

CONSTRAINTS AND POLICY IN EDUCATION AND PUBLIC
BUDGET LIMITS

by

ERIC STEPHENS

A thesis submitted to the
Department of Economics
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada

November, 2009

Copyright © Eric Stephens, 2009

Abstract

This dissertation investigates the impacts of constraints on optimal government policy. The first two chapters provide a general introduction and review of the literature. Chapters three and four analyze education spending and institutional structure in an economy with informational asymmetries and employment matching frictions. The fifth investigates the impact of politics on government decisions over taxation and spending programs more generally.

Chapter three analyzes the situation where governments can target education funds to specific observable groups (referred to as low and high productivity neighbourhoods/regions). The results suggest that it may be optimal to employ educational transfers, rather than cash transfers amongst individuals, to achieve social welfare objectives. This is because the former can reduce distortions created by the tax system. However, the value of educational spending in mitigating these information frictions may not be that large, and may in fact exacerbate such distortions. This suggests that an optimal education policy may be more regressive when there is a distortionary tax system in place. Further, we show that even if “equalizing opportunities” is deemed optimal in the static problem, it may not be a reasonable policy goal when we extend the analysis to include dynamics.

Chapter four is joint work with Afrasiab Mirza. We analyze an economy where heterogeneous individuals are uncertain about their endowments. The education system trades off the desire to capitalize on talent through specific skills training with the need to provide individuals with opportunities to learn about their career preferences through a broader education. The results consider the implications of various educational institutions for the income distribution and consequently welfare.

Chapter five analyzes the dynamics of public spending, taxation and debt in a political agency model. Choices are made by short-lived politicians who can be only partially controlled through the electoral process. The main focus is to consider the impacts of binding limits on the public budget. The value of imposing this additional friction depends both on the extent to which politicians' goals deviate from their constituencies and how effectively the electoral process disciplines them when they misbehave. The results also suggest that the value of such a restriction depends on the fiscal position at the time in which it is imposed.

Co-Authorship

Chapter 4 of this thesis was co-authored with Afrasiab Mirza.

Acknowledgments

I sincerely thank Robin Boadway, Sumon Majumdar and Charles Beach for their guidance, insights and encouragement.

I thank the seminar participants of the Canadian Economic Association Meetings, Queen's University, University of Alberta, Wilfrid Laurier University, Singapore Management University and the Australian National University.

Finally, I am very grateful to my family and friends. Especially the Wren Lake Pirates.

Table of Contents

Abstract	i
Co-Authorship	iii
Acknowledgments	iv
Table of Contents	v
List of Tables	viii
List of Figures	ix
Chapter 1:	
General Introduction	1
Chapter 2:	
Literature Review	7
Chapter 3:	
Education vs. Optimal Taxation: The Cost of Equalizing Oppor- tunities	20

3.1	Introduction	20
3.2	Individuals	24
3.3	Education policy	25
3.4	Social mobility, equal opportunities and dynamics	55
3.5	Conclusions	66

Chapter 4:

	Education Systems	69
4.1	Introduction	69
4.2	Model	72
4.3	Benchmark	76
4.4	Uncertain returns	79
4.5	Adverse selection	82
4.6	Coordination and matching	85
4.7	Conclusions	90

Chapter 5:

	Leviathan Governments and Public Debt	92
5.1	Introduction	92
5.2	The model	94
5.3	Optimal debt limits	105
5.4	Infinite horizon	110
5.5	Conclusions	111

Chapter 6:

Summary and Conclusions	113
6.1 References	118

Appendix A:

Appendix for Chapter 3	125
A.1 Tagging and welfare	125
A.2 Proofs	128

Appendix B:

Appendix for Chapter 4	139
B.1 Screening and the reduced form	139
B.2 Proofs	140

Appendix C:

Appendix for Chapter 5	143
C.1 Proofs	143

List of Tables

3.1	Optimal policy under dynamic and static objectives	65
-----	--	----

List of Figures

3.1	$\frac{w_r(1)}{w_r(1-\delta_E)}$ as a function of r and α	38
3.2	$\frac{w_r(1)}{w_r(1-\delta_E)} = \sqrt{2}$	39
3.3	Case 1	54
3.4	Case 2	54
4.1	Specialized system	77
4.2	General system	78
4.3	Specialized system with uncertainty	80
4.4	General system with uncertainty	81
4.5	Specialized system with uncertainty and screening costs	83
5.1	Saving behaviour	106
A.1	Tagging case 1	127
A.2	Tagging case 2	127

Chapter 1

General Introduction

Spending on government programs represents a huge portion of GDP in all modern economies. Each individual's contribution varies, according to their resources and the existing tax system. What determines an optimal policy will depend on the resources in the economy, along with the objectives of the policy maker and the information available to them. We are interested in analyzing the effects that various policies will have on social welfare, in an economy where there may be a variety of imperfections caused by informational asymmetries, the process of job search and matching as well as politics.

In much of the literature on taxation and redistribution, the relevant heterogeneity between individuals is due to differences in their productivity or human capital. Presumably there is a direct link between productivity and education and in chapters three and four we consider education policy specifically.

Chapter three considers education spending in the standard non-linear taxation framework originally developed by Mirrlees (1971). In this set-up, individual type is not observable to the tax setter, which introduces an information friction and results in a second-best outcome. In the standard model, the distribution of types (productivity), is considered fixed. In this chapter we allow education policy to determine this distribution and thus manipulate welfare outcomes directly through individual productivity, rather than solely through money transfers via the tax system. This approach has received little attention in the literature and has important implications for both education policy and the tax system itself. One reason for the dearth of existing work may be that the effects of changes in relative productivity are not obvious and can be difficult to characterize. This is because changes in productivity will not only affect individual wages, but also the tax system itself and the distortions created by it.

Specifically, we consider a situation where governments can target education funds to observable groups (which are referred to throughout the chapter as low or high productivity neighbourhoods/regions). Targeting funds to these groups has direct impacts through individual productivity and indirect impacts through the structure of optimal taxes and the resulting distortions they bring about. The results suggest that educational transfers *may* be valuable relative to money transfers as a tool to maximize social welfare in the standard consumption/leisure framework.

It seems reasonable that educational transfers that reduce productivity differences, and thus reduce transfers through the tax system, can be welfare enhancing in a second-best world. Indeed, this is true for some functional forms in the model, and thus calls for a more a progressive education policy when there are distortions resulting from the tax

system. This is not true generally, however, and there are a couple of reasons for this. The first and most obvious, is that spending on lower productivity types is inefficient, simply because it results in less human capital on aggregate. Secondly, and less obvious, is that such transfers may actually make it more difficult for a planner to differentiate types, and thus increase the distortions to individual labour supply caused by the tax system. In this case, optimal education policy will be more regressive in the presence of tax and transfer redistribution schemes.

Often the justification for policies that create greater equality in educational outcomes go beyond simple consumption/utility arguments, so we interpret these results with a grain of salt. For example, we may consider human capital as intrinsically valuable, or consider the impacts of the distribution of human capital on externalities such as crime, etc. However, presumably consumption and leisure are still important to any social objective and the analysis points to a possible conflict or harmony between the equalization of educational attainments and utility levels.

A further caution regarding so-called equal opportunity policies is found when we extend the analysis to include dynamics. We see that “equalizing opportunities” (which amounts to equating productivity across regions in the model) is not optimal even when it is deemed so in a modified static problem. This is because people are free to choose where they live and any attempt to redistribute to lower productivity regions is confounded when those with high human capital choose to leave the region after receiving education.

Chapter four considers education policy from a different perspective. Here we allow for the reality that individuals do not have perfect information about their own abilities and career suitability. Individuals must choose their educational paths under uncertainty

and the choices they face will depend on the educational institution under which they find themselves. We are interested in the value of education systems that allow individuals to learn more about their abilities, relative to those that focus on enhancing a given skill. This issue has been ignored thus far in the literature on human capital and has interesting implications for the design of education policy and institutions.

In particular, we consider the relative merits of two institutions we call specialized and general school systems. In the model, individuals are heterogenous with respect to their talents and are also initially uninformed about their particular draw from the talent distribution. We characterize a fundamental trade-off between various educational system designs that weighs the returns to enhancing specific skills with the benefits of a broader education that reduces risk in skill attainment.

The trade-off described above is characterized in a frictionless labour market which we refer to as first-best. As we are often interested in labour market frictions, we revisit this trade-off in a second-best world. The second-best results from both informational asymmetry as well as matching frictions between potential workers and employers.

Informational asymmetry arises because firms cannot perfectly observe individual productivity. The framework used to model asymmetric information is a reduced-form version of the standard screening model originally found in Rothschild-Stiglitz (1976). Our model predicts that a specialized system imposes a larger adverse selection cost than a general one. This is because such a system will generate ex-post differences amongst ex-ante identical individuals.

Along with the adverse selection problem, we also consider another labour market friction which is especially relevant to the discussion of education systems. This is the

existence of coordination failures that can arise when individuals seek employment. Using a simple version of the framework outlined by Montgomery (1991), we consider the impact of different educational institutions have on this labour market friction, and find that a general system leads to a larger coordination failure than a specialized one. This is because, due to the general education, we have individuals who are indifferent over which sectors of the labour market to enter, which exacerbates existing coordination problems.

In chapters three and four we consider optimal education policy through the division of a fixed education budget across groups with observable characteristics, as well as the educational institution itself, focusing on various frictions that can exist in the economy. Throughout the analysis, we assume policy makers' preferences are captured by some Social Welfare function. While this approach can be very fruitful, we also recognize that politics play an important role in policy formation, and in chapter five we consider a world where policy choices (such as education budgets) are made by elected officials, who can be only partially controlled through the electoral process. So again, we are interested in the second-best, which arises here due to politician self-interest rather than market imperfections.

In the model we analyze the dynamics of public spending, taxation and debt in a political agency model similar to that originally developed by Barro (1973) and Ferejohn (1986). Policy choices are made by short-lived politicians who can be only partially controlled through the electoral process. The main focus is to consider the impacts of binding limits on the public budget. Examples of such restrictions can be seen through balanced budget requirements that are imposed on various governments. While these types of restrictions are often discussed (and indeed exist), there is little in the way of theoretical

support for such arguments. This chapter is an attempt to fill this gap.

Unlike the chapters three and four, where frictions limited policy makers to second-best outcomes, here it may be that imposing another friction on the model can be welfare improving. It turns out that the value of such a limit depends both on the extent to which politicians' goals deviate from their constituencies and how effectively the electoral process disciplines them when they misbehave. We also see that the value of such a restriction depends on the fiscal position at the time in which it is imposed.

The remainder of the thesis proceeds as follows: Chapter two presents a review of the relevant literatures addressed in each of the following chapters. Chapter three considers education policy in a standard optimal tax framework. The fourth chapter considers the merits of various educational institutions when there are various labour market frictions that arise from informational asymmetries and the imperfect matching of workers and employers. The fifth chapter models politics explicitly and makes budget decisions endogenous, with the goal of analyzing optimal debt controls.

Chapter 2

Literature Review

This dissertation touches on a variety of different areas. Chapters three and four contribute to the literatures on optimal taxation, job search and matching, equality of opportunity and the economics of education. Chapter five is a contribution to the literature on political agency. We now go through each in turn.

Education, taxation and redistribution

Mirrlees (1971) describes an optimal taxation problem that has formed the backbone of a large literature on redistribution. One of the objectives in chapter three is to contribute to this body of work.

In the standard non-linear tax model, there is a heterogeneous population who differ in their productivity, which is defined by some fixed distribution. These individuals work

and consume according to their tastes, productivity and the existing tax scheme. A planner anticipates individual behaviour and employs an unrestricted tax and transfer function to optimize its objective, which is defined by a concave social welfare function. It is commonly assumed that individual types are unobservable to the tax authorities, which requires that the optimal tax scheme be incentive compatible (so that individuals of each productivity type must find it optimal to choose the bundle assigned to them). Incentive compatibility generally leads to a second-best outcome, as incentives to work must be distorted at the optimum.

As mentioned, the distribution of skills is generally considered fixed, which is relaxed in this chapter. We describe an education technology that defines this distribution explicitly. In chapter three we use this extension of the standard framework to further investigate redistributions through individual skills or wages *as well* as taxes. This obviously has implications for the literature on optimal taxation, but can also be viewed as a contribution to the discussion on equality of opportunity, as is discussed below.

