^{Q(T)} \int_{\theta(q)}^T \dot{\lambda}[x - x^b] \,\mathrm{d}t\,\mathrm{d}q,
\end{aligned}$$

where we use the side conditions for x, x^σ and λ (see also (2.6) and (2.7)) and have denoted
$\theta^\sigma(q) := \theta[u^\sigma](q)$ and $\theta(q) := \theta[u](q)$. Changing back the order of integration and denoting
$f^\sigma(t,q) = f(t,q,x^\sigma,Q^\sigma,y^\sigma,u^\sigma)$ we obtain from the above equalities that

$$\begin{aligned}
&\int_0^T \int_0^{Q^\sigma(t)} \lambda(t,q)[f^\sigma(t,q) - \dot{x}^b(t)] \,\mathrm{d}q\,\mathrm{d}t \\
&- \int_0^T \int_0^{Q(t)} \lambda(t,q)[f(t,q) - \dot{x}^b(t)] \,\mathrm{d}q\,\mathrm{d}t
\end{aligned}$$

[4] It is to be mentioned that $o(\sigma)$ is not necessarily of second order with respect to σ. It can be of order $3/2$.

$$
\begin{aligned}
&= -\int_0^T \int_0^{Q^\sigma(t)} \dot{\lambda}(t,q)[x^\sigma(t,q) - x^b(t,q)]\,\mathrm{d}q\,\mathrm{d}t \\
&\quad + \int_0^T \int_0^{Q(t)} \dot{\lambda}(t,q)[x(t,q) - x^b(t,q)]\,\mathrm{d}q\,\mathrm{d}t. \qquad (2.43)
\end{aligned}
$$

In a similar way as for L and using the same notational convention we represent the left-hand side of (2.43) as

$$
\begin{aligned}
&\int_0^T \int_0^{Q(t)} \lambda\,[f_x \Delta x + f_Q \Delta Q + f_y \Delta y + \Delta_u f]\,\mathrm{d}q\,\mathrm{d}t \\
&+ \int_0^T \int_{Q(t)}^{Q^\sigma(t)} \lambda\,[f - \dot{x}^b(t)]\,\mathrm{d}q\,\mathrm{d}t + o(\sigma). \qquad (2.44)
\end{aligned}
$$

The right-hand side of (2.43) can be rewritten as

$$
-\int_0^T \int_0^{Q(t)} \dot{\lambda}\Delta x\,\mathrm{d}q\,\mathrm{d}t - \int_0^T \int_{Q(t)}^{Q^\sigma(t)} \dot{\lambda}(t,q)[x^\sigma(t,q) - x^b(t)]\,\mathrm{d}q\,\mathrm{d}t. \qquad (2.45)
$$

Now we shall argue that the second term in the last expression is $o(\sigma)$. We estimate this term by

$$
\begin{aligned}
&\int_0^T \left| \int_{Q(t)}^{Q^\sigma(t)} |x^\sigma(t,q) - x^b(t)|\,\mathrm{d}q \right| \mathrm{d}t\, \|\dot{\lambda}\|_{L_\infty} \\
\leq\; &\int_0^T \left| \int_{Q(t)}^{Q^\sigma(t)} |x^\sigma(t,q) - x^\sigma(\theta^\sigma(q),q)|\,\mathrm{d}q \right| \mathrm{d}t\, \|\dot{\lambda}\|_{L_\infty} \\
&+ \int_0^T \left| \int_{Q(t)}^{Q^\sigma(t)} |x^\sigma(\theta^\sigma(q),q) - x^b(\theta^\sigma(q))|\,\mathrm{d}q \right| \mathrm{d}t\, \|\dot{\lambda}\|_{L_\infty} \\
&+ \int_0^T \left| \int_{Q(t)}^{Q^\sigma(t)} |x^b(\theta^\sigma(q)) - x^b(t)|\,\mathrm{d}q \right| \mathrm{d}t\, \|\dot{\lambda}\|_{L_\infty}. \qquad (2.46)
\end{aligned}
$$

The second term in the right-hand side equals zero due to the definition of x. Let us consider the first term. Since for a.e. q the function $x^\sigma(\cdot,q)$ is Lipschitz with a constant C (independent of q) we have

$$
\begin{aligned}
&\int_0^T \left| \int_{Q(t)}^{Q^\sigma(t)} |x^\sigma(t,q) - x^\sigma(\theta^\sigma(q),q)|\,\mathrm{d}q \right| \mathrm{d}t \\
\leq\; & C\int_0^T \left| \int_{Q(t)}^{Q^\sigma(t)} |t - \theta^\sigma(q)|\,\mathrm{d}q \right| \mathrm{d}t \\
=\; & C\int_0^T \left| \int_{Q(t)}^{Q^\sigma(t)} |\theta^\sigma(Q^\sigma(t)) - \theta^\sigma(q)|\,\mathrm{d}q \right| \mathrm{d}t \\
\leq\; & \frac{C}{\alpha_0}\int_0^T \left| \int_{Q(t)}^{Q^\sigma(t)} |Q^\sigma(t) - q|\,\mathrm{d}q \right| \mathrm{d}t \leq \frac{C}{2\alpha_0} T c^2 \sigma^2,
\end{aligned}
$$

where we have used that θ^σ is Lipschitz continuous with constant $1/\alpha_0$ and also (in the last inequality) that $\|\Delta Q\|_{C[0,T]} \leq c\sigma$ (cf. (2.8)). The same argument applies to the last term in (2.46) since x^b is Lipschitz.

Thus we obtain from (2.43), (2.44) and (2.45) the equality

$$\int_0^T \int_0^{Q(t)} [\dot{\lambda}\Delta x + \lambda(f_x\Delta x + f_Q\Delta Q + f_y\Delta y + \Delta_u f)]\,\mathrm{d}q\,\mathrm{d}t \\ + \int_0^T \int_{Q(t)}^{Q^\sigma(t)} \lambda[f - \dot{x}^b(t)]\,\mathrm{d}q\,\mathrm{d}t = o(\sigma), \tag{2.47}$$

where σ is close to zero.

Now we introduce an absolutely continuous function $\mu : [0,T] \mapsto \mathbf{R}$ satisfying $\mu(T) = 0$. By the same argument that we used in obtaining (2.47) we obtain the equality

$$\int_0^T [\mu(g_Q\Delta Q + g_y\Delta y) + \dot{\mu}\Delta Q]\,\mathrm{d}t = o(\sigma), \tag{2.48}$$

where σ is close to zero (note that ΔQ and Δy depend on σ). As before, $o(\sigma) \approx \sigma^{3/2}$.

Finally we introduce an $(1 \times m)$-dimensional function $\nu \in L_\infty(0,T)$ and obtain from (2.3) by the same argument as before that

$$-\int_0^T \nu\Delta y\,\mathrm{d}t + \int_0^T \int_0^{Q(t)} \nu\Delta_u h\,\mathrm{d}q\,\mathrm{d}t + \int_0^T \int_{Q(t)}^{Q^\sigma(t)} \nu h\,\mathrm{d}q\,\mathrm{d}t = o(\sigma), \tag{2.49}$$

where σ is close to zero. Summing up the equalities (2.42), (2.47), (2.48), (2.49) we obtain (2.10).

Chapter 3

On the relation of country size to political system and wealth inequality

3.1 Introduction

There are few historical examples of territory trade between countries. The biggest are the purchases by the United States of America of Louisiana (2,147,000 km^2 for \$15,000,000) from France in 1803 and Alaska (1,717,854 km^2 for \$7,200,000) from Russia in 1867. In XX-th century the Treaty of Petrópolis between Bolivia and Brazil, signed on November 11, 1903 gave Brazil the territory of Acre (191,000 km^2), in exchange for over 3,000 km^2 of Brazilian territory between the Abunâ and Madeira rivers, a monetary payment of two million British pounds, and a pledge of a rail-link between the Bolivian city of Riberalta and the Brazilian city of Porto Velho.