With asymmetric information, redistributions through the tax system are generally costly, and policy must weigh any distortionary effects against the benefits, which depend on the planner's social welfare function (aversion to inequality). This is the often discussed efficiency/equity tradeoff. A legitimate criticism of this type of analysis is that there is rarely, if ever, much that can be said about the degree of inequality aversion. In particular, it depends crucially on the reasons for different outcomes as well as the way in which society aggregates individual preferences. Success or failure can be the result of luck, talent, effort, etc., each of which have different implications for any redistributive policy. We simply do not know why people have different outcomes, and this seriously

limits the value of any proposed scheme designed to facilitate greater equality. Even in the extremely unrealistic case of perfect information, it is difficult to imagine a consensus on exactly what is deserving of transfers and what is not (i.e., how do we define a social welfare objective?).

One way in which to address this criticism, and a motivation for the analysis in chapter three, is to use what is referred to as an equal opportunity objective. The idea of equal opportunity, although not a new concept, is not very well understood. This chapter is an attempt to help shed some light on this issue as well.

Equal opportunity has evolved over the years and earlier treatments such as Dworkin (1981) and Sen (1985, 1995) have been built upon by Roemer (1998), Fleurbaey and Maniquet (2005), among others. Following Sugden (2004), we can decompose the basic ideas behind equality of opportunity in two ways. First, “starting-line equality” is the view that every person should have access to the same set of options from which one’s path can be chosen. This is stronger than just a lack of discrimination and generally implies a redistribution of resources. The second is that equal efforts should yield equal rewards (a principle Sugden argues is incompatible with a market economy). Throughout chapter three, by equal opportunity we shall be referring to the relatively weak concept of “starting-line equality”, the exact meaning of which will be made precise below. This will amount to equalizing the *distribution* of outcomes across the observable characteristic, which we refer to as the region. Fundamentally we wish to equalize productivities to the extent possible. With perfect information, this is equivalent to equalizing productivity completely over all individuals (all of whom have identical preferences). When there is unobserved heterogeneity across groups this is not possible, and equal opportunity in this case amounts

equalizing the distributions of skills across an observable characteristic (region or group etc.).

Although we are interested in the idea of equal opportunity, as discussed earlier the analysis in chapter three uses the standard non-linear tax model. However, the framework we impose will allow us to capture an equal opportunity objective through a specific parameterization of the typical social welfare function. Thus we can consider different objectives using the orthodox approach and contrast the two in a relatively simple way. This is a particularly nice, unique feature of the model developed here that doesn't exist elsewhere.

Theoretically, heterogenous outcomes can be divided into two categories. First, there are those such as preferences, for which we are deemed "responsible", and do not call for compensation. Second, there are those for which we are held "not responsible", such as talent or family wealth, which may call for compensation. Fleurbaey and Maniquet (2005) provides a comprehensive survey of these ideas. In chapter three we assume that there is an *observable* difference between the initial conditions of the two groups (e.g., rich and poor) for which they are not responsible. This allows us to consider an equal opportunity objective in the standard setting as we shall see. Specifically we let each individual transform educational resources into human capital with varying effectiveness, depending on both the neighbourhood or region in which they attend school (which is observable) as well as their individual productivity (which is not directly observed, however the distribution of types is known as well as individual income).

There is a vast empirical literature on the impact of "local effects" on children's academic outcomes. Examples include Goux and Maurin (2007), Ding and Lehrer (2007) or

Hoxby (2000). Related theoretical contributions that provide explanations for this include C. de Bartolome (1990) and Benabou (1996a,b). Although it is generally not clear what one is “responsible” for, we take the view that the positive impact that parents’ socioeconomic status has on their children’s educational outcomes is “unfair”. This seems to be one instance where there is a general consensus and can be seen in policies such as the “Head Start” program in U.S.

If a rich child is more productive than a poor one, then under an incentive-compatible tax scheme the former are better off in terms of both standard utility and any non-pecuniary benefits of education (other things equal). The planner does not know why you are productive or not ex-post, and so even with redistributive transfers those born poor are always at a disadvantage. Throughout we will assume that the tax system is not conditioned on where you were educated (as in the literature on tagging), although relaxing this assumption is considered in the appendix to the chapter. In fact, even if the government could distinguish those born poor, it is not clear that they deserve compensation ex-post because their outcome is influenced by other factors that may be just as arbitrary as parents’ income, like talent or luck. Consider a talented and wealthy person who came from modest beginnings. Should they be compensated at the expense of those who were born privileged, but have low talent (or bad luck) and are relatively poor? Justification for such a transfer ex-post is unclear. In particular, if such a transfer were possible, it would violate the so-called principal of horizontal equity, but compensation may be warranted ex-ante. Finally, we note that this type of redistribution may be more politically feasible than money transfers. While politics are not modeled here, this is certainly of importance.

There is a large literature on education and redistribution that is relevant to the analysis

here. In a seminal paper, Arrow (1971) analyzes the division of spending when there is no further redistribution and describes the progressivity/regressivity of optimal policy. In later papers (as in this thesis), education is assumed to take place in some initial period and taxation/redistribution in a second stage. This framework generally implies that the optimal policy is that which maximizes the size of the “pie”, which is redistributed through the tax system as in Bruno (1976), Ulph (1977) and Ulph and Hare (1979). As increasing the pie often means spending on those with better endowments, these models imply more regressive education spending.

A more recent paper by Cremer et al. (2008), argues further value in regressive educational expenditures, and is very closely related to the analysis here. This is the only other paper we’re aware of that directly treats individual productivity as endogenous in the optimal non-linear tax problem.¹ Due to the difficulty in deriving explicit solutions, they are forced to make comparisons of specific policies. Their results are restricted further by some very strong assumptions about human capital production as well as the focus on only two types of individuals. Regardless, they show that inequality in productivities can be welfare improving due to non-convexities in the optimal tax problem (which is central to the results in this thesis as well). The intuition is that, although a more equal distribution of productivity implies less absolute redistribution, that redistribution which is undertaken is more distortionary, substantially so, as incentive compatibility constraints tighten.

Chapter three can be seen as complementary to this work and differs in a variety of ways. Due to the difficulty that arises from a general specification of preferences, in this thesis we restrict attention to quasi-linear-in-consumption utility functions. Further, we

¹Fleurbaey et. al (2002) consider a related model in which policy influences productivity indirectly, which we discuss below.

allow for a more general human capital function and heterogeneity within groups targeted for education funds, rather than just two individuals (although the two-type case is considered in-depth). Given this framework, we are able to derive globally optimal policies for specific functional forms (rather than just compare extremes) and consider equal opportunity arguments in greater depth. Our results suggest that distortions caused by the tax system may in fact result in a more *progressive* optimal education policy in some cases, which is contrary to previous results . However, our analysis is compatible with that in Cremer et. al (2008), and we derive sufficient conditions for their results in our framework. Which again suggest more regressive education funding.

In a similar spirit, Krause (2006) shows that subsidies for skilled education can be consistent with an optimal redistributive program, as these may facilitate incentive compatibility in labour supply decisions.

In another related paper, Fleurbaey et al. (2002) use a mechanism-design approach to consider the progressivity/regressivity of schooling expenditures where ability is private information and agents with different talents invest in education. The planner is endowed with a fixed sum of money which is distributed either in cash or in kind (through educational “help” which reduces the cost of acquiring education). Thus, individual productivity is endogenous here as well, just indirectly.

They show that, when effort and help are substitutes, the second-best optimum involves more education spending on low-types (at the expense of less effort from these types), along with money transfers to reduce consumption inequality. If effort and help are complementary, then the high-types receive more education help. Thus the regressivity/progressivity of education spending depends on the particular specification. In either

case, however, education levels and individual efforts increase with talent. Importantly they find that the more averse to inequality is the planner, the more inequality we see in educational attainment. So in this case, equalization of individual educational attainments may conflict with that of utility levels, albeit for somewhat different reasons.

Finally, we note that the impacts of taxation on educational investments is an important issue, but one not considered here, as we focus on early education and do not allow for individual choice over the amount of human capital. It should be noted that ignoring individual investments avoids the reality that governments are unable to commit to a tax scheme in the future. Boadway et al. (1996) show that, when government cannot commit to tax policy once education is obtained, the result is too little investment in human capital. Similarly, Bovenberg and Jacobs (2005) argue for education subsidies as a way to alleviate tax distortions to human capital accumulation (although there is no issue of commitment in their model).

Educational institutions and labour market frictions

Often, discussions of education policy revolve around the division of a budget (as in chapter three). In chapter four we consider the impact of the educational institution itself. The institution becomes important when we relax the unrealistic assumption that individuals are perfectly informed about their endowments when making educational choices. The timing and set of choices available to students will in turn influence the distribution of human capital, which is of course of great importance to policy makers. Uncertainty over type, and in particular the importance of institutions is something that has not been

discussed in the literature thus far.

There is a vast literature (most of which is empirical) on human capital that addresses a broad array of related issues. Typically, economists focus on the impacts that education spending may have on the growth and distribution of income. Examples include Heckman and Krueger (2003) and Gradstein, Justman and Meier (2005). Some authors have considered the structure of the education system explicitly. For example, Krueger and Kumar (2004) develop a model where there are varying degrees of what they refer to as vocational and general education. Their results suggest that Europe's relatively poor performance vis a vis the U.S may be partially explained by an education system that favors the former. The analysis of education institutions themselves, however, seems to be confined to this simple dichotomy.

A somewhat related literature is that on "streaming" or "tracking". The value and fairness of such systems, which separate students based on some measure of innate ability, have sparked much debate in recent years. Some examples include Epple, Newlon and Romano (2002), Brunello and Giannini (2001), and Bertocchi et al. (2004).

Rather than imposing a mechanism that sorts individuals by talent, we consider a system that gives all individuals access to the same resources regardless of their endowments. What we label a specialized system gives individuals complete flexibility in choosing the focus of their studies. For example, one may choose a course of study that specifically favors analytic over verbal acumen. On the other hand, what we call a general system is a fixed curriculum that requires study of all subjects by all students (i.e., no choices). In a sense our environment allows for a deeper analysis of what is considered "general education". The resulting outcomes and disparities between individuals can be attributed to a

different process than that found in existing models and has different implications for the size and distribution of productive resources.

MacDonald (1980) provides an interesting analysis that considers the role of education in the acquisition of information (which is an important feature of our model in chapter four). He considers the matching of firms and workers with different productivity characteristics, taking the view that each group of individuals has some comparative advantage in a particular task. Emphasis is on the acquisition of education to increase information about type, which may lead to better labour market matches. His paper describes a different type of friction than ours and does not consider the importance of institutions as is the focus here.

As mentioned earlier, in chapter four we consider institutional structure when there are both perfect and imperfect labour markets. Allowing for such imperfections provides further insight into these issues and we consider two types of frictions specifically. The first is asymmetric information and we use a reduced-form version of the classic Rothschild-Stiglitz (1976) framework. Essentially, we allow for the fact that employers do not have perfect information over the abilities of prospective employees, and must undergo some costly screening practice to reveal the types. As the labour market is assumed competitive, the cost is borne by the high types through a lower wage.

The second friction we consider in our analysis is particularly relevant to the discussion of educational institutions. This is a coordination problem, which is originally modeled by Montgomery (1991). In this analysis, firms post wages to attract workers and workers in turn apply to a job. As there is no coordination between workers or firms, there is no way to ensure that all vacancies are filled in a timely manner, which creates a cost. The size

of this cost depends on the size of the pool of potential applicants, which in turn depends on the education system as we shall see. We use a simple version of Montgomery's model to characterize the extent of any coordination failure, and we contrast the specialized and general institutions accordingly. This approach seemed the most appropriate, although a number of interesting frictions were considered from the literature on employment search and matching, an excellent survey of which can be found in Rogerson, Shimer and Wright (2005). We leave further development of these ideas to future work.

Politics and debt

In the final chapter of this dissertation, we relax the assumption that policy makers are benevolent planner's optimizing a welfare objective. This section of the dissertation is a contribution to the literature on political agency. In particular, we are interested in the validity of arguments for constitutional debt restrictions on incumbent governments. Although such arguments are commonly discussed and in fact have been imposed in a variety of cases, they generally lack a theoretical foundation, which we attempt to rectify here.

That politicians can rarely, if ever, be suitably labelled benevolent planner's is the subject of a large literature on which this chapter draws. For a survey of these issues see, Besley (2006) or Persson and Tabellini (2000). In particular, we consider the behaviour of taxes and government debt within the widely-used electoral accountability model originally developed by Barro (1973) and Ferejohn (1986).

In this framework, there is an incumbent that is drawn from of a group of homogeneous

politicians, who provide effort in the production of public services at a cost to themselves. Their efforts are assumed indistinguishable from random economic fluctuations and that in the interest of retaining office, they must “behave”. The voters are modeled as a single individual (they all prefer the politician to do as much as possible). Voting behaviour and the equilibrium services provided are quite simple, in that both are characterized by a voting rule which the voters are assumed to agree upon. It is further assumed that the incumbent has correct expectations about the rule. The rule is defined such that if the incumbent provides services equal or above that described by the rule, they are re-elected. Otherwise they are ousted and an identical challenger is brought in.

In the analysis of this thesis, we use a somewhat modified version of the above and assume that the policy maker derives utility from spending, rather than shirking their duties, a so-called leviathan government. Further, we allow politicians to borrow and save as well as generate tax revenue. The voting rule here is a tax, borrowing and spending level such that should the incumbent deviate from the rule, they are ousted in favour of an identical challenger. If incumbents are indeed inclined to spend too much, a natural question to ask is whether we should constrain them through some sort of constitutional restriction? This is the question we ask in chapter five and the model outlines a simple framework in which it can be addressed.

We note that the focus here is not only in capturing the deviation of government spending from some optimum (generally spending is too high in this framework), but also in the mix of financing between debt and taxation. In this regard, we will employ a tax smoothing approach, as in the classic paper by Barro (1979).

Relevant literature includes a recent paper by Besley and Smart (2008), who consider

the implications that yardstick competition, transparency, tax limits and changes to the marginal cost of public funds have on the political equilibrium when there is both moral hazard and adverse selection. Bassetto and Sargent (2006) consider the value in constraining the government to issue debt only to finance capital items and not for ordinary budgets.

Battagliani and Coate (2008) propose a dynamic model with tax smoothing where public decisions are made via legislative bargaining. In their model, the value of balanced budget restrictions (a special case of the debt controls considered here) depends on the size of the tax base relative to desired public good spending. The larger this ratio, the more beneficial is such a restriction.

Alesina and Tabellini (2007) allow for debt in an agency model very similar to the one presented in this thesis (also using a version of Ferejohn 1986). They use this theoretical framework to explain the existence of pro-cyclical fiscal policy observed in many developing nations. However, due to the nature of information in their model, incumbents borrow as much as possible every period, so debt is not an effective instrument to smooth consumption (which of course doesn't allow for an interesting discussion on debt controls).

Chapter 3

Education vs. Optimal Taxation: The Cost of Equalizing Opportunities

3.1 Introduction

What defines optimal policy depends crucially on the rule by which various outcomes are appraised. Recently, the concept of equality of opportunity as a normative criterion has gained some attention. This standard, it is argued, bridges the gap between egalitarianism and the ideals of responsibility and freedom.¹ We consider the impact of policies designed to equalize productivities or “opportunities”, rather than utilities as in standard welfare economics. Our focus, however, is more general than defining and arguing a specific policy or social objective. We wish to examine the merits of such an approach using the standard tools found in the optimal tax and redistribution literature. In particular, we

¹Although often discussed, there is no general consensus on the exact meaning of the term. For a good discussion, see Roemer (1998) or Fleurbaey and Maniquet (2005).

investigate the solution to an education spending problem, which we can then use to define and analyze policies designed to bring about greater equality in human capital outcomes.

A quality education system has long been seen as integral to the health of a society. Schooling not only increases productivity, but can also facilitate social mobility. In this way, it seems a natural tool for redistribution and especially relevant when considering the importance of “equalizing opportunities” as it directly impacts the distribution of skills.² Indeed, in most societies it is uncontroversial for the state to play an important role in the provision of education. The reasons for and the implications of this, however, are far from obvious. In particular, the distinction between economic and philosophical/ethical arguments is often blurred.

In the model, individuals vary in their innate talents and locations. We define a human capital technology which maps both characteristics, along with public spending on education, into individual productivity. Education policy shapes the human capital distribution, which in turn impacts welfare as defined by the solution to a social planner’s optimal tax problem. We consider optimal education spending when public funds are distributed between two different locations and when one location has more productive students, for example rich and poor neighbourhoods. It seems reasonable that greater equality in human capital between regions may be optimal, as welfare is defined in a second-best world. Greater equality in productivity, which implies less redistribution through taxation, could mean less distortion to labour markets and an increase in social welfare. The results suggest this intuition can be correct; however, we see that the equal opportunity policy, defined

²Many have condemned the often substantial inequality we observe in education budgets across various districts (largely funded by the local tax base) on grounds of efficiency, morality and legality. For a discussion of these and related issues, see Berne (1988) or Fernandez and Rogerson (1996).

as that which compensates completely for local advantages, is never optimal. This is true regardless of the planner's aversion to inequality. In fact, it may be the case that the optimal policy is that which spends nothing in the low-productivity region, *regardless of social preferences*. Although education transfers can mitigate the distortions brought about by taxation, it may not be wholly effective in this regard and greater equality in human capital outcomes may come at the cost of lower social welfare.

Redistributing to individuals from low-productivity areas through education spending may be limited for a variety of reasons. There will be a reduction in the amount available for money transfers as total output declines. Obviously, it is inefficient to spend too many resources on lower-productivity students. Less obvious is the role played by the tax system, and we see that, for a given set of skills, increasing the productivity of low types can tighten the incentive constraints in the tax problem, which leads to *greater* distortions in the labour market. Further, we note that to satisfy incentive compatibility in the tax problem, low types are discouraged from working. Thus any benefit from increasing their relative productivity, which is costly to begin with, is not fully realized.

Education policy obviously has important implications that extend outside of the simple framework used to derive the results discussed above. These include self-esteem, crime, voter savvy, general social cohesion, etc. Such considerations provide justification for equalizing opportunities even if these policies are not optimal in terms of a standard consumption/leisure view of utility. Whether or not we should equalize opportunities for such reasons is a moral/ethical question for which current economic models provide little insight. Such policies may or may not, however, reduce welfare in the standard model,

which is presumably of importance to the debate. We do not attempt to answer this question here, but note that, if such policies are indeed deemed optimal, it is not clear that they remain so when we extend the model to include dynamics. In particular, we see that equal opportunity type policies as defined in the static setting are not consistent with dynamic behaviour. This is because individuals are heterogeneous and free to live where they please. Once educated and working, they will segregate according to income and the problem repeats next period. Previous spending influences the composition of neighbourhoods/regions, and we see that spending “too much” in the low-productivity area is sub-optimal as it leads to lower outcomes in the next generation for *everyone*. If the objective is to sever the link between socio-economic status and human capital outcomes, simply equalizing for local effects in each period is not reasonable. Optimal policy must account for the effects on the long-run distribution of human capital, while being aware of the exodus of high productivity types from the low-productivity area.

In a standard framework, we consider policies which allow us to bolster the relative productivity of those born less fortunate; the question is whether or not this is optimal. The current literature suggests not, when there is a tax system already performing the role of utility equalization. This is consistent with the analysis here, given certain assumptions regarding the disutility of labour as well as the productivity technology. Importantly, there are specifications in our model which suggest tax distortions lead to more *progressive* education policy.

Further, the dynamic analysis suggests that, if we employ an equal opportunity type policy (for reasons outside of the standard redistribution framework like the importance of self-esteem etc.), this may be inconsistent with dynamic goals and should be modified

accordingly. The analysis makes clear the need to be precise about the objectives of education policy and to be aware of the effects it will have on the distribution of skills in a dynamic world. The rest of the chapter is organized as follows. In section 3.2, the basic analytical environment is described. Section 3.3 analyzes education policy in a static model. The framework is then extended to consider dynamic implications in section 3.4, and section 3.5 concludes. All proofs can be found in the appendix.

3.2 Individuals

For each individual, we denote units of labour supplied by l and consumption by c , which consists solely of a single good whose price is normalized to one. Individual utilities are defined by the same quasi-linear in consumption function of the bundle $(l, c) \in \mathbb{R}_+ \times \mathbb{R}_+$,

$$\tilde{u}(c, l) = c - f(l), \tag{3.1}$$

where the function $f(l)$ satisfies $f'(l) > 0$ and $f''(l) < 0$. The labour market is competitive and it is assumed that aggregate production is simply the sum of individual productivity. So once an individual is educated, human capital equals the wage w . At a given wage, each individual optimizes utility given by (3.1) and supplies l units of labour, which implies a pre-tax income of

$$y = wl. \tag{3.2}$$

Pre-tax income could also be considered one's labour supply in efficiency units. Using (3.2), utility can be written in terms of consumption and pre-tax income as is common in

the optimal tax literature

$$u(c, y) = c - f\left(\frac{y}{w}\right). \quad (3.3)$$

3.3 Education policy

Wages and productivity are synonymous and a function of individual characteristics and government education spending. Importantly, we assume that human capital attainment is determined by purely exogenous factors from the individuals perspective³. Education spending is targeted to specific neighbourhoods/regions and takes place in an initial time period, after which students enter the working world. The planner not only makes an education spending choice, but also employs a tax and transfer scheme to achieve redistributive goals between individuals once they enter the workforce. Throughout, welfare is assumed to be aggregated by a standard concave social welfare function. Education spending will impact welfare at the optimum as defined by the tax problem. Note that as a result of the timing, skills are considered fixed for a government solving an optimal tax problem affecting working adults.

Although integral to the results, our focus is not on the tax problem but over how to allocate a fixed education budget B between the two neighbourhoods⁴. Define e_j as the amount of educational spending in location $j \in \{p, r\}$. The subscripts $\{p, r\}$ represent the

³We consider this primary/secondary education and take the view that children's aversion to schoolwork is irrelevant. This is in line with Roemer (1998) and is likely inappropriate if considering higher education.

⁴To determine the size of this budget requires a general theory of social justice, which is ignored here as the focus is on the composition of spending. This was made endogenous in an earlier version and the optimal size of the education budget was seen to be positively related to the total value received from education. In particular, if we impose "sub-optimal" education policies, these are generally accompanied by a reduction in the size of the budget.

low-productivity (“poor”) and high-productivity (“rich”) locations respectively. Let the fraction of B spent in the low-productivity neighbourhood be denoted by δ and normalize B to 1, so that $e_p = \delta$ and $e_r = 1 - \delta$.

Throughout the rest of section 3.3, we describe the education spending problem. The next two subsections consider simplified models in which there are no individual differences, so that there are two productivity types corresponding to the two locations. These frameworks portray the intuition and allow analytic results that are relatively easy to interpret and relate to the current literature. In section 3.3.3 we discuss the more general case where there is individual heterogeneity within each location.

3.3.1 Two endogenous wages

If there is no individual heterogeneity within each location, then human capital is a function solely of the location in which one is educated. Thus there are two types, as there are two locations. The idea of location is only maintained here to be consistent with the later analysis and current literature. The interpretation can be more general and the basic assumption is that low and high productivity students can be observed and funded separately. Once working, students educated in the high-productivity region receive wage $w_r(1 - \delta)$, while those from the low-productivity region receive $w_p(\delta)$. We assume that the functions $w_j(e_j)$ are increasing and concave, that $w(0) = 0$, and that $\forall x, w_r(x) \geq w_p(x)$. Further we assume the wage/productivity technology exhibits a constant elasticity defined as $\epsilon_w = e \frac{w'(e)}{w(e)}$. Before describing the optimal education spending choice, we define the equal opportunity policy δ_E , as that which equalizes the wages of students in the two locations. Thus $w_r(1 - \delta_E) = w_p(\delta_E)$.

First-best

In the first-best, the planner can observe the individual productivity of working adults and make lump sum transfers to achieve any redistributive goals. As individual utility is separable and linear in consumption, labour supply is solely a function of productivity and there is no income effect. Thus, regardless of the degree of inequality aversion, the optimal education policy in the first-best is simply that which maximizes “total surplus”.⁵

$$S = w_p l_p(w_p) + w_r l_r(w_r) - f(l_p(w_p)) - f(l_r(w_r)) \quad (3.4)$$

Differentiating with respect to δ , noting that the indirect impact of changes in δ to the labour supply is zero at the optimum (by the envelope theorem), gives

$$\frac{\partial S}{\partial \delta} = \frac{\partial w_p}{\partial \delta} l_p(w_p) + \frac{\partial w_r}{\partial \delta} l_r(w_r). \quad (3.5)$$

The conditions for a maximum are not necessarily satisfied here, however, and in fact often will not hold. The second derivative is

$$\frac{\partial^2 S}{\partial \delta^2} = \sum_{j=p,r} \left(\frac{\partial^2 w_j}{\partial \delta^2} l_j(w_j) + \left(\frac{\partial w_j}{\partial \delta} \right)^2 \frac{\partial l_j(w_j)}{\partial w_j} \right). \quad (3.6)$$

The first term in the sum is non-positive and the second positive. Thus the second-order condition implies a minimum if the curvature in $w_j(e_j)$ is not too strong relative to the disutility of labour. For example, when productivity is linear, $\partial^2 w_j / \partial \delta^2 = 0$, so we have a minimum and the optimal policy is $\delta = 0$.⁶

⁵There are no incentive problems and money can be transferred freely. As labour supply is not dependent on the income level (which includes the transfer), the objective is simply to maximize “total output” less “total labour disutility”, regardless of the degree of curvature in the social welfare function.