We study how wealth inequality and the decision making mechanism, adopted by a country, influence its wellbeing, measured by a social welfare function or by the amount of resources (e.g. land) accumulated in the country. We also suggest the condition under which the purchase of the resources (land) happens voluntarily rather than under military threat. We try to imagine why, for example, Alexander II, the Emperor of Russia, could have sold Alaska, and more importantly, why the United States could have bought it.

Indeed, one could argue that usually countries conquer land rather than buy it. Territorial disputes, that accompany trade, have a higher probability of going to war than other kinds of disputes, for example policy disputes (see, e.g. Vasquez & Henehan, 2001). We consider conquest of the territory as a special case of trading, when the buyer country pays not to the seller country but to the third player, a mercenary army, that conquers territory for the buyer, unless the seller pays to the mercenary at least the same amount to keep its territory.

Which tactics would countries prefer: to fight or to deal? The same dilemma has been studied (e.g. H. I. Grossman & Mendoza, 2001a, 2001b) in the economic theory of empire building using examples of the Roman and other empires, where three strategies where considered: Uncoerced Annexation, Coerced Annexation, and Attempted Conquest. In an *Uncoerced Annexation* the Romans would compensate the Barbarians sufficiently to induce the Barbarians to agree to the annexation of their country by Rome. In a *Coerced Annexation* the Romans would induce the Barbarians to agree to the annexation of their country under the threat that the Romans will attack and try to conquer the country. In an *Attempted Conquest* the Romans would attack the Barbarian country. In contrast with choice-theoretic explanation by Grossman and Mendoza we tackle this problem with a game-theoretic model, where both players have the same strategic possibilities, but could have different parameters. We are to find the condition at which countries would trade their territory rather than conquer it. We try to correspond the results with terms Uncoerced Annexation, Coerced Annexation, and Attempted Conquest.

We study these questions on the example of two neighboring countries which use their territories as a production factor (capital). So the countries look like firms competing for a limited resource. Land is deviled between two states. Private good produced in a country is distributed among its citizens proportionally to their shares of the state economy. These shares are inherited from the parents and, in general, not equal. The agents are assumed to be selfish and consuming all their product before end of their life. Hence, product is not accumulated though generations. Generations of agents are assumed to be not overlapping. Thus, each generation starts its life with zero amount of product, but with inherited heterogeneous distribution of shares in the future domestic production.

Either country can trade part of its territory for the part of other's country production. Countries exchange territory for product, or *vice versa*, according to the voting rules corresponding to their forms of government. We model three forms of government: *monarchy* (one monarch gets all the product and makes decisions), *oligarchy* (citizens vote by their shares of the economy), and *democracy* (one citizen – one vote). Notice, that monarchy, in this definition, is equivalent to the situation when decision can be accepted only unanimously, i.e. every citizen has *veto* power. Thus, citizens choose the size of the land to buy and the amount of country's product to be given to a group of neighboring country's citizens for their decision to cede the part of their county's territory.

In the presented study we try to avoid some not very realistic assumptions such as exogenous income, uniform distribution of the territory among agents, and free will or full mobility of every agent, when one independently decides which country to join. It is typical for the literature to use such assumptions for allocation of public facilities (Cremer, Kerchove, & Thisse, 1985) or for the similar problem of the equilibrium size and number of nations on the continuum of uniformly distributed individuals (Alesina & Spolaore, 1997), where each individual at the border between two countries can choose which country to join with her land. (Alesina & Spolaore, 1997) have also considered the coalition equilibrium as well as (Bolton & Roland, 1997). Charles M. Tiebout in his paper (Tiebout, 1956) considered a concept of equilibrium, where the consumer-voter is fully mobile and will move to one of the fixed number of communities where her preference pattern is best satisfied. This concept has been criticized by (Bewley, 1981). He argued that if one tries to generalize the rigorous version of Tiebout's theory in a number of interesting directions, then equilibria may no longer exist or may no longer be Pareto optimal. The existence proof of "strong Tiebout equilibrium" in (Greenberg & Weber, 1986) makes use of the notion of "consecutive games" which the authors introduce and show that for such games there always exists a partition with a nonempty core. More general research (e.g. Haimanko, Breton, & Weber, 2004) also needs special assumptions about the structure of the model in order to prove existence of equilibrium.

Seems that perfect mobility and free will of agents make it difficult to find an equilibrium, therefore we will refrain from such assumptions. The more so because there are always some mobility constraints in the real world. That is why, and also for the sake of simplicity, in the presented model, migration is not allowed. Notice, that in our model a citizen does not own any particular piece of land. She receives her share of country's production, which depends on country's territory. These shares do not need to be equal for all citizens. Thus, we study the problem with heterogeneous wealth distribution among the agents with endogenous income.

This work elaborates on some early author's results in (Belyakov, 2007b).

3.2 Mechanisms of border moving by land trade

We consider two neighboring countries $i \in \{1,2\}$ as firms with production functions $f_i(S_i)$ depending on the countries territories S_i in the following way

$$f_i(S_i) \geq 0, \quad \frac{df_i(S_i)}{dS_i} > 0, \quad \frac{d^2 f_i(S_i)}{dS_i^2} < 0 \quad \text{for all } S_i \geq 0. \tag{3.1}$$

So functions f_i are positive, strictly increasing, strictly concave, and defined for the positive territory. Each country i is populated with constant number of citizens $N_i > 2$ which have their shares θ_{ji} of domestic production. Constant N_i means that no migration is allowed even if a part of the land is transmitted to the other country and no population growth. One can imagine that all population of a country lives in the city which always stays on the territory of the country. For convenience we assume that θ_{ji} are sorted in ascending order over index i and normalized, i.e.

$$\theta_{ji} \leq \theta_{j+1\,i} \quad \text{for all } j = 1,\ldots,N_i - 1, \quad \text{and } \sum_{j=1}^{N_i} \theta_{ji} = 1 \quad \text{for all } i = 1,2,$$

Agents have linear utilities with respect to the produced private good. Private good is the only means of payment in the ceonomy. The territory of the country stays in state property, so its citizens do not own any particular part of it. Thus, besides territory we could similarly consider any other limited natural resource or production factor such as rights to exploit deposits of gas, oil, etc. Citizens can be considered as shareholders.

3.2.1 Territory exchange mechanism and timing

The agents live only one period, and they are selfish, i.e. they do not care about future generations. So citizens of country i inherit from their parents only common country territory S_i and rights for the shares θ_{ji} of domestic production.

During the life period of one generation the countries go through the following sequence of major events.

1. Countries produce private good $f_i(S_i)$ using their inherited territories S_i.

2. Countries transfer the amounts of land ΔS_i which they choose in the result of trade or conquest.

3. Countries produce private good $f_i(S_i + \Delta S_i)$ using their changed territories.

Countries produce product twice (events 1 and 3) and may exchange territory in between (event 2).[1] In the beginning of the period countries have zero amount of private good, because each generation is selfish and consumes all good to the end of its life. So country can pay for the increase of its territory only after it produces good. Hence payments can be made after any of these three events. Citizens do not discount their consumption during the time period.

All land is used and its total amount is constant $S_1 + S_2 = const$, hence the sum of territory changes ΔS_i is zero, meaning

$$\Delta S_1 = -\Delta S_2. \tag{3.2}$$

The government of each country can buy land from its neighbor by paying the citizens of the neighboring country compensations for loses of their income due to decreased production.

Government of each country can transfer the payment (private good) personally to citizens of the other country. We assume that the payment is made before the product is redistributed among the citizens of the county that buys territory. Thus, we can say that, a proportional tax is imposed on the citizens of the buyer country, while individual compensations to citizens of the seller country look like bribes. These are the key assumptions allowing to compensate only to the selected part of the seller country society its loss of utility, that lead to "asymmetry" between buyer and seller.

Sum of product transactions t_i to the country should be zero

$$t_1 + t_2 = 0, \tag{3.3}$$

unless there is no additional expense.