⁶Cremer et al. (2008), show with general individual preferences, that when the wage technology is linear, wage differentiation ($\delta = 0$) is preferred to wage equalization, even when there are no local advantages.

Second-best

In the first-best, we may redistribute incomes without imposing any efficiency costs. Now we consider the imperfect information case, where only income (and not wages or labour supply) is observable, and thus any redistribution scheme must be incentive compatible. This is the situation considered in most of the optimal tax literature. In our case, this is equivalent to the assumption that in the initial stage, we can spend on education knowing which is the high and low productivity type, but once they enter the workforce we can no longer differentiate between the two. If we allow for the planner to observe each type in the second stage, then we are back to the first-best where the planner imposes lump sum transfers. For the more general case in section 3.3.3, where there is individual heterogeneity, the results of relaxing this are less obvious and this is discussed in Appendix A.⁷

We see in the first-best it is optimal to maximize the size of the “cake” which may involve spending nothing (or certainly less) in the poor neighbourhood depending on the assumptions made regarding human capital acquisition. This strategy may no longer be optimal here as inefficiencies from redistribution will eat into the “bigger cake”. It is conceivable then that education spending be used to redistribute if inefficiencies from transferring utilities through the tax system are large relative to those brought about by transferring productivities. Before looking at the impacts of education on social welfare, we must first consider the second-best tax problem which defines it. Social welfare can be represented by the solution to the following Pareto-optimizing problem.⁸

⁷This issue is considered in the literature on tagging; see for instance Boadway and Pestieau (2006)

⁸The optimum is determined given the utility of the high type, characterized by m . Changes in m amount to changes in the planner’s aversion to inequality and this maps out the utility possibility frontier.

$$\max_{c_p, c_r, y_p, y_r} c_p - f\left(\frac{y_p}{w_p}\right) \quad \text{subject to} \quad (3.7)$$

$$y_p + y_r \geq c_p + c_r \quad (3.8)$$

$$c_r - f\left(\frac{y_r}{w_r}\right) \geq c_p - f\left(\frac{y_p}{w_p}\right) \quad (3.9)$$

$$c_r - f\left(\frac{y_r}{w_r}\right) \geq m \quad (3.10)$$

where (3.8) and (3.9) are the budget and incentive constraints respectively and m is the predetermined utility level of the high type. The incentive constraint requires that the optimal allocation for a high type provide at least as much utility as is attained by the high type mimicking the low (as types are not observable). Constraint (3.10) implements social preferences by forcing a minimum utility requirement for the high type. The corresponding lagrangian expression is

$$\mathcal{L} = c_p - f\left(\frac{y_p}{w_p}\right) + \lambda(y_p + y_r - c_p - c_r) + \gamma\left(c_r - f\left(\frac{y_r}{w_r}\right) - c_p + f\left(\frac{y_p}{w_p}\right)\right) + \quad (3.11)$$

$$\mu\left(c_r - f\left(\frac{y_r}{w_r}\right) - m\right). \quad (3.12)$$

For example, if $m = 0$ we have a maxmin or Rawlsian objective function.

The optimum is described by the constraints and the first-order conditions which reduce to

$$\gamma = 1 - \lambda, \quad \mu = 2\lambda - 1 \quad (3.13)$$

$$\frac{f'(\frac{y_p}{w_p})}{w_p} = \lambda + (1 - \lambda) \frac{f'(\frac{y_p}{w_r})}{w_r} \quad (3.14)$$

$$f' \left(\frac{y_r}{w_r} \right) = w_r. \quad (3.15)$$

Note that (3.15) represents the familiar “no distortion at the top” condition and, given the quasi-linear form for utility, completely determines y_r . Using the optimal conditions, we can derive the following comparative static results:

$$\frac{\partial y_p}{\partial m} = \frac{2}{\left[1 - \frac{f'(\frac{y_p}{w_r})}{w_r}\right]} > 0 \quad (3.16)$$

$$-\frac{\partial \gamma}{\partial m} = \frac{\partial \lambda}{\partial m} = \frac{2 \left(\frac{f''(\frac{y_p}{w_p})}{w_p^2} - (1 - \lambda) \frac{f''(\frac{y_p}{w_r})}{w_r^2} \right)}{\left[1 - \frac{f'(\frac{y_p}{w_r})}{w_r}\right]^2} > 0. \quad (3.17)$$

These inequalities hold as $f(l)$ is convex, $w_r \geq w_p$ and we know that $\lambda \in [\frac{1}{2}, 1]$ from the first-order conditions. We see that as inequality aversion increases (m decreases), the before-tax income of the low type decreases as labour supply is more distorted. Also, we see that λ is increasing in m which implies that γ is decreasing in m . As would be expected, the incentive constraint tightens as redistribution increases.

The optimal tax scheme characterized above is the subject of a vast literature which we will draw on, but is not the focus here. Rather, our interest is on the social welfare impacts

of manipulating the distribution of wages/productivities through the education system.⁹

The impact of a change in δ on welfare, at the optimum as defined by the tax problem is

$$\frac{\partial \mathcal{L}}{\partial \delta} = \underbrace{f' \left(\frac{y_p}{w_p} \right) \frac{y_p}{w_p^2} \frac{\partial w_p}{\partial \delta} + \mu f' \left(\frac{y_r}{w_r} \right) \frac{y_r}{w_r^2} \frac{\partial w_r}{\partial \delta}}_{\text{Direct impact} \geq 0} + \underbrace{\gamma \left(f' \left(\frac{y_r}{w_r} \right) \frac{y_r}{w_r^2} - f' \left(\frac{y_p}{w_p} \right) \frac{y_p}{w_p^2} \right) \frac{\partial w_r}{\partial \delta}}_{\text{Impact on the incentive constraint} < 0}.$$

(3.18)

Education policy affects welfare directly through the first term which may be positive or negative as the impact on the poor and rich depends on the functional forms and the planner's aversion to inequality. It is this term that represents the possible value in spending on the poor. If spending in the poor region is not too inefficient and the planner has a high degree of aversion to inequality, this term will be positive (with a maxmin this is unambiguously positive as $\mu = 0$). The effect on the incentive constraint, however, is always negative.

Defining the optimal education policy in the general problem is often not feasible. However, It is interesting to consider the impacts of deviations from the first-best policy, in a second-best environment. From this, we can see how changes in spending from the first-best optimum will affect social welfare when individual productivity is unobservable to the policy maker. We have the following result:

⁹Recent related papers by Brett and Weymark (2007) and Simula (2007) consider the impact of changes in productivities to the optimal tax scheme in a model with quasi-linear preferences and a weighted utilitarian objective function (the former linear in leisure, the latter in consumption).

Proposition 1 *Denote the solution to the first-best problem as δ^{fb} (which maximizes total surplus).*

- 1. When $\delta^{fb} = 0$, increasing spending in the low-productivity region beyond the first best, reduces welfare in the second-best.*
- 2. If $\delta^{fb} > 0$, so that some spending on the low-productivity region is optimal in the first-best, then increasing spending from the first-best increases welfare in the second-best.*

Proof: See appendix.

Proposition 1 tells us that increases in spending on the low-productivity region beyond the first-best will increase welfare in the second-best, unless the low-productivity type is producing nothing. We see the presence of distortions in the economy does not necessarily lead to a more regressive education policy, as in previous research like Cremer et. al (2008). In their analysis, regressive policies are optimal because the first-best is always zero. We see that when we allow for a more general productivity technology, education spending can have benefits in a second-best world that don't exist when individual productivity is observable in the tax problem.

It is also interesting to consider how the impacts of policy changes differ for various levels of inequality aversion. Using again the conditions for optimality in the tax problem, we have

$$\frac{\partial^2 \mathcal{L}}{\partial \delta \partial m} = \frac{\epsilon_w}{\delta(1-\delta)} \left[\underbrace{\frac{\partial y_p}{\partial m} \left(\frac{2y_p}{w_p^2} - \delta \lambda \right)}_{>0} + \underbrace{\frac{\partial \gamma}{\partial m} \delta (y_p + y_r)}_{<0} \right] \quad (3.19)$$

where $\epsilon_w = \frac{e}{w(e)} \frac{\partial w}{\partial e}$ is the elasticity of productivity with respect to education funding (which is assumed to be constant). We see there are two conflicting forces influencing the size of the impact of education policy changes. Neither of these effects dominates for all δ , so it is impossible to describe the importance inequality aversion has on a change in δ without making further assumptions. We can, however, gain some intuition by examining these effects in turn. The first is that the higher is m , the less redistribution and thus the less distortion on the low-wage type. Therefore, changes in δ have a larger impact on the poor because they are producing more. Noting that the “no distortion at the top” condition (3.15) implies there is no impact on the output of the high type. Secondly, as increases in δ tighten the incentive constraint, distortions are smaller when m is larger, as there is less redistribution.

Numerical example

If we are to say anything more concrete about the impacts of education spending, we must first characterize consumption and income under the optimal tax scheme. To this end, we impose some further structure on the problem which is assumed throughout the rest of this subsection (the importance of these assumptions and the consequences of relaxing them are discussed below). Let the disutility of leisure be given by $f(l) = l^2/2$, which allows us to solve the optimal tax problem explicitly. In accordance with proposition 1,

we consider the cases where the first-best policy is at the corner ($\delta = 0$) and then that in which the first-best policy has some spending on the low-productivity.

First-best: $\delta = 0$

First, we make the following assumption regarding the wage technology, which is a sufficient condition for a corner solution of $\delta = 0$ to the first-best problem (described in equation 3.5).

$$\frac{\partial w_j(e_j)^2/e_j}{\partial e_j} \geq 0. \quad (3.20)$$

Further, we assume the following, which is often implied by the latter condition on $w_j(e_j)$.

$$\frac{w_r(1)}{w_r(1 - \delta_E)} \geq \sqrt{2} \quad (3.21)$$

The inequality (3.21) holds when the efficiency loss from an equal opportunity policy is “large enough”. In particular, when (3.21) is satisfied, welfare under a Rawls (maxmin) objective is higher for $\delta = 0$ than for $\delta = \delta_E$ (see appendix for details). It would seem that (3.21) will cease to hold if diminishing returns are high enough. However, δ_E is also increasing with the curvature in $w_j(e_j)$, so it is not immediately obvious when this will be violated. In fact, inequality (3.21) is satisfied by a large class of functions as discussed below in section 3.3.1. We have the following result.

Proposition 2 *Given the assumptions above, for both extremes of inequality aversion (utilitarian and maxmin), the optimal education policy spends nothing in the low-productivity region ($\delta = 0$).*

Proof: See appendix.

Our assumptions allow us to go beyond local optimality conditions and solve for the optimal policy, at least for the two extremes of the social welfare function. We see that when the first-best policy is $\delta^{fb} = 0$, then all educational resources are better spent in the high-productivity region, which provides a larger cake to divide in the second period. This is true for not only for a utilitarian, who has no preference for redistribution (this is true by assumption here), but also for a maxmin objective which only applies weight on the utility of the low type. Thus, even with extreme aversion to inequality, the value of education spending on low types is outweighed by the costs. For intermediate levels of inequality aversion it is difficult to derive the globally optimal policy, although the results of proposition 2 appear to hold for all planner's preferences between the extremes (the details are discussed in the appendix).

Proposition 2 suggests that when the first-best is at a corner, education spending is inferior to money transfers as a tool for equalizing utility. In fact for a large class of functions, any spending in the low-productivity location is inefficient, even when the planner only values the welfare of these individuals. Spending on the low-types reduces total output (and consequently transfers), as that of high-types is decreased, while that of the low-types increases relatively less. For a fixed set of skills, low-types are discouraged from working by the tax scheme, so the benefit from increasing their human capital is not fully realized.

Finally, increasing the productivity of the low-types tightens the incentive constraint in the tax problem which creates an even greater distortion in the labour market.

First-best: $\delta > 0$

Now consider the case when the first-best policy has at least some spending on the low-productivity type. Obviously, assumptions (3.20) and (3.21), from the last section are violated. Even with simplified functional forms, solving for the optimal education policy is difficult. However, we know from the earlier proposition that increasing δ beyond δ^{fb} is welfare increasing. Further we can show that welfare is increasing in δ at the equal opportunity policy (the upper bound on education spending in the low-productivity region). Thus, if welfare is a monotone function of δ (which is possible but not guaranteed), the optimal policy is the equal opportunity policy. Regardless, the optimal second-best policy in this case spends more on the low-productivity region than in the first-best. More formally we have

Proposition 3 *When the first best policy involves positive spending on the low-productivity type, the optimal education policy for a maxmin (Rawls) objective is the equal opportunity policy. This is true provided the difference between types is not too large.*

Proof: See appendix.