3.2.2 Forms of governments and decision making

The territory exchange needs approval from the citizens of both countries in the case of trade and only from the citizens of attacking country in the case of conquest. The approval is given according to the national decision making system.

Due to the territory change ΔS_i the production of country i changes at the following amount

$$\Delta f_i \equiv f_i(S_i + \Delta S_i) - f_i(S_i). \tag{3.4}$$

[1] Two production events, instead of one, will be essential in Section 3.4, where due to possibility to cheat, countries should have some good to pay before the exchange. Notice, that exchange of territory after the last production would make no sense for selfish population.

Let us first consider the country that sells its territory. In order person j of country i to agree with the deal her loss of utility $\theta_{j\,i}\,\Delta f_i$ should be compensated. We consider the following formes of government implying different voting procedures requiring minimal total compensation $\alpha_i \Delta f_i$:

- *Monarchy* implies that only one person possesses the production and makes decisions (meaning absolute monarchy), i.e. $\theta_{N_i\,i} = 1$ and $\theta_{j\,i} = 0 \quad$ for all $j = 1, \ldots, N_i - 1,$. Hence, agent N_i is the Monarch. Thus, in order to buy some land $-\Delta S_i$ from monarchy i one needs to compensate all loss of its domestic production $-\Delta f_i$. The same total compensation is required in the case of *unanimous rule* when each citizen has *veto* power and therefore is to get compensation, thus

$$\alpha_i = \sum_{j=1}^{N_i} \theta_{j\,i} = 1.$$

- *Oligarchy* means that citizens vote with their shares. A decision is made if people voting for it own together more than a half of domestic production. Here the buyer needs to compensate more than half of the seller country production losses, hence

$$\alpha_i = \min_{J \subset \{1,\ldots,N_i\}} \left(\sum_{j \in J} \theta_{j\,i} \,\middle|\, \sum_{j \in J} \theta_{j\,i} > \frac{1}{2} \right).$$

- *Democracy* has the rule of majority voting. It is sufficient for the buyer to compensate the poorest half of the seller country i society (plus one person if N_i is an even number), with the total shares

$$\alpha_i = \sum_{j=1}^{[N_i/2]+1} \theta_{j\,i}.$$

Thus, the buyer needs to pay for the same amount of land its full cost to a Monarchy, more than half of its cost to an Oligarchy, and the smallest amount to a Democracy. It follows from the following inequalities

$$\sum_{j=1}^{[N_i/2]+1} \theta_{j\,i} \leq \min_{J \subset \{1,\ldots,N_i\}} \left(\sum_{j \in J} \theta_{j\,i} \,\middle|\, \sum_{j \in J} \theta_{j\,i} > \frac{1}{2} \right) \leq 1. \tag{3.5}$$

Parameter $\alpha_i \in (0, 1]$ is the minimal relative share of country loses, that one needs to compensate in order to get agreement of the country to sell its territory. This parameter reflects the decision mechanism in a country. When

$\alpha_i = 1$ then we model country i as a Monarchy, when $\alpha_i \in \left(\frac{[N_i/2]+1}{N_i}, 1\right)$ we consider an Oligarchy, when $\alpha_i \in \left(\frac{1}{2}, \frac{[N_i/2]+1}{N_i}\right]$ it can be either an Oligarchy or a Democracy, and when $\alpha_i \in \left(0, \frac{1}{2}\right]$ it is a Democracy. Coefficient α_i can also be a measure of inequality in country i, if it is a Democracy.

Situation when country i buys land is different. Since shares θ_{ji} of future production remain constant all citizens with $\theta_{ji} > 0$ get benefit $\theta_{ji}\Delta f_i$ from the increase of country territory $\Delta S_i > 0$. Hence, all citizens agree to give their future $\theta_{ji}\Delta f_i$ for the territory increase ΔS_i regardless of the government form in their country. Thus, each country i sells its territory for not less than $-\alpha_i \Delta f_i > 0$ and buys territory for not more than $\Delta f_i > 0$, because otherwise the deal would not be approved. It means that if $\alpha_i < 1$ then country i can sell territory cheaper than buy it. This "asymmetry", as we mentioned in Section 3.2.1, comes from ability of buyer country to compensate only selected citizens of seller county their losses.

Group of country citizens, that needs minimal compensation to promote a deal, will be called the *ruling coalition* of the country. We can say that government of each country maximizes the wealth of its ruling coalition. Thus, α_i is the sum of shares of ruling coalition.

3.3 Fair land trades

Let us consider fair land trading, where no conquest or cheating is allowed. So one cannot refuse to pay for the purchased land even after the final production (event 3 in section 3.2.1). Let us suppose that county 1 buys territory from country 2. The consumption of the buyer consists of the production in the first event $f_1(S_1)$, the production in the third event $f_1(S_1) + \Delta f_1$, and negative payment $t_1 \leq 0$. The maximization problem of the buyer with respect to ΔS_1 and t_1 can be written omitting the constant $2f_1(S_1)$

$$\Delta f_1 + t_1 \quad \to \quad \max_{\Delta S_1 \in [0, -\Delta S_2],\, t_1 \in [-t_2, 0]}, \tag{3.6}$$

$$\text{s.t.} \quad \Delta f_1 + t_1 \quad \geq \quad 0, \quad \text{(participation constraint)} \tag{3.7}$$

$$2f_1(S_1) + \Delta f_1 + t_1 \quad \geq \quad 0, \quad \text{(budget constraint)} \tag{3.8}$$

where maximization of social welfare gain $\Delta f_1 + t_1$ is equivalent to the maximization of the ruling coalition welfare gain $\alpha_1 \Delta f_1 + \alpha_1 t_1$, because all citizens are ready to pay for the increase of their country's territory, since all of them benefit from it proportionally to their constant shares θ_{j1}. Participation constraint (3.7) implies positive social welfare gain and hence ensures budget constraint (3.8) so we can disregard the budget constraint here.

In the seller country, the not ruling part of the society is worse off, because it does not get any compensation for their loss in future production share, in all $(1-\alpha_2)\Delta f_2$. Thus, the seller's problem is as follows

$$\alpha_2\Delta f_2 + t_2 \quad \to \quad \max_{\Delta S_2\in[-\Delta S_1,0],\, t_2\in[0,-t_1]}, \tag{3.9}$$

$$\text{s.t.} \quad \alpha_2\Delta f_2 + t_2 \quad \geq \quad 0. \tag{3.10}$$

Expressions (3.6)–(3.10) form a bargaining problem.

We assume that this problem can be solved by the governments, so that they both are interested in maximization of their total revenue, $\Delta f_1 + \alpha_2\Delta f_2$, from the trade. Indeed, it is easy to show that any reasonable bargaining solution (e.g. Nash bargaining solution) defining the payment $t_1^* = -t_2^*$ for given territory exchange $\Delta S_2 = -\Delta S_1$ provides incentives for both governments to consider such territory exchange that maximizes the total revenue $\Delta f_1 + \alpha_2\Delta f_2$. It follows from the positive dependance of the resulting objective functions $\Delta f_1 + t_1^*$ and $\alpha_2\Delta f_2 + t_2^*$ on the total revenue $\Delta f_1 + \alpha_2\Delta f_2$. Hence, common maximization problem is the following

$$\Delta f_1 + \alpha_2\Delta f_2 \quad \to \quad \max_{\Delta S_1=-\Delta S_2\geq 0}, \tag{3.11}$$

$$\text{s.t.} \quad \Delta f_1 + \alpha_2\Delta f_2 \quad \geq \quad 0, \tag{3.12}$$

where trade possibility condition (3.12) is obtained from the sum of budget constraints (3.7) and (3.10) taking into account relation (3.3). Problem (3.11)–(3.12) does not contain moneyproduct transfers and defines only territory allocation which is an interest of ours.