If the initial difference between types is not too large (if it were, the first-best would not be positive to begin with), then a maxmin planner would implement the equal opportunity policy in the second-best. Not only does education spending on the low-types have value

in the second-best, but it is ideal to spend up to the point where wages are completely equalized.

Generally, if diminishing returns in $w(e)$ are large enough and/or the difference between communities is not that large, then a utilitarian optimum (which is equivalent to the first-best described above) will not be at the corner and it is optimal to spend some funds in the low-productivity community. In fact, if inequality (3.21) does not hold, then the optimal second-best policy for a maxmin objectives is indeed to equalize productivity *completely* (which is the equal opportunity policy δ_E). To further illustrate, continue to assume that utility be quadratic and that the wage (human capital) functions are

$$w_p = \delta^\alpha, \quad w_r = r(1 - \delta)^\alpha$$

where the constant $r \geq 1$ represents the advantage of the high productivity region.

First consider the problem facing a utilitarian planner, which is equivalent to the first-best problem. For this specification, total surplus as described by (3.4) is

$$S = \frac{1}{2} \sum_j w_j^2.$$

The second-order conditions imply that surplus S is convex whenever $\alpha \geq 1/2$ and concave otherwise. As mentioned above, if the wage function is linear ($\alpha = 1$) then surplus is convex, but generally the problem is convex when diminishing returns are not too large relative to the disutility of labour.¹⁰ If the problem is convex then the optimal policy is $\delta = 0$. If it is concave, then there is some lower range over which spending on the low type is beneficial. Proposition 2 states that when surplus is convex, optimal policy in the second-best

¹⁰For instance, we could set $f(l) = \frac{1}{\epsilon} l^\epsilon$, where $\epsilon \geq 2$. Under this specification, surplus is convex when $\alpha \geq 1 - \frac{1}{\epsilon}$.

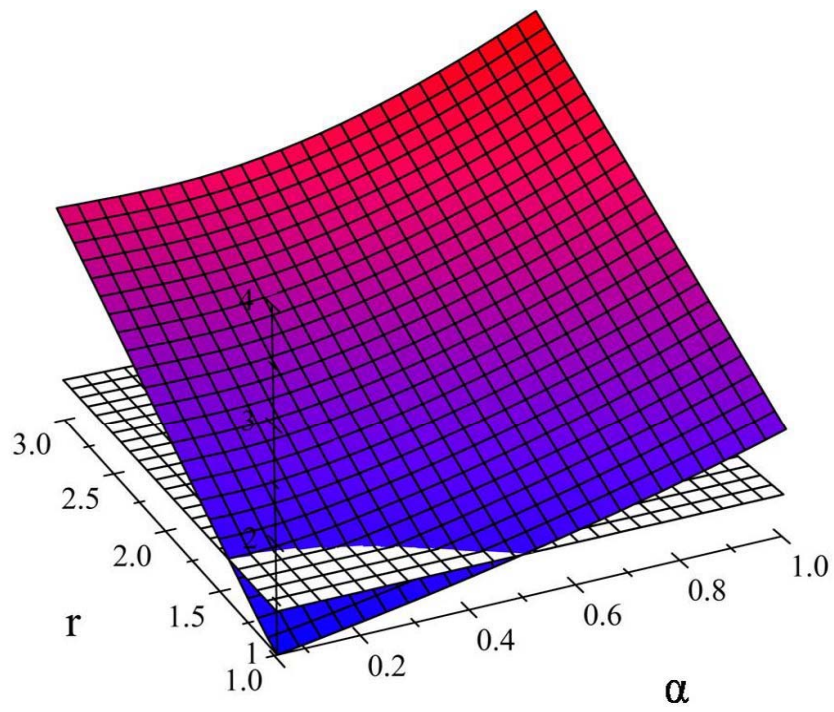


Figure 3.1: $\frac{w_r(1)}{w_r(1-\delta_E)}$ as a function of r and α

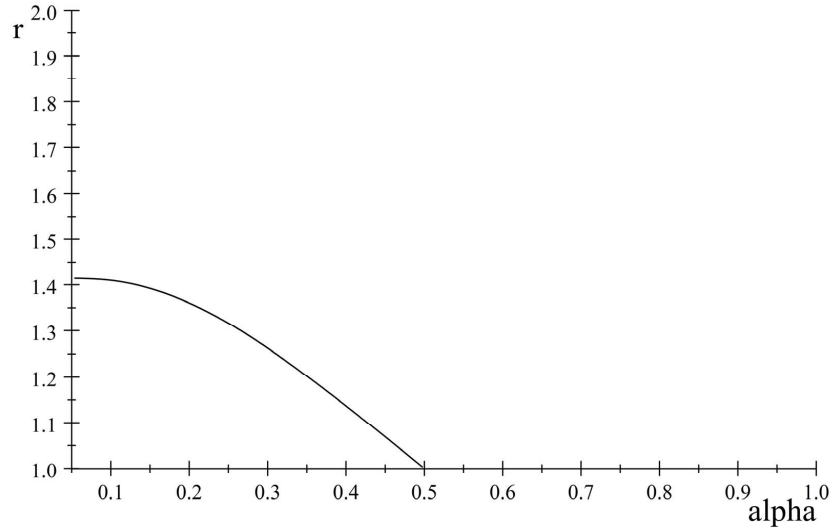


Figure 3.2: $\frac{w_r(1)}{w_r(1-\delta_E)} = \sqrt{2}$

spends nothing in the low-productivity region, *even with extreme aversion to inequality*. However, proposition 3 states that when δ^{fb} is not zero, the optimal second-best policy for a Rawls objective is the equal opportunity policy. For a Maxmin planner, given the functional forms in this subsection, we can describe the regions over which these very different policies are optimal. Figure 3.1 describes the function $w_r(1)/w_r(1-\delta_E)$ over various combinations of α and r . When this function lies above the plane at $\sqrt{2}$, the optimal policy with a Rawls objective is $\delta = 0$. However, the function lies below the plane at $\sqrt{2}$ for some combinations of r and α . For such values, (3.21) does not hold and it is optimal to completely equalize productivity between the two types (δ_E). This range of parameters is easier to see in two dimensions. Figure 3.2 graphs the level curve of the function $w_r(1)/w_r(1-\delta_E) = \sqrt{2}$. With values of r and α below the curve, the difference

between locations is not too large and there is diminishing returns. In this case, with a Rawls objective, equal opportunity (equal productivity) is optimal.

Finally, we note that with these functional forms, when the first-best problem is convex (so that $\delta^{fb} = 0$), then assumption (3.21) *always* holds and the optimum in the second-best is also $\delta = 0$.

3.3.2 Two fixed wages

In this sub-section, we wish to highlight an important feature of policies designed to equalize productivities. This is that the beneficiaries of such policies are generally those in the low-productivity community who are able to make the best use of educational resources (i.e., the “smart ones”). This was impossible in the previous section as individuals were homogenous within locations. We will consider the more general problem where there is individual heterogeneity within locations in section 3.3.3, but we can highlight this particular point very starkly using a simpler model.

Let there be two types of jobs in the economy, high and low skill, which are characterized by two *fixed* wages w_2 and w_1 respectively. The fraction of high and low skill workers in the economy is denoted by π and $1 - \pi$, and is a function of education spending in the initial period. In particular, define the total fraction of skilled workers π as the sum of those originating from the low and high-productivity regions which are denoted by $p(e)$ and $r(e)$ respectively, so that

$$\pi(\delta) = p(\delta) + r(1 - \delta) \tag{3.22}$$

We assume that both $p(e)$ and $r(e)$ are increasing and concave in education spending e so that the total fraction of skilled workers π is concave in δ . Although this is a reduced

form approach and this is not explicit, the interpretation is that those on the “higher end of the distribution” within each location will attain a skilled job. Further, it is assumed that $r'(e) > p'(e) \forall e$ so that with the same resources, students from the high-productivity (rich) area are more likely to attain a skilled job.

First-best

The first-best outcome is again that which maximizes “total surplus”, which in this case is given by

$$S = (1 - \pi)(w_1 l(w_1) - f(l_1(w_1))) + \pi(w_2 l(w_2) - f(l_2(w_2))). \quad (3.23)$$

In the interior, the first-order condition describes the first-best δ^{fb} .

$$\frac{\partial S}{\partial \delta} = \frac{\partial \pi}{\partial \delta} (w_1 l(w_1) + w_2 l(w_2) - f(l_1(w_1)) - f(l_2(w_2))) = 0. \quad (3.24)$$

Unlike the previous case, the second-order condition is satisfied and this indeed describes the optimal policy.¹¹ Denote the first-best policy as that which satisfies

$$p'(\delta^{fb}) = r'(1 - \delta^{fb}). \quad (3.25)$$

Note that by the definition of $p(e)$ and $r(e)$, we have $0 \leq \delta^{fb} < \frac{1}{2}$.

Second-best

As in the previous section, the second-best problem is characterized by the solution to the following Pareto-optimizing problem.

¹¹Although if $p(e)$ and $r(e)$ are linear we are at a corner and $\delta^{fb} = 0$.

$$\max_{c_1, c_2, y_1, y_2} c_1 - f\left(\frac{y_1}{w_1}\right) \quad \text{subject to} \quad (3.26)$$

$$(1 - \pi)(y_1 - c_1) + \pi(y_2 - c_2) \geq 0 \quad (3.27)$$

$$c_2 - f\left(\frac{y_2}{w_2}\right) \geq c_1 - f\left(\frac{y_1}{w_2}\right) \quad (3.28)$$

$$c_2 - f\left(\frac{y_2}{w_2}\right) \geq m. \quad (3.29)$$

Where again (3.27) and (3.28) are the budget and incentive constraints respectively and m is the predetermined utility level of the high types. The incentive constraint requires that the optimal allocation for high types provide at least as much utility as is attained by mimicking the low type (as types are not observable). Constraint (3.29) implements social preferences by forcing a minimum utility requirement for the high type. The corresponding lagrangian expression is

$$\begin{aligned} \mathcal{L} = & c_1 - f\left(\frac{y_1}{w_1}\right) + \lambda((1 - \pi)(y_1 - c_1) + \pi(y_2 - c_2)) \\ & + \gamma\left(c_2 - f\left(\frac{y_2}{w_2}\right) - c_1 + f\left(\frac{y_1}{w_2}\right)\right) + \mu\left(c_2 - f\left(\frac{y_2}{w_2}\right) - m\right). \end{aligned}$$

We see that in the first-best, it is optimal to maximize the size of the cake. This may involve spending nothing, or certainly less, in the low-productivity area depending on the assumptions made regarding the function π . Since the wages are assumed fixed, it would seem that the optimal fraction spent on the poor in the second-best problem is the same as that from the first-best, as there is no redistributive motive for spending on the poor (i.e., there is no way to increase the wage of the low type and any deviation from δ^{fb} shrinks the cake that can be redistributed).

Proposition 4 *The optimal second-best education policy is unique and equals that implied by the first-best for any social preferences.*

Proof: See appendix.

The optimal education policy is independent of inequality aversion because the wages are fixed. Thus spending in the low-productivity region above the first-best level is never optimal, regardless of the planner's objective. However, the impact on welfare from various education policies that are not optimal from this perspective *is* differing in the social objective. We saw above that the impact of changes in δ were generally not the same for all social preferences. The exact relationship was ambiguous, however, as the impact, characterized in (3.19), was of indeterminate sign. In this case, however, this is not true, and we can see how the degree of inequality aversion influences the impact of changes in δ . The following proposition assumes that social preferences take an exponential form and in particular exhibit constant absolute inequality aversion.

Proposition 5 *Education policies that spend a sub-optimally high proportion of funds in the low-productivity region have a larger negative impact on low types than high types.*

Proof: See appendix.

An increase in the fraction of unskilled workers causes the utility possibility frontier to shift inward from the laissez faire (utilitarian) point. Low-skill types are hurt more by a deviation from the first-best because it reduces the ability of the planner to redistribute. This is because there are fewer to tax and more recipients. There is no relative impact in

the utilitarian case as there is no redistribution (of course, total welfare is reduced simply because there are fewer skilled types). So we see in this simplified case that increasing education in the poor region shrinks the cake, which has a negative impact on welfare directly as there are fewer skilled types and indirectly as the scope for redistribution is decreased. Not only is welfare decreased but the worst off are hit harder.