We can define the price

$$p_i \equiv \frac{\partial \Delta f_i}{\partial \Delta S_i} = \left.\frac{\partial f_i(S)}{\partial S}\right|_{S=S_i+\Delta S_i},$$

using expression (3.4). Thus, p_i is the maximal price at which country i can buy small part of territory. Then, $\alpha_i p_i$ is the price at which country i can sell small part of its territory. In Fig. 3.1 the prices p_i are drown with solid lines and the prices $\alpha_i p_i$ with dashed lines. In these terms the first order condition for problem (3.11) takes the following form

$$p_1 - \alpha_2 p_2 \quad = \quad 0 \quad \text{for } \Delta S_1 > 0, \tag{3.13}$$

$$p_1 - \alpha_2 p_2 \quad \leq \quad 0 \quad \text{for } \Delta S_1 = 0. \tag{3.14}$$

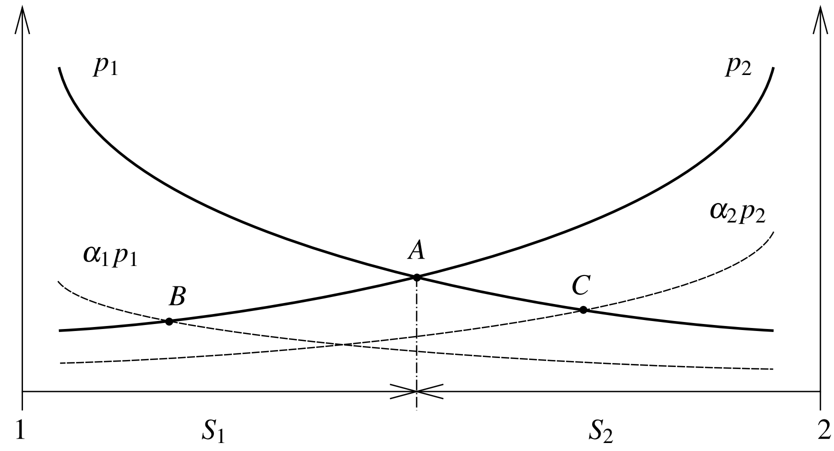


Figure 3.1: Functions p_1 and p_2 of territories S_1 and S_2 are the marginal productivities of territory of countries 1 and 2. Functions $\alpha_1 p_1$ and $\alpha_2 p_2$ are the marginal production shares of ruling coalitions in countries 1 and 2.

Condition (3.13) means that if the initial border between the countries is on the left from the absciss of point C in Figs. 3.1 and 3.2 then the countries will move the border to C. Condition (3.14) means that if the border is on the right of C country 1 will not buy any land.

In order to solve the same problem when country 2 buys territory from country 1 we need only to change indexes in (3.13) and (3.14)

$$p_2 - \alpha_1 p_1 = 0 \quad \text{for } \Delta S_2 > 0, \tag{3.15}$$

$$p_2 - \alpha_1 p_1 \leq 0 \quad \text{for } \Delta S_2 = 0. \tag{3.16}$$

Condition (3.15) means that if the initial border between countries is on the right from the absciss of point B in Figs. 3.1 and 3.2, then the countries will move the border to B. Condition (3.16) means that if the border is on the right of B country 2 will not buy any land.

From the solutions (3.13)–(3.16) we can conclude that if the initial border position is on the left from point B, the only option is for country 2 to sell its lend to country 1, so that new border is at point C, and *vice versa*, if the initial border position is on the right from point C, the only option is for country 1 to sell its lend to country 2, so that new border is at point B. The only question left is at which point, B or C, countries will move their border if its initial position is between B and C. The answer, which country to be the buyer and which the seller, depends on the particular bargaining conditions. But we can assume, as before, that governments maximize the total revenue from the trade. Then

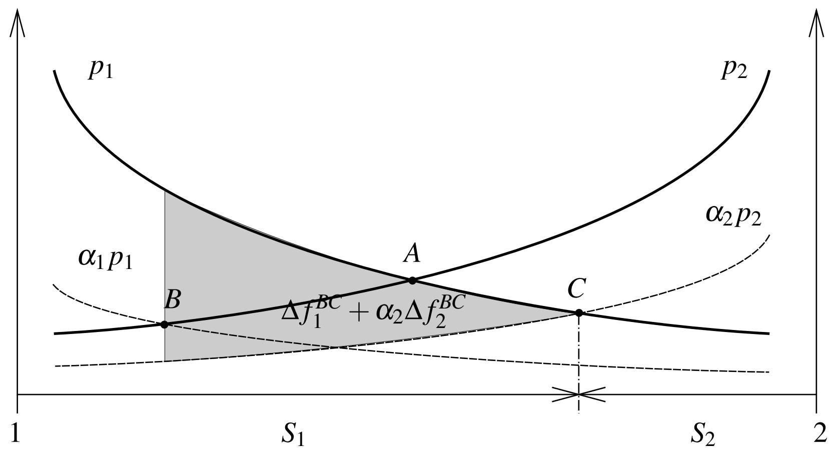


Figure 3.2: Country 1 buys territory from country 2 moving the border from point B to point C. Total trade revenue is the gray area $\Delta f_1^{BC} + \alpha_2 \Delta f_2^{BC}$.

the border near or at point B will be moved to point C, and *vice versa*, the border near or at point C will be moved to point B, which creates a cycle in generations. For example, when country 1 is the buyer, moving border from B to C, the total revenue is $\Delta f_1^{BC} + \alpha_2 \Delta f_2^{BC}$, which is the gray area in Fig. 3.2, where we denote $\Delta f_i^{BC} = f_i(S_i^C) - f_i(S_i^B)$ according to (3.4). In the next life period the situation is the opposite, and new generations of citizens move the border from C to B.

Thus, we get a cyclical solution which is a sequence of bargaining results on the border position (of events 2, section 3.2.1), one in each generation's life time. Generally speaking, this solution is asymptotically stable in Lyapunov's sense (Liapounoff, 1907).

Definition 1 *Let $x(\tau)$ be the solution and $\tilde{x}(\tau)$ be the perturbed solution, where τ is the time. Solution $x(\tau)$ is (asymptotically) stable if for every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that $|\tilde{x}(0) - x(0)| < \delta$ is sufficient for $|\tilde{x}(\tau) - x(\tau)| < \varepsilon$ for all $\tau > 0$ (and $\lim\limits_{\tau \to \infty} \tilde{x}(\tau) = x(\tau)$).*

The only case when solution could be unstable is when the initial border position is at the threshold, where countries can not decide which country should sell or buy, because both options give equal total trade revenue.

Let us introduce average size of the the country i

$$\langle S_i \rangle = \frac{S_i^B + S_i^C}{2}, \tag{3.17}$$

where S_i^B is the size of country when the border is at point B, while S_i^C is that at C. We can conclude from Figs. 3.1 and 3.2, that minimal territory of country i positively depends on its parameter α_i, while maximal size of the country does not depend on α_i

$$\frac{dS_1^B}{d\alpha^1} > 0, \quad \frac{dS_2^C}{d\alpha^2} > 0, \quad \frac{dS_1^C}{d\alpha^1} = 0, \quad \frac{dS_2^B}{d\alpha^2} = 0. \tag{3.18}$$

Hence, the average size of country i positively depends on parameter α_i, i.e.

$$\frac{d\langle S_i\rangle}{d\alpha_i} > 0. \tag{3.19}$$

It means that, *caeteris paribus*, Monarchy in average is bigger than Oligarchy, which is, in tern, bigger than Democracy. The greater the inequality in the Democracy, the smaller its territory could be.