If the return to education funding is very low for those on the low end of a distribution of individual endowments, then equal opportunity policies will hurt these types the most. This is because they receive little benefit from education funding (in this case they receive none) and further they see a reduction in money transfers through the tax system. If our redistributive goals are to help the least fortunate (which is implied by most social welfare functions), policies that attempt to equalize opportunities will certainly not be optimal. It may be that the *talented* poor, who will likely achieve a reasonable outcome regardless, are the only beneficiaries of these policies.

3.3.3 Individual heterogeneity

We now extend the model of section 3.3.1 and let individuals differ within each region. Specifically, we allow for two types of heterogeneity amongst people. First, we define the individual-specific endowment θ_i . This can represent innate talents, good family environment, charisma, or any such combination of the like. Let $\theta_i \in \{\theta_l, \theta_h\}$, with $\theta_h > \theta_l$. It is assumed throughout that θ is *not* observable. Secondly, as with the simpler models presented above, individuals differ in their location. We will represent location by the

constants L^j where $L^j \in \{L^p, L^r\}$ and $L^r \geq L^p$. Wages (human capital) are defined by

$$w = \theta_i h(L^j, e_j) \quad (3.30)$$

where the function $h(\cdot)$ is increasing and concave in both arguments, and as above e_j denotes education funding in region j . To reduce notation we define $h_j = h(L^j, e_j)$.¹² In the simpler models we had only two types which were defined by their location. When we allow for individual-specific differences within locations, there are four types of persons with wages $\{\theta_l h_p, \theta_h h_p, \theta_l h_r, \theta_h h_r\}$. We assume that students with no funding have zero productivity so that $h(L, 0) = 0$. Finally let $\partial^2 h / \partial L \partial e > 0$, so that for a given level of resources a rich student has a higher marginal return than a poor one.¹³

Our focus is on the optimal education policy, which again impacts welfare which as defined by the solution to an optimal tax problem. However, as we now have more than two types, we can no longer solve the tax problem using the Pareto-optimizing characterization as above. We thus define social welfare explicitly as

$$\sum_{k=1}^4 \Psi \left(c_k - f \left(\frac{y_k}{w_k} \right) \right) \quad (3.31)$$

where $\Psi(\cdot)$ is a standard concave social welfare function. The tax problem is to optimize social welfare through a tax and transfer scheme for a given set of individual productivities, where individuals are denoted by the subscript k . In particular, we note that redistributions that occur through the tax system are accounting for differences across both L and θ . Spending on education, however, accounts only for differences in L , which is important

¹²In the static model, as L^r and L^p are fixed, this notation is unnecessary, but maintained to be consistent with the dynamic framework of section 3.4.

¹³The impacts of relaxing this are discussed below.

depending on the interpretation of θ . Below, we characterize the optimal δ , but we are also interested in the equal opportunity policy.

Definition 1 *Let the equal opportunity policy, denoted δ_E , be that which satisfies*

$$h(L^p, \delta_E) = h(L^r, 1 - \delta_E). \quad (3.32)$$

The policy δ_E equalizes the local component of productivity so that those with equal θ have the same outcome. This policy accounts solely for differences in location and not for differences in individual characteristics θ . In this sense, δ_E is a somewhat weak definition as it equalizes opportunities between locations, but not between individuals within locations. If we could target funds to the individual rather than just the location, we could define the equal opportunity policy as that which implies equal productivities for all types. In sections 3.3.1-3.3.2 above, we ignored individual heterogeneity, so that definition 1 is consistent with this view. In what follows, the equal opportunity policy is taken as given, and is considered fundamentally different than any policy implied by optimizing a social welfare function.

First-best

With perfect information about individual productivity, the planner again maximizes “total surplus”, which in this case is

$$S = \sum_k (w_k l_k - f(l_k)). \quad (3.33)$$

Differentiating with respect to δ we have

$$\frac{\partial S}{\partial \delta} = \sum_k \frac{\partial w_k}{\partial \delta} l_k(w_k). \quad (3.34)$$

As above, we consider the second-order conditions.

$$\frac{\partial^2 S}{\partial \delta^2} = \sum_k \left(\frac{\partial^2 w_k}{\partial \delta^2} l_k(w_k) + \left(\frac{\partial w_k}{\partial \delta} \right)^2 \frac{\partial l_k(w_k)}{\partial w_k} \right) \quad (3.35)$$

Analogous with the simpler case above, this is not generally a “nice” concave function; and depending on the assumptions made regarding $h(\cdot)$ and $f(\cdot)$, it may have a maximum at the corner where $\delta = 0$. Again, this is true when productivity is linear, so that $\partial^2 w_k / \partial \delta^2 = 0$.

Second-best

In the first-best, redistribution imposes no efficiency costs. Consider now the imperfect information case, where only income, and not wages or labour supply, is observable. The optimal policy in the first-best may involve spending nothing on the education of the poor, and certainly less than on the rich. Will such a policy remain optimal once we introduce information asymmetry? In the general model with asymmetric information regarding types, the optimal tax problem is

$$\max_{c_k, y_k} W = \sum_{k=1}^4 \Psi \left(c_k - f \left(\frac{y_k}{w_k} \right) \right) \quad \text{subject to} \quad (3.36)$$

$$\sum_k y_k \geq \sum_k c_k \quad (3.37)$$

$$c_2 - f \left(\frac{y_2}{w_2} \right) \geq c_1 - f \left(\frac{y_1}{w_2} \right) \quad (3.38)$$

$$c_3 - f \left(\frac{y_3}{w_3} \right) \geq c_2 - f \left(\frac{y_2}{w_3} \right) \quad (3.39)$$

$$c_4 - f\left(\frac{y_4}{w_4}\right) \geq c_3 - f\left(\frac{y_3}{w_4}\right). \quad (3.40)$$

The inequalities (3.37)-(3.40) represent the budget constraint and three incentive constraints which must be satisfied by the tax scheme. Again, the incentive constraints require that at the optimum, each individual attains at least as much utility from their own bundle as from mimicking the type immediately below them.¹⁴ The corresponding lagrangian expression is

$$\begin{aligned} \mathcal{L} = & \sum_k \Psi \left(c_k - f\left(\frac{y_k}{w_k}\right) \right) + \lambda_1 \left(\sum_k y_k - \sum_k c_k \right) \\ & + \lambda_2 \left(c_2 - f\left(\frac{y_2}{w_2}\right) - c_1 + f\left(\frac{y_1}{w_2}\right) \right) + \lambda_3 \left(c_3 - f\left(\frac{y_3}{w_3}\right) - c_2 + f\left(\frac{y_2}{w_3}\right) \right) \\ & + \lambda_4 \left(c_4 - f\left(\frac{y_4}{w_4}\right) - c_3 + f\left(\frac{y_3}{w_4}\right) \right). \end{aligned}$$

Again, we are interested in the welfare impacts of manipulating the distribution of productivities through the education system. As in the simpler model, there will be a “direct effect” on utilities at the optimum which is of ambiguous sign (although again strictly positive for a maxmin planner). We note, however, that the impact on the incentive constraints is no longer unambiguously negative in this case as it was above.

Before analyzing education policy, we must first discuss a complication that arises in the more general case with individual heterogeneity. That education policy will impact the ordering of wages between the types in each location. The following definition is useful.

Definition 2 *Let $\bar{\delta}$ be that policy which satisfies the following*

$$\theta_h h_p(\bar{\delta}) = \theta_l h_r(\bar{\delta}). \quad (3.41)$$

¹⁴Technically, there are many more of these constraints, but these can all be shown to be redundant here.

The policy $\bar{\delta}$ is that which equalizes the productivity of high-types in poor location with that of low-types in the rich location. Refer to values of $\delta < \bar{\delta}$ as case 1, and values of $\delta \geq \bar{\delta}$ as case 2. In case 1, poor students will end up as the bottom two types in the income distribution. As $\delta \in [0, \bar{\delta}]$, the wage ordering $0 \leq w_1 \leq w_2 < w_3 < w_4$ corresponds to $\theta_l h_p(\delta) < \theta_h h_p(\delta) < \theta_l h_r(1 - \delta) < \theta_h h_r(1 - \delta)$. Whereas in case 2 the ordering $0 \leq w_1 \leq w_2 < w_3 < w_4$ corresponds to $\theta_l h_p(\delta) < \theta_l h_r(1 - \delta) < \theta_h h_p(\delta) < \theta_h h_r(1 - \delta)$.

Regardless of the orderings, analysis of the conditions for an optimum yield the following result.

Proposition 6 *Independent of the planner's degree of inequality aversion, the equality opportunity policy spends more than is optimal on the low-productivity region.*

Proof: See appendix.

Proposition 6 tells us that the equal opportunity policy involves spending more on the low-productivity area than is optimal for a social welfare function with *any* non-negative aversion to risk (including infinite). Thus in the general case there are no social preferences that give δ_E as the optimum. Note that we do not require specific restrictions as above with Proposition 2. It is interesting, however, to note that when we relax the assumption that $\partial^2 h / \partial L \partial e > 0$ it is possible that the equal opportunity policy is optimal (at least locally). This is true because for some specifications of the human capital function, redistributive and efficiency goals coincide (at least locally). This is stated in the following corollary.

Corollary 1 *When education funding and local effects are substitutes, such that $h(L, e) = h(L + ke)$, where k is some positive constant, the equal opportunity policy is locally optimal for any social preferences.*

Proof: See appendix.

We see that when neighbourhood quality and education funding are complementary, the equal opportunity policy is too extreme from a social welfare perspective. Given this, a reasonable question is whether education has any value as a redistributive tool when there is a redistributive tax system. Here we have the analogue to proposition 1.

Proposition 7 *Defining the solution to the first-best problem as δ^{fb} , we have the following:*

1. *When $\delta^{fb} = 0$, increasing education spending in the low-productivity region above the first-best is strictly welfare decreasing. This is true regardless of the degree of inequality aversion.*
2. *When $\delta^{fb} > 0$:*
 - (a) *For case 1, increasing education spending in the low-productivity region above the first-best is strictly welfare increasing. This is true regardless of the degree of inequality aversion.*
 - (b) *For case 2, increasing education spending in the low-productivity region above the first-best MAY BE welfare increasing. This is true regardless of the degree of inequality aversion.*

Proof: See appendix.

Increasing education in the low-productivity region from $\delta = 0$ has a negative impact on welfare. Thus proposition (7) states that deviations from the first-best policy in a second-best world are welfare decreasing. So when diminishing returns to education funding are not too large (see equation 3.35), education is not ideal as a redistributive tool in the more general model.

If $\delta^{fb} > 0$, the optimal education policy in the second-best will involve higher spending in the low-productivity region. This is unambiguous in case 1 as defined above, but not certain under case 2 (although certainly plausible, see appendix for details).

We have thus far discussed the impact of education spending on welfare around certain important policies. Deriving the globally optimal δ in the second-best problem is more difficult in the general model than with the previous simplifications. We proceed by describing optimal education spending to the extent possible and then consider a numerical example with specific functional forms and parameter values in section 3.3.3.

Consider the effect of a change in δ on welfare under the optimal tax policy. After some manipulation we can show that welfare is increasing in δ whenever

$$\delta < \frac{(1 - T_2)(\frac{\lambda_3}{\lambda_1} + 1)y_2 + y_1}{\sum_k y_k} \quad (3.42)$$

and decreasing otherwise, where y_k , λ_k and T_2 are determined by the optimal tax problem (denote the marginal tax on type k as T_k). We see that the larger the share of output produced by those from the low-productivity neighbourhood, the more valuable is education spending in that region, as we would expect. Also note that when $\delta = 0$, $y_1 = y_2 = 0$, which is consistent with the case where the second-best problem has an optimum at $\delta = 0$. For the two extremes of social preferences (utilitarian and maxmin), which will be considered below, condition (3.42) is

Utilitarian:

$$\delta < \frac{y_2 + y_1}{\sum_k y_k} < \frac{1}{2} \quad (3.43)$$

Maxmin:

$$\delta < \frac{3(1 - T_2)y_2 + y_1}{\sum_k y_k}. \quad (3.44)$$

The optimum for a utilitarian is equivalent to the first-best as there is no redistribution and thus no concerns regarding information. Since spending on the low-productivity region is inefficient, we see that the impact of increasing δ above equal funding is never optimal. At the other extreme, a maxmin planner places weight only on the welfare of type 1. Even so, it is not clear that education spending is beneficial as this has negative impacts on the incentive constraints and the amount which may be redistributed through the tax system.