3.3.1 Welfare analysis

Even without knowledge about exact values of product transfers t_i we can say something about the dependance between social welfare of the country and its type of government. Since agents have linear utilities the utilitarian social welfare function can be evaluated as production plus product transfer $W_i = f_i(S_i) + t_i$. Let us introduce the average social welfare of country i as

$$\langle W_i\rangle = \frac{f_i(S_i^B) + t_i^B + f_i(S_i^C) + t_i^C}{2}, \tag{3.20}$$

where S_i^B and t_i^B are the territory and product transfers of country i, if the border is at point B; while S_i^C and t_i^C are those at C in Fig. 3.2. Thus, regardless of the governments' bargaining powers, we can write the lower bounds of countries' welfares

$$\langle W_1\rangle \geq f_1(S_1^B), \quad \langle W_2\rangle \geq f_2(S_2^C). \tag{3.21}$$

The global average social welfare can be calculated precisely from (3.20), due to the payment balance (3.3):

$$\langle W_1 + W_2\rangle = \sum_{i=1}^{2} \frac{f_i(S_i^B) + f_i(S_i^C)}{2}. \tag{3.22}$$

If we differentiate expression (3.22) with respect to parameters α_i, then using conditions (3.1) and (3.18), we get that the global average social welfare (3.22) positively depends on both parameters α_1 and α_2, i.e.

$$\frac{d\langle W_1 + W_2\rangle}{d\alpha_i} > 0 \quad \text{for all } i = 1, 2. \tag{3.23}$$

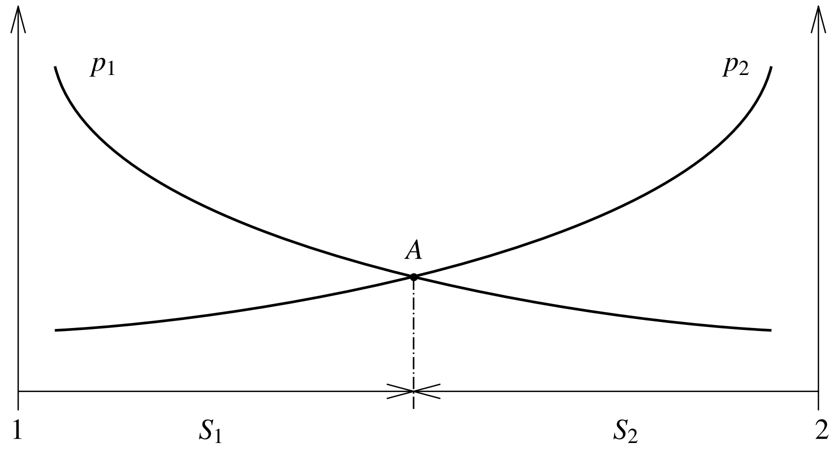


Figure 3.3: Functions p_1 and p_2 of territories S_1 and S_2 are the marginal productivities of territory of countries 1 and 2.

Thus, we can get the unique global social optimum for the corner solution $\alpha_1 = \alpha_2 = 1$, which implies that each country is Monarchy or has unanimous voting rule. So the counties locate the border at point A.

3.3.2 Two Monarchies choose the stable social optimum

When two Monarchies trade territory the buyer have to compensate all domestic production loss to the seller since $\alpha_1 = \alpha_2 = 1$. That is why maximization problems (3.6)–(3.7) and (3.9)–(3.10) coincide:

$$\Delta f_1 + \Delta f_2 \quad \rightarrow \quad \max_{\Delta S_1 = -\Delta S_2}, \tag{3.24}$$

$$\text{s.t.} \quad \Delta f_1 + \Delta f_2 \quad \geq \quad 0, \tag{3.25}$$

and have the same unique solution $p_1 = p_2$, positioning the border at point A in Fig. 3.3.

This territory allocation is a global social optimum as it actually maximizes the sum of production in both countries

$$\Delta S_1^A = \arg \max_{\Delta S_1 = -\Delta S_2} f_1(S_1 + \Delta S_1) + f_2(S_2 + \Delta S_2) = \arg \max_{\Delta S_1 = -\Delta S_2} \Delta f_1 + \Delta f_2.$$

That is why this allocation is Pareto efficient. We note that the territory allocation of two Monarchies is a steady state equilibrium which means that if time period starts with such allocation the Monarchies will not change it. This steady state solution is also asymptotically stable in Lyapunov's sense; see definition 1, which means that if period starts with slightly different allocation then it converges to that with border at point A.

3.4 Effect of possibility to cheat and conquest

Let us add to the model the possibilities of not committing the obligation on a contract (cheating) by the seller and conquest of the territory by the buyer. The conquest is a two step sequential game, see Fig. 3.4. In the first step country 1 (conqueror) instead of paying for the territory, pays $-t_1^a \geq 0$ for hiring the army to conquer a part of neighbor's territory. In the second step, if the country under attack does not pay equal or greater amount of product $-t_2^a$ to the army to defend its territory, then the territory goes to the conqueror.

We recall that private good is produced twice in the time period: in the beginning and in the end (see section 3.2.1). We assume that between these two productions (event 2 in section 3.2.1) countries play the game in Fig. 3.5, which includes two conquest subgames, see Fig. 3.4. Thus, before the game each country is endowed with the good from the first production. Only that amount of good can be used for land purchasing, because no one would rely on the promise to pay in the end of the period if cheating is allowed. Hence, there is the initial product constraint

$$f_i(S_i) + t_i^A \geq 0, \tag{3.26}$$

which matters only for the conqueror as will be shown later on.

Conquest looks like country 1 buys land from Monarchy (see Fig. 3.1), having budget constraint (3.26). The only difference is that the seller does not get the payment

$$-t_1^a = \min\left\{-\Delta f_2^A, f_1(S_1)\right\}, \tag{3.27}$$

which goes to the Army. The Army could be treated as the third player, which can not gain utility from the land itself but only conquers territory for the country which pays the highest price.

In the conquest game in Fig. 3.4, country 2 will wish to defend its territory ΔS_2 with minimal military expanses $-t_2^a = -t_1^a$ if conqueror pays $-t_1^a < -\Delta f_2$, otherwise country 2 retreats with $t_2^a = 0$.

Proposition 1 *Initial production $f_i(S_i)$ of country i is greater than loss $-\Delta f_i > 0$ from any territory decrease $\Delta S_i \in [-S_i, 0)$, i.e. $f_i(S_i) > -\Delta f_i$.*

Proof I t follows from definition (3.4) of symbol Δf_i and the first property of function $f_i(S_i)$ in (3.1). □

Corollary 1 *In order to conquer the territory ΔS_1 country 1 should transfer $t_1^a = \Delta f_2$*

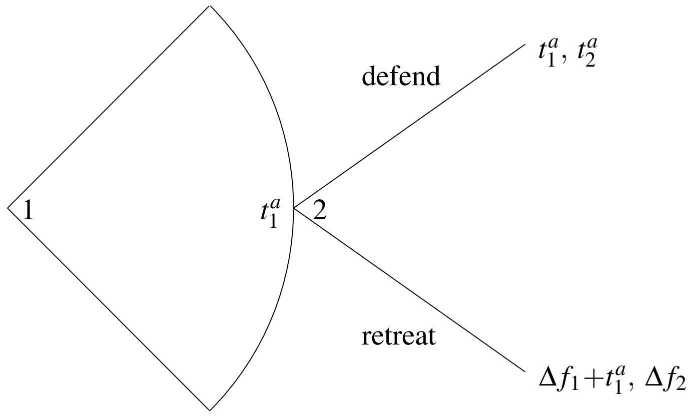


Figure 3.4: Conquest game. Payoffs of countries are sums of payoffs of all its citizens. Conqueror chooses military transfer $t_1^a < 0$. Attacked country defends spending $t_2^a = t_1^a < 0$ if $t_1 > \Delta f_2$, or retreats spending $t_2^a = 0$ if $t_1^a \leq \Delta f_2$. That is why in SPNE conqueror chooses $t_1^a = \Delta f_2$.