Once spending in the low-productivity region exceeds $\bar{\delta}$, this reverses the order of the untalented rich and talented poor in the income distribution, and in this range of expenditures, policy changes will have a different impact because of the effects on incentives in the tax problem. Refer to this as case 2. In case 2, $\delta \in (\bar{\delta}, \delta_E)$, which implies the wage ordering is $\theta_l h_p(\delta) < \theta_l h_r(1 - \delta) < \theta_h h_p(\delta) < \theta_h h_r(1 - \delta)$. For case 2, we have $\partial \mathcal{L} / \partial \delta > 0$ (and negative otherwise) whenever

$$\delta < \frac{\frac{\Psi'(v_1)}{\lambda_1}(1 - T_1)y_1 + (1 - T_3)\left(\frac{\lambda_4}{\lambda_1} + 1\right)y_3 + y_2(1 - (1 - T_2)\left(\frac{\lambda_3}{\lambda_1} + 1\right))}{\sum_k y_k}. \quad (3.45)$$

In case 2, the two extreme cases for social preferences are

Utilitarian:

$$\delta < \frac{y_3 + y_1}{\sum_k y_k} < \frac{1}{2}. \quad (3.46)$$

Maxmin:

$$\delta < \frac{4(1 - T_1)y_1 + 2(1 - T_3)y_3 + (1 - 3(1 - T_2))y_2}{\sum_k y_k}. \quad (3.47)$$

As in case 1, the larger the contribution of output of those from the poor neighbourhood, the more valuable spending is in that neighbourhood. Ideally, we could use these expressions and those above to characterize education policy for different social preferences. However, unless we have analytic solutions for the tax problem that are simple enough to work with, this yields little. To gain further insight, we solve the problem for a specific set of functions and parameters.

Numerical example

Let the functional forms be as in the example of section 3.3.1. Figures 3.3 and 3.4 describe welfare at the optimum in the second-best problem for different values of the education spending parameter δ . The value functions represent the two extremes of social preferences, utilitarian (solid curve) and maxmin (dashed curve). These are generated with the function parameters $\alpha = 0.6$, $\epsilon = 2$. The heterogeneity between types is given by the parameters $L^p = \theta_l = 1$, $L^r = 3$ and $\theta_h = 2$. We see that, for this characterization, welfare in the second-best problem is strictly decreasing in δ , even in the case with infinite aversion to inequality. The value function traced out in the graphs is one of a variety of combinations of parameters all of which have similar results.

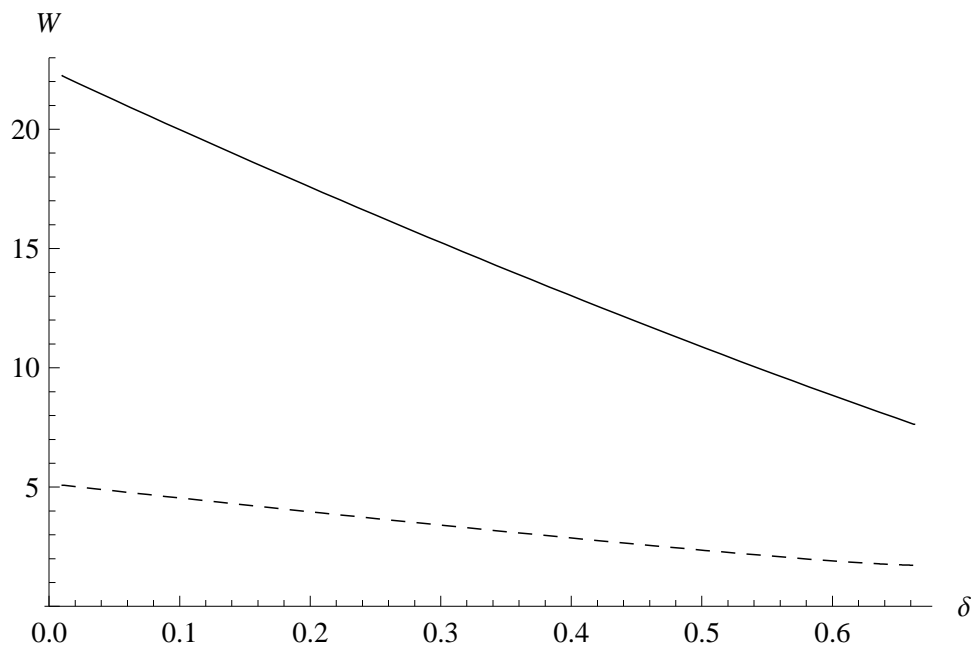


Figure 3.3: Case 1

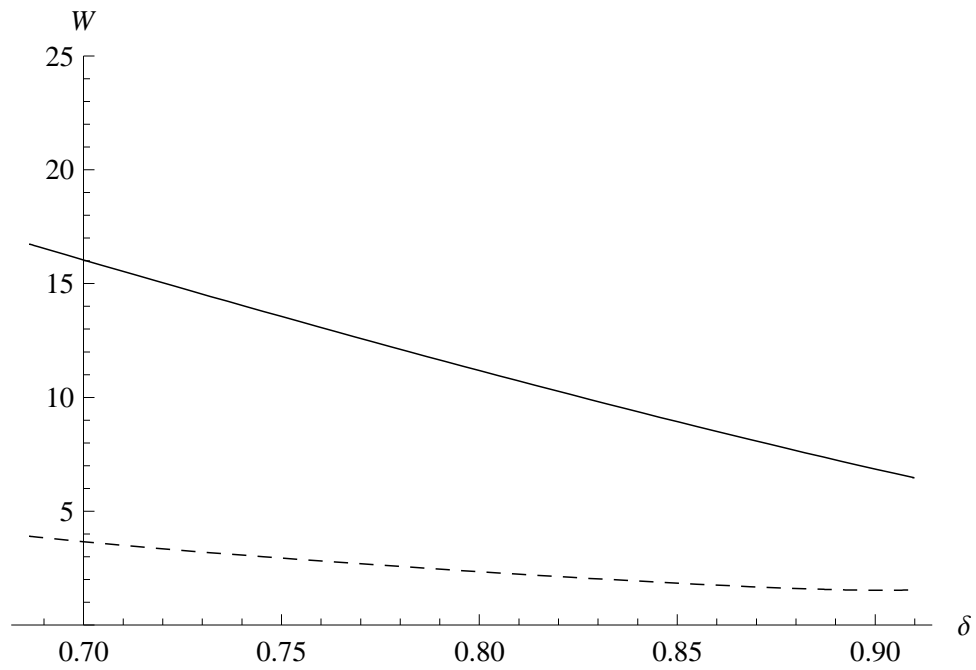


Figure 3.4: Case 2

The results in the more general case are in line with those found above in the simpler models. In this case the first-best is $\delta^{fb} = 0$ and we see that the second-best policy is the same. For these parameters, arguments for the equalization of school funding, or more extremely equalization of opportunities, need to be based on reasons beyond the redistribution of utility in the consumption/leisure sense. Figures 3.3 and 3.4 imply that for this specification, education is inferior to cash transfers even in the second-best problem and with extreme aversion to inequality.

3.4 Social mobility, equal opportunities and dynamics

Our results thus far suggest that education funding should not be used to redistribute utilities between individuals in the standard sense. However, there are many valid arguments supporting such policies that go beyond the analysis above, for example the importance of self-esteem. In this section, we take as given the validity of such a policy and assume that social welfare is concerned solely with the individual's human capital. Then we consider more deeply the importance of education policy in a dynamic setting. First, we modify the above framework slightly and let the population be defined by a continuum rather than four types and, as above, the distribution of θ is the same in the two communities. We continue to assume that the population is equally divided between the two locations. Thus, the human capital of a child i born in neighborhood j is as described in section 3.3.3:

$$w = h_{ij} = \theta_i h(L^j, e_j)$$

with the exception that $\theta \in [0, \bar{\theta}]$ is continuous and distributed according to the measurable function $F(\theta)$, where the total population is normalized to one so that $F(\bar{\theta}) = 1/2$.

3.4.1 Static education spending

We consider first the static spending problem in which the planner divides a fixed budget (again normalized to one) over the two neighborhoods. Define aggregate human capital as the sum of individual human capital in both neighborhoods.¹⁵

$$H = \int_{\theta} [\theta h(L^p, \delta) + \theta h(L^r, 1 - \delta)] dF(\theta) \quad (3.48)$$

Definition 3 Let δ^* denote the spending policy that maximizes aggregate human capital H , implicitly defined by

$$\frac{\partial h(L^p, \delta^*)}{\partial \delta} = -\frac{\partial h(L^r, 1 - \delta^*)}{\partial \delta}. \quad (3.49)$$

We define the planner's aversion to inequality over differences in human capital rather than utility as with a standard social welfare function. In this way the government has preferences solely over human capital and not over some subset of the results of human capital (i.e., consumption and leisure). Implicit in many arguments surrounding equal opportunity policy in education are issues rarely treated in economic analysis. For instance, non-pecuniary effects such as self-esteem, crime, voter savvy, general social cohesion, etc. Consumption is still important in that consumption is presumably rising in human capital, but this is not the only factor of importance. Thus we ignore redistributions over consumption in this section. Further, we put more structure on the planner's preferences and

¹⁵We can think of H as equivalent to aggregate income given the assumptions made regarding wages in section 3.2

assume they are given by a CES function with parameter $\rho = 1 - \sigma$, where $\sigma \in (-\infty, 1]$. Consequently, we define the static problem as

$$\max_{\delta} W = \left(\int_{\theta} ([\theta h(L^p, \delta)]^{\sigma} + [\theta h(L^r, 1 - \delta)]^{\sigma}) dF(\theta) \right)^{\frac{1}{\sigma}}. \quad (3.50)$$

From the planner's perspective, the unborn face a gamble with two outcomes: either born rich or born poor. Education spending influences the expected return of the gamble. Spending affects today's children, who work and live tomorrow. This policy is undertaken ex ante (before adult working life begins), unlike taxes and transfers (after working life begins). In this way, spending on the poor can be taken as insurance against poverty to the not-yet-born. *This insurance is separate from that offered by a tax system which compensates for θ as well as socio-economic status.* In short, larger aversion to differences in human capital across neighbourhoods results in more resources spent on the education of the poor. In particular, as $\rho \rightarrow \infty$ the optimal policy is the equal opportunity policy described in definition (1).

Lemma 1 *Let $\hat{\delta}$ denote the optimal policy in the static problem, as defined by (3.50); then both*

$$\delta_E \geq \hat{\delta} \geq \delta^* \text{ and } \frac{\partial \hat{\delta}}{\partial \rho} > 0. \quad (3.51)$$

This bounds optimal policy choices and describes the “equity/efficiency” tradeoff in this problem. The larger is ρ , the larger is δ and the smaller is aggregate human capital relative to the possible maximum (at δ^*). Furthermore, as H is concave in δ , this “inefficiency” is increasing and convex.

Example

Let human capital production take the following form

$$h(\theta, L, e) = \theta L^\alpha e^{1-\alpha}. \quad (3.52)$$

Further, let differences between neighbourhood spillovers be given by $L^p = 1$ and $L^r = \phi > 1$. The planner's problem is

$$\max_{\delta} W = \max_{\delta} \left(\int ([\theta \delta^{1-\alpha}]^\sigma + [\theta \phi^\alpha (1-\delta)^{1-\alpha}]^\sigma) dF(\theta) \right)^{\frac{1}{\sigma}}. \quad (3.53)$$

The optimal policy is given by

$$\hat{\delta} = \frac{a}{a+1} \quad \text{where } a = \phi^{\frac{\alpha\sigma}{\sigma(1-\alpha)-1}}. \quad (3.54)$$

From Lemma (1), the range of solutions for differing values of $\sigma \in (-\infty, 1]$ is given by $[\delta^*, \delta_E]$ which in this case is

$$\hat{\delta} \in \left[\frac{1}{1+\phi}, \frac{1}{1+\frac{1}{\phi^{\frac{1}{1-\alpha}}}} \right]. \quad (3.55)$$

We see that the larger are both ϕ and α , the larger is the difference between efficiency and redistributive goals. We turn now to the dynamic problem.

3.4.2 Dynamic policy and mobility

The static model described in the previous section lays out a simple framework in which to analyze optimal education spending in this context. This provides a simple equity/efficiency tradeoff we can use to justify policies like those in definition (1). Now we consider the dynamic implications of such policies.

Education spending occurs in two periods: 0 and 1, which are denoted by subscripts. The problem is solved recursively and period 1, which is equivalent to the static problem above, is considered first. Up to this point, we've assumed that people are born into one of the two neighborhoods and this has impacted their human capital outcomes. Government spending on education was only concerned with a single generation of children. However, if local human capital externalities are truly important and education spending can generate "social mobility", then we must consider any impacts on the distribution of skills of a "mobile" population.