Proof I t follows from proposition 1, that country 2 has enough initial endowment $f_2(S_2)$ to defend its territory, which it wishes to do only if $-t_1^a < -\Delta f_2$. Hence, inequality (3.26) is strict and $-t_1^A = -\Delta f_2$ is the minimal payment for conquest. □

Using corollary 2, we can write the maximization problem of the conqueror in the game Fig. 3.4

$$\Delta f_1 + \Delta f_2 \quad \to \quad \max_{\Delta S_1 = -\Delta S_2 \geq 0}, \tag{3.28}$$

$$\text{s.t.} \quad f_1(S_1) + \Delta f_2 \quad \geq \quad 0, \tag{3.29}$$

which differs from (3.24)–(3.25) only by constraint (3.29). Constraint (3.29) can be active. To specify the Subgame Perfect Nesh Equilibrium (SPNE) of the conquest game in Fig. 3.4 in cases of active and inactive constraint (3.29) we introduce the following definition.

Definition 2 *We will say that country 1 is* ***rich*** *if its initial production is greater than total loss of production in country 2 after moving the border to point A, i.e. the following inequality is satisfied*

$$f_1(S_1) > -\Delta f_2^A. \tag{3.30}$$

Thus, country 1 is ***not rich*** *if inequality (3.30) is not satisfied.*

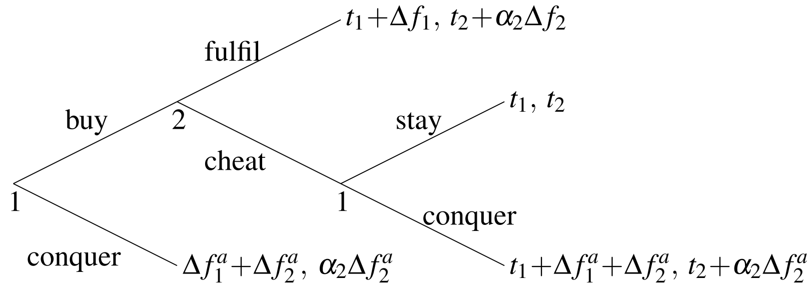


Figure 3.5: The game territory exchange is played each time period by new generation. Payoffs of country 1 are sums of payoffs of all its citizens. Payoffs of country 2 are sums of payoffs of its ruling coalition members.

The SPNE of the game in Fig. 3.4 consists of the payments to the army (3.27), $t_2^a = 0$, and territory changes ΔS_i which can be found in both cases of *rich* and *not rich* country 1.

1. If country 1 is *rich* (see definition (2)), then it conquers ΔS_1^A. Conqueror's profit is $\Delta f_1^A + \Delta f_2^A \geq 0$ while the profit of country 2 is $\Delta f_2^A \leq 0$ and profit of its ruling coalition is $\alpha_2 \Delta f_2^A \leq 0$.

2. If country 1 is *not rich*, then inequality (3.30) is not satisfied, which implies that country 1 conquers such amount of territory $\Delta S_1^a \leq \Delta S_1^A$ that $f_1(S_1) = -\Delta f_2$.

We will analyze the game in Fig. 3.5 for both internal solution for rich conqueror and corner solution for not rich one.

If the government of country 1 wants to buy land ΔS_1, then it proposes to its citizens a project of purchase, for which amount of product $-t_1$ is collected from the citizens who will benefit from the territory increase. It means that everybody, who have strictly positive share θ_{j1} of production, would agree to pay not more than $\theta_{j1}\Delta f_1$ and, hence, would vote for the project. Then the government of country 1 gives product $t_2 = -t_1 \leq \min\{-f_1(S_1), \Delta f_1\}$ to the ruling coalition of country 2.

The ruling coalition of country 2 receives the amount of product t_2 and can execute its duty, giving required territory $-\Delta S_2$ to the buyer, or it can cheat, refusing to give the territory. In the first case the profit of the seller's ruling coalition is $t_2 + \alpha_2 \Delta f_2$, profit of the buyer is $t_1 + \Delta f_1$, where transfers are balanced according to (3.3), and the game is finished. In the case of cheating, the profit of seller is t_2, while the profit of buyer is $t_1 < 0$. But the buyer has the possibility to conquer part ΔS_1^a of the disputed territory, paying for this Δf_2^a as much as the seller maximally can pay to keep that territory. In that case, being rational, the seller country retreats without fight, and its ruling coalition have profit $t_2 + \alpha_2 \Delta f_2^a$ while the profit of the buyer is $t_1 + \Delta f_1^a + \Delta f_2^a$, and the game is finished.

Proposition 2 *Country* 1 *buys only such amount of land* ΔS_1*, which it would be able to conquer in the case of cheating by county* 2*.*

Country 1 buys land when it surely is not going to be cheated, because otherwise it would conquer it in the first place. If county 1 buys more than it could conquer after has been cheated, then it gives country 2 incentives to cheat, because country 2 would have more land rest after conquest than after executing its duties. □

Corollary 2 *Country* 1 *buys land of amount* ΔS_1*, not exceeding the amount* ΔS_1^A*, that locate the border at point A, like it would trade with Monarchy.*

Since point A is the optimal position of the border for unconstrained conquest (3.28) country 1 would never conquer more than ΔS_1^A after being cheated. It follows from proposition 2 that country 1 also do not buy more then ΔS_1^A. □

It follows from corollary 2, that maximization problem for buying the lend

$$\Delta f_1 + t_1 \quad \to \quad \max_{\Delta S_1 \geq 0,\, t_1 \leq 0}, \tag{3.31}$$

$$\text{s.t.} \quad f_1(S_1) + \Delta f_2 + t_1 \quad \geq \quad 0, \tag{3.32}$$

has more restrictive budget constraint (3.32) than (3.7). If both countries prefer to trade the bargaining problem is composed by (3.31)–(3.32) and (3.9)–(3.10) with conquest (3.28)–(3.29) as a disagreement point.

1. When country 1 is *not rich* (see definition (2)) then inequality (3.32) becomes an equality. Hence, ΔS_1^a is also a solution of maximization problem (3.31)–(3.32), which yields profit $\Delta f_1^a + \Delta f_2^a + t_1$, where $t_1 \leq 0$. Country 1 chooses between conquest and purchase, comparing their profits $\Delta f_1^a + \Delta f_2^a$ and $\Delta f_1^a + \Delta f_2^a + t_1$, therefore, it would buy only for free, i.e. $t_1 = 0$, which is the disagreement point for seller country 2. Thus, not rich country conquers territory rather then buys it.

2. When country 1 is *rich* (inequality (3.30) is satisfied) it buys ΔS_1^A according to corollary 2, with payment $t_2 = -t_1 \leq f_1(S_1) + \Delta f_2^A$.

These strategies are SPNE of the game in Fig. 3.5 which can be described in terms introduced in (H. I. Grossman & Mendoza, 2001a, 2001b). We distinguish the type of annexation according to the values of t_2.

- We can say that if $t_2 + \alpha_2 \Delta f_2^A \geq 0$, then it is an Uncoerced Annexation, because the ruling coalition of country 2 is better off in the end of the period.
- If $t_2 + \alpha_2 \Delta f_2^A < 0$ while $t_2 > 0$, then it is a Coerced Annexation, because the ruling coalition of country 2 is worse off after the trade. Which means that country 1 has threatened country 2 with possibility of conquest. Though, we have not any additional costs introduced for Coerced Annexation, like the cost of deploying Legions in (H. I. Grossman & Mendoza, 2001a, 2001b).
- If the countries have nothing to bargain about, because country 1 is not rich, then it simply conquers the territory ΔS_1^a, which we could call an Attempted Conquest, where $t_2 = 0$.

Thus, not rich yet more efficient country conquers territory generation by generation until it becomes rich enough to buy territory of the globally optimal size S_i^A.

The main result of this section is that the possibility to cheat and conquest completely removes the dependence of the country size on its type of government. More then that, this possibility results in the Pareto efficient and asymptotically stable steady state territory allocation like that between two Monarchies with border at point A in Fig. 3.1.

Another interesting result is that "not rich" countries expand by concurring territory. While "rich" countries buy territory, because other countries would not dare to concur it. That could be the reason why Alaska was sold to the United States rather than absorbed in North-West Territories of future Canada. Although Russian-American Company, that was a monopoly in Alaska, was evaluated 3,721,400 silver rubles, see (Okun, 1939), while Alaska was sold for approximately 11 million silver rubles. Thus, clearly we have a case of Uncoerced Annexation.

3.5 Effect of additional costs

Until now we believed that the there is no costs for adaptation of new territory or moving the border. Let us assume that there are adaptation cost $d \geq 0$ per

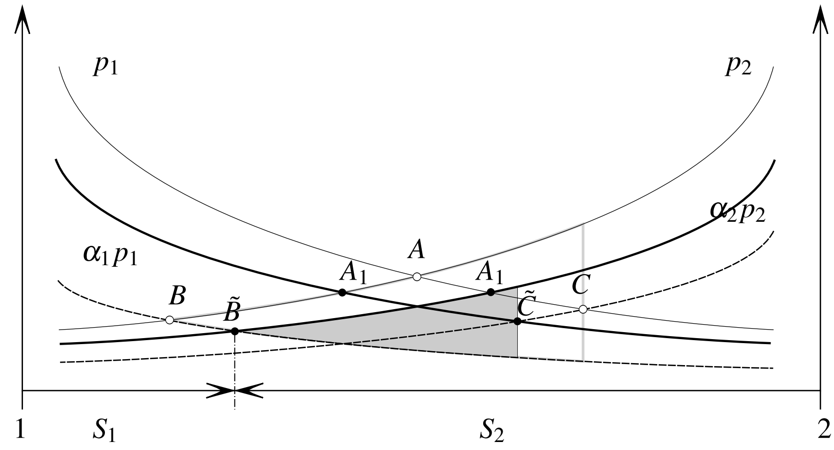


Figure 3.6: Country 1 buys territory from country 2 moving the border from point B to point C. Total trade revenue $\Delta f_1^{\tilde{B}\tilde{C}} + \alpha_2 \Delta f_2^{\tilde{B}\tilde{C}} - c - d\left|\Delta S_1^{\tilde{B}\tilde{C}}\right|$ is the gray area minus constant c.

unit of traded territory and constant cost $c \geq 0$ of rebuilding the border, then product balance equation (3.3) takes the following form:

$$t_1 + t_2 = -c - d\,|\Delta S_1|\,. \tag{3.33}$$

Hence, maximization problem (3.11)–(3.12) also changes:

$$\begin{aligned} &\Delta f_1 + \alpha_2 \Delta f_2 - c - d\,|\Delta S_1| \quad \to \quad \max_{\Delta S_1 = -\Delta S_2 \geq 0}, && (3.34)\\ \text{s.t.}\quad &\Delta f_1 + \alpha_2 \Delta f_2 - c - d\,|\Delta S_1| \quad \geq \quad 0, && (3.35) \end{aligned}$$

The first order condition for problem (3.34) takes the following form:

$$\begin{aligned} p_1 - \alpha_2 p_2 - d \quad &= \quad 0 \quad \text{for } \Delta S_1 > 0, && (3.36)\\ p_1 - \alpha_2 p_2 \quad &\leq \quad 0 \quad \text{for } \Delta S_1 = 0. && (3.37) \end{aligned}$$

We see that only per unit adaptation cost d influences the first order condition (3.36)–(3.37), and hence, moves points B and C to $\tilde{B}$ and $\tilde{C}$, see Fig. (3.6), while constant c in the maximization constraint (3.12) determines whether trade will happen or not. Thus, oscillatory solutions can occur only if both of the following inequalities are satisfied

$$\begin{aligned} \Delta f_1^{\tilde{B}\tilde{C}} + \alpha_2 \Delta f_2^{\tilde{B}\tilde{C}} - c - d\left(S_1^{\tilde{C}} - S_1^{\tilde{B}}\right) \quad &\geq \quad 0, && (3.38)\\ \Delta f_2^{\tilde{C}\tilde{B}} + \alpha_1 \Delta f_1^{\tilde{C}\tilde{B}} - c - d\left(S_2^{\tilde{B}} - S_2^{\tilde{C}}\right) \quad &\geq \quad 0. && (3.39) \end{aligned}$$

It is easily seen form Def. 1, that if inequalities (3.38) and (3.39) are strict for all points of a solution, then the oscillatory solution is asymptotically stable. Oscillatory solutions could coexist with multiple steady state solutions, these are all such points S_i for which the following inequalities are satisfied:

$$\Delta f_1^{\tilde{C}} + \alpha_2 \Delta f_2^{\tilde{C}} - c - d\left(S_1^{\tilde{C}} - S_1\right) \quad < \quad 0, \qquad (3.40)$$

$$\Delta f_2^{\tilde{B}} + \alpha_1 \Delta f_1^{\tilde{B}} - c - d\left(S_2^{\tilde{B}} - S_2\right) \quad < \quad 0. \qquad (3.41)$$

These steady state solutions are stable but not asymptotically. Here it means that if we change the initial point a little (so that inequalities (3.40) and (3.41) are still satisfied), then the perturbed solution neither goes further away from the unperturbed solution, nor converges back to it.

If cheating and conquest are allowed, then the same game theoretic reasoning, as in the previous section, gives us the existence of a steady state points S_i, which should satisfy the following inequality:

$$\Delta f_1^{\tilde{A}} + \Delta f_2^{\tilde{A}} - c - d\left|S_i^{\tilde{A}} - S_i\right| \quad < \quad 0. \qquad (3.42)$$

It means that all points between A_1 and A_2 are stable steady state solutions.

We can conclude that additional linear costs reduce the amplitude of territory oscillations. Hence, they decrease dependence between size of the country and its type of government. Such costs can also cause the appearance of multiple stable steady state positions of the border.

3.6 Conclusion

In this work the mechanism governing the territory exchanging between groups of people such as countries is proposed, based on trading with approval of both sides under particular voting rule (veto rule, majority rule,...).

We have obtained a cyclical in time asymptotically stable solution, which is the sequence of bargaining results on the border position, one in each generation life time. We also have found that the average size and social welfare of the Monarchy are bigger than those of Oligarchy, which in turn bigger than those of Democracy. The greater inequality in a Democracy, the smaller its territory could be.

Possibility to cheat and conquest completely removes the dependence of the country size on its type of government. More than that, this possibility results in the Pareto efficient and asymptotically stable steady state territory allocation, like one between two Monarchies or countries with veto voting rules (when decisions can be made only unanimously).

It was found that only rich countries, that produce more product than sellers could lose, Definition 2, can afford to buy territory rather then conquer it. Monarchy sells its territory only when it is to difficult to protect, for instance, because of the high distances, like in the case of Alaska, that would have been difficult to defend form British Empire. We may say, that it was the US, that was rich enough to be capable of defending its new territory.

Additional linear costs reduce the amplitude of territory oscillations. Hence, they decrease the dependence of the size of a country on its form of government. Such costs can also cause the appearance of multiple stable steady state positions of the border.

References

Aghion, P., & Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, *60*(2), 323–351.

Aghion, P., & Howitt, P. (1996). Research and development in the growth process. *Journal of Economic Growth*, *1*(1), 49-73.

Aghion, P., & Howitt, P. (1998). *Endogenous growth theory*. Cambridge, MA: MIT Press.

Alesina, A., & Spolaore, E. (1997). On the number and size of nations. *The Quarterly Journal of Economics*, *112*(4), 1027–1056.

Aubin, J.-P. (1991). *Viability theory*. Boston, Basel, Berlin: Birkhäuser.

Aubin, J.-P., & Frankowska, H. (1990). *Set-valued analysis*. Boston, Basel, Berlin: Birkhäuser.

Barro, R., & Martin, X. Sala-i. (2004). *Economic growth*. Cambridge, MA: MIT Press.

Belyakov, A. O. (2007a). About mathematical description of the people's association development processes (ethnoses, firms staff etc.). *Economics and Mathematical Methods*, *43*(2), 118–122.

Belyakov, A. O. (2007b). *On the dynamics of people's unions* (Working Paper No. BSP/2007/094E). New Economic School. Available from `http://www.nes.ru/dataupload/files/programs/econ/preprints/2007/Belyakov_eng.pdf`

Belyakov, A. O. (2009, June). On the relation of country size to the form of government. In *Ecore summer school on market evolution and public decision.* Brussels. Available from `http://www.ecares.org/ecaresdocuments/ws/summerschool2009/papers/belyakov.pdf`

Belyakov, A. O., Davydov, A., & Veliov, V. M. (2013). Optimal cyclic exploitation of renewable resources [Research Report]. (2013-07). Available from `http://orcos.tuwien.ac.at/fileadmin/t/orcos/Research_Reports/2013-07b.pdf`

Belyakov, A. O., Feichtinger, G., & Simon, C. (2011). *General* *equilibrium model with horizontal innovations and heterogeneous products* (Research Report No. 2011-13). Operations Research and Control Systems, Institute of Mathematical Methods in Economics, Vienna University of Technology. Available from http://orcos.tuwien.ac.at/fileadmin/t/orcos/Research_Reports/2011-13.pdf

Belyakov, A. O., Haunschmied, J. L., & Veliov, V. M. (2012). *General equilibrium model with horizontal innovations and heterogeneous products* (Research Report No. 2012-01). Operations Research and Control Systems, Institute of Mathematical Methods in Economics, Vienna University of Technology. Available from http://orcos.tuwien.ac.at/fileadmin/t/orcos/Research_Reports/2012-01_Bely-VV-HS.pdf

Belyakov, A. O., Tsachev, T., & Veliov, V. M. (2011). Optimal control of heterogeneous systems with endogenous domain of heterogeneity. *Applied Mathematics & Optimization*, *64*(2), 287–311.

Belyakov, A. O., & Veliov, V. M. (2014). Constant versus periodic fishing: age structured optimal control approach [Research Report]. (2014-01). Available from http://orcos.tuwien.ac.at/fileadmin/t/orcos/Research_Reports/2014-01.pdf

Bewley, T. F. (1981). A critique of tiebout's theory of local public expenditures. *Econometrica*, *49*(3), 713–740.

Blanchard, O. (1985). Debt, deficits and finate horizons. *Journal of Political Economy*, *93*, 223–247.

Bolton, P., & Roland, G. (1997). The breakup of nations: A political economy analysis. *The Quarterly Journal of Economics*, *112*(4), 1057–1090.

Boucekkine, R., de la Croix, D., & Licandro, O. (2002). Vintage human capital, demographic trends, and endogenous growth. *Journal of Economic Theory*, *104*(2), 340 - 375.

Boucekkine, R., Río, F. del, & Licandro, O. (2005). Obsolescence and modernization in the growth process. *Journal of Development Economics*, *77*(1), 153–171.

Cass, D., & Yaari, M. E. (1967). Individual saving, aggregate capital accumulation, and efficient growth. In K. Shell (Ed.), *Essays on the theory of optimal economic growth* (p. 233-268). Cambridge, MA: MIT.

Corduneanu, C. (1971). *Principles of differential and integral equations*. Boston: Allyn and Bacon.

Cremer, H., Kerchove, A.-M. D., & Thisse, J.-F. (1985). An economic theory of public facilities in space. *Mathematical Social Sciences*, *9*(3), 249–262.

d'Albis, H., & Augeraud-Véron, E. (2011). Continuous-time overlapping generations models. In R. Boucekkine, N. Hritonenko, & Y. Yatsenko (Eds.), *Optimal control of age-structured population in economy, demography, and the environment. routledge explorations in environmental economics* (pp. 45–69). Taylor and Francis.

Diamond, P. A. (1965). National debt in a neoclassical growth model. *American Economic Review*, *55*, 1126–1150.

Dinopoulos, E., & Thompson, P. (1998). Schumpeterian growth without scale effects. *Journal of Economic Growth*, *3*(4), 313–335.

Dixit, A. K., & Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *American Economic Review*, *67*(3), 297–308.

Faruqee, H. (2003). Debt, deficits, and age-specific mortality. *Review of Economic Dynamics*, *6*(2), 300 - 312.

Gale, D. (1973). Pure exchange equilibrium of dynamic economic models. *Journal of Economic Theory*, *6*, 12–36.

Greenberg, J., & Weber, S. (1986). Strong tiebout equilibrium under restricted preferences domain. *Journal of Economic Theory*, *38*(1), 101–117.

Grossman, G. M., & Helpman, E. (1989). Product development and international trade. *Journal of Political Economy*, *97*(6), 1261–1283.

Grossman, G. M., & Helpman, E. (1991a). *Innovation and growth in the global economy*. MIT Press.

Grossman, G. M., & Helpman, E. (1991b). *Innovation and growth in the global economy*. Cambridge, MA: MIT Press.

Grossman, H. I., & Mendoza, J. (2001a, February). *Annexation or conquest? the economics of empire building* (Working Paper No. 8109). National Bureau of Economic Research. Available from `http://www.nber.org/papers/w8109`

Grossman, H. I., & Mendoza, J. (2001b). Butter and guns: Complementarity between economic and military competition. *Economics of Governance*, *2*(1), 25–33.

Haimanko, O., Breton, M. L., & Weber, S. (2004). Voluntary formation of communities for the provision of public projects. *Journal of Economic Theory*, *115*(1), 1–34.

Jones, C. I. (1976). Product selection, fixed costs, and monopolistic competition. *Review of economic studies*, *43*(2), 217–235.

Jones, C. I. (1995a). R&d-based models of economic growth. *Journal of Political Economy*, *103*(4), 759–784.

Jones, C. I. (1995b). Time-series tests of endogenous growth models. *Quarterly Journal of Economics*, *110*, 495–525.

Jones, C. I. (1999). Growth: With or without scale effects? *American Economic Review*, *89*, 139–144.

Judd, K. L. (1985). On the performance of patents. *Econometrica*, *53*, 567–585.

Kortum, S. S. (1997). Research, patenting, and technological change. *Econometrica*, *65*(6), 1389-1419.

Liapounoff, A. (1907). Problème général de la stabilité du mouvement. *Annales de la faculté des scienses de Toulouse Sér. 2*, *9*, 203–474.

Okun, S. B. (1939). *Russian-american company*. Moscow - Leningrad: SotsEkGiz.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, *98*, 71–102.

Samuelson, P. A. (1958). An exact consumption-loan model of interest with or without the social contrivance of money. *Journal of Political Economy*, *66*, 467–482.

Segerstrom, P. S. (1998). Endogenous growth without scale effects. *American Economic Review*, *88*(5), 1290–1310.

Simon, C., Belyakov, A. O., & Feichtinger, G. (2012). Minimizing the dependency ratio in a population with below-replacement fertility through immigration. *Theoretical Population Biology*, *82*(3), 158-169.

Skritek, B., Tsachev, T., & Veliov, V. M. (2014). Optimality conditions and the hamiltonian for a distributed optimal control problem on controlled domain. *Applied Mathematics & Optimization*, 1-24.

Sorger, G. (2011). Horizontal innovations with endogenous quality choice. *Economica*, *78*(312), 697–722.

Tiebout, C. M. (1956). Voluntary formation of communities for the provision of public projects. *Journal of Political Economy*, *64*(5), 416–424.

Vasquez, J., & Henehan, M. T. (2001). Territorial disputes and the probability of war, 1816-1992. *Journal of Peace Research*, *38*(2), 123-138.

Veliov, V. M. (2008). Optimal control of heterogeneous systems: Basic theory. *Journal of Mathematical Analysis and Applications*, *346*(1), 227 - 242.

Yaari, M. E. (1965). Uncertain lifetime, life insurance and the theory of the consumer. *The Review of Economic Studies*, *70*, 83–101.