Population size remains fixed and each person lives 2 periods, the first of which is childhood where education takes place and the second adulthood where one works and has one child. Once an adult, people choose a neighborhood in which to live. It is assumed that the rich neighborhood is more desirable (for reasons beyond but including better schools). Essentially, if one can afford it, they will live in the rich neighborhood and the poor are kept out. There are numerous papers both empirical and theoretical that describe this phenomenon, and it is simply assumed here.¹⁶

The idea of "social mobility" is commonly referred to as a policy goal which is akin to equal opportunity in that it is designed to sever the link between parents' status and their children's outcomes. However, when you increase "social mobility", people must move in both directions, so it is unclear what the ethical rationale for such a policy would be

¹⁶Theoretical references include Benabou (1996a), Rogerson and Fernandez(1996). Recent empirical papers include Bayer, McMillan and Reuben (2005) and Bayer, Ferreira and McMillan (2007). In particular, it is assumed that moving between neighbourhoods is costless, which is appropriate within a city but perhaps not between countries (as in say the EU) or even states/provinces within a union. Adding moving costs will mitigate the negative impacts described below in Proposition 8 but will not eliminate them.

unless there is some efficiency argument underlying it.¹⁷ In the dynamic case, however, education spending not only influences the outcomes of tomorrow's adults' incomes, but also neighborhood composition, which influences their *children's* outcomes. Before considering optimal policies, we must characterize the period 1 local effects L_1^p, L_1^r which are endogenous from a period 0 perspective.

As income is monotonically increasing in human capital and there are only two neighborhoods, we can divide the population into two groups: those below and those above the median in the human capital distribution. This implies cutoff levels of θ in each neighborhood labelled θ_p^m and θ_r^m . If one's θ is above the cutoff, they will earn enough (exceed the median) to live in the rich neighborhood, and vice-versa.

Definition 4 *The talent cutoff levels θ_p^m and θ_r^m are defined by the following conditions (where L_0^p, L_0^r are exogenous initial differences between the neighborhoods):*

$$\theta_p^m h(L_0^p, \delta_0) = \theta_r^m h(L_0^r, 1 - \delta_0) \quad (3.56)$$

and

$$F(\theta_p^m) + F(\theta_r^m) = \frac{1}{2}. \quad (3.57)$$

Further, we assume that the local externality is given by the average human capital in the neighborhood,¹⁸ so that local effects in period 1 are defined by

¹⁷For instance, models in which communities fund education and the poor community is equally productive but is credit constrained and so invests too little.

¹⁸This is done for simplicity, but a more general specification could be a weighted sum that puts more emphasis on the lower tail (bad apples) or on the higher tail (role models), etc. It would be interesting to consider the impact of different specifications for the local effect.

$$L_1^p = \int_0^{\theta_p^m} [\theta h(L_0^p, \delta_0)] dF(\theta) + \int_0^{\theta_r^m} [\theta h(L_0^r, 1 - \delta_0)] dF(\theta) \quad (3.58)$$

$$L_1^r = \int_{\theta_p^m}^{\bar{\theta}} [\theta h(L_0^p, \delta_0)] dF(\theta) + \int_{\theta_r^m}^{\bar{\theta}} [\theta h(L_0^r, 1 - \delta_0)] dF(\theta). \quad (3.59)$$

Assuming F is uniform over $[0, \bar{\theta}]$, the expressions for L_1^p and L_1 (L_1^r is given by default as $L_1 - L_1^p$) are:

$$L_1 = H_1 = \frac{\bar{\theta}}{2} (h(L_0^r, 1 - \delta_0) + h(L_0^p, \delta_0)) \quad (3.60)$$

$$L_1^p = \frac{\bar{\theta}}{8} \frac{h(L_0^r, 1 - \delta_0)h(L_0^p, \delta_0)}{(h(L_0^r, 1 - \delta_0) + h(L_0^p, \delta_0))}. \quad (3.61)$$

Regardless of what education policy is in period 0, the rich will separate themselves and their children will benefit from being raised in the good neighborhood. This is true even if the planner is able to completely eliminate neighborhood differences between rich and poor in period 0's children. This is a fact of life in a free society and will always be true unless we force people to integrate, in which case it would cease to be a free society. A forward-looking policy will consider the impacts of spending on future neighborhood characteristics, which will impact the extent to which children will have the same chances in the future.

It can be shown that both L_1 and L_1^r are strictly decreasing in δ_0 (for $\delta \geq \delta^*$), which is not particularly surprising. It is generally accepted that equity may come at the price of efficiency, but at some point too much spending on today's poor will hurt the *next generation* of poor.¹⁹

¹⁹Not to mention the negative impact on the tax base which is not modelled here.

Lemma 2 *Reducing the fraction spent on the poor in period 0 from δ_E will increase average human capital in both rich and poor neighborhoods in period 1.*

Proof: See appendix.

A static equal opportunity argument corresponds to extreme aversion to inequality in this model ($\sigma = -\infty$). In fact, the equal opportunity policy δ_E is so redistributive that the effect on tomorrow's poor children is negative (let alone the impact on everyone else). Not only is it negative, but could be substantially so depending on the initial difference in neighborhood composition. Equal opportunity policy prescriptions based on static concepts of justice will be too extreme and inconsistent with optimal dynamic behaviour. The use of education as a redistributive tool should seek to optimize more than just *this generation* of poor. This reasoning extends beyond the extreme case of an equal opportunity policy as is shown below.

2 period problem

The period 0 problem is the choice of δ_0 that provides spending for period 0 children, but also considers the impact of spending on the distribution on human capital. Ignoring discounting, the problem is

$$W = \max_{\delta_0} \left(\underbrace{\int_{\theta} ([\theta h(L_0^p, \delta_0)]^{\sigma} + [\theta h(L_0^r, 1 - \delta_0)]^{\sigma}) dF(\theta)}_{W_1(\delta_0)} \right)^{\frac{1}{\sigma}} \quad (3.62)$$

$$+ \left(\underbrace{\int_{\theta} ([\theta h(L_1^p, \hat{\delta}_1)]^{\sigma} + [\theta h(L_1^r, 1 - \hat{\delta}_1)]^{\sigma}) dF(\theta)}_{W_2^*(\delta_0)} \right)^{\frac{1}{\sigma}}. \quad (3.63)$$

$W_2^*(\delta_0)$ is welfare under the optimal static policy $\hat{\delta}_1$ (described above). Note that $L_1^j = L_1^j(\delta_0)$ and $\hat{\delta}_1 = \hat{\delta}_1(\delta_0)$. While the general optimal policy is difficult to solve explicitly, we can gain some valuable insight from examining the optimal conditions, particularly the way in which our definition of equal opportunity may be inappropriate in the dynamic problem.

Assumption 1 *Let $h(\cdot)$ satisfy the following condition*

$$\frac{\partial h(L_1^p, \hat{\delta}_1)}{\partial L} \geq \frac{\partial h(L_1^r, 1 - \hat{\delta}_1)}{\partial L}. \quad (3.64)$$

Assumption 1 requires that, at the period 1 optimum, an increase in community quality has at least as large an effect on the poor neighbourhood. This seems perfectly reasonable, and is in fact satisfied by a variety of functions $h(\cdot)$, including the Cobb Douglas production function considered above with any value $\alpha \in (0, 1)$.²⁰

²⁰Although not necessary, this simplifies the analysis substantially.

Proposition 8 *A planner maximizing a static objective (as defined by (3.50))*

1. *Will spend too little on the poor (with respect to the dynamic optimum) when $\rho = 0$.*
2. *Always redistributes too much (with respect to the dynamic optimum); for $\rho \geq 2$, the extent to which is increasing with ρ .*

Proof: See appendix.

If there is no aversion to risk or ex-ante inequality, then optimal static policy in the first period (δ_0^*) spends too little on the poor region. For $\rho \in (0, 2)$ the difference between static and dynamic objectives is ambiguous and depends on the function $h(\cdot)$. For large enough aversion to inequality, $\rho \geq 2$, the static policy always spends too much on the poor.

Corollary 2 *If the human capital technology is defined such that $h(L, e) = h(L + ke)$, where k is some positive constant, then the optimal static and dynamic policies are equivalent.*

Proof: See appendix.

We see that, if $h(\cdot)$ takes the form above, then the value of spending on the poor region is higher at the margin regardless of aversion to inequality. In essence, there is no conflict between redistributive and efficiency goals because investment in the poor neighbourhood is more efficient. The optimal policy for any degree of inequality aversion is in fact the equal opportunity policy.²¹ To firm up the intuition, consider the following example.

²¹This case is more in line with models involving credit constraints in the investment of human capital.

Numerical example

We define a specific human capital production function as well as the other relevant parameters described above and solve numerically for the optimal spending mix δ_0 for a variety of risk aversion parameters ρ . Let the $h(\cdot)$ be given by

$$h(L, e) = L^\alpha e^\beta. \quad (3.65)$$

The following table characterizes the solutions to the static problem (described in (3.50)) and dynamic problem (described in (3.62)); these are denoted by δ_0^s and δ_0^d respectively. The orders of magnitude are of no relevance, of course, but results are instructive nonetheless.

Table 3.1: Optimal policy under dynamic and static objectives

ρ	δ_0^s	δ_0^d	$\delta_0^s - \delta_0^d$
0	.355	.364	-.009
0.15	.384	.384	0
0.5	.439	.435	.004
1	.5	.495	.005
1.5	.545	.534	.011
2	.579	.559	.020
3	.627	.592	.035
5	.680	.630	.050
10	.733	.677	.056
100	.793	.735	.058

$$(\alpha = 0.5, \beta = 0.3, \phi = 2, \bar{\theta} = 50)$$

First, we see that Proposition 8 holds in that a planner with no aversion to risk spends too little on the poor and too much when $\rho \geq 2$. Also note that the difference between static and dynamic optima is increasing in ρ . For this set of parameters, it is only for

very low levels of aversion to risk that spending on the poor is too low in the static case. In particular, planner's with risk aversion parameter $\rho = 0.15$ have it "right". Notably, for $\rho = 1$ (the log case which implies equal funding across neighbourhoods), we have overspending in the static problem. This case is often discussed in the literature and may represent a real constraint in this problem in that, for political reasons, a planner could never spend more on the rich.

3.5 Conclusions

The results suggest that redistribution through education may or may not be optimal when policy objectives are defined by a standard consumption/leisure model. In particular, when the optimal first-best (perfect information) policy is non-zero, the second-best policy will be more progressive than the first-best. Implying that education spending on low-productivity types will mitigate distortions generated by the tax system.

For a large class of functions however, the optimal first-best policy will indeed be to spend nothing on the low-productivity types. In this case it is ideal to spend all educational resources on high-productivity students and make up for inequities through money transfers even in the second-best. There are a variety of reasons why this is the case. Spending in the low-productivity location results in a reduction in total output (and consequently transfers), as that of high-types is decreased, while that of the low-types increases relatively less. It is obvious that spending on less productive students can reduce output; less obvious is the role played by the tax system. For a fixed set of skills, low-types are discouraged from working by the tax scheme, so any benefit from increasing their human

capital is not fully realized. Further, increasing the productivity of the low-types tightens the incentive constraint in the tax problem, which creates an even greater distortion in the labour market.

Finally, we note that regardless of functional forms, there are no social preferences under which an equal opportunity policy is called for when there is individual heterogeneity amongst groups.

A further important caveat to educational redistribution, is that the *talented* poor, who will possibly achieve a reasonable outcome regardless, may be the major beneficiaries of these policies. This will be true when the return to education funding is small for those on the low end of the distribution of individual endowments. We have seen that these types are hurt the most by equal opportunity policies, as they will see little but a reduction in money transfers through the tax system. If our redistributive goals are to help the least fortunate (which is implied by the standard social welfare analysis), policies that attempt to equalize opportunities will certainly not be optimal in this case.

The point of this chapter is not to argue for or against policies designed to bring about more equality in productivity, but rather to bring to light the conflict between various objectives. The merits of equal opportunity policies clearly extend beyond the standard limited concepts of utility maximization. The value of education undoubtedly transcends its impact on consumption outcomes both individually and in the aggregate. The impacts of educational spending on self esteem, crime, voter savvy, general social cohesion, etc., are central to the policy debate. Although sometimes difficult to model in an economic framework (at least in an interesting way), these concerns are of great importance nonetheless.

We have shown that, if we accept the validity of such arguments and choose policies

which create greater equality in productivities amongst socio-economic groups, then it is imperative that spending decisions take into account dynamic implications. In particular, when individuals are mobile and there are local externalities in the production of human capital, then a complete equalization of opportunities will be too extreme. If the objective is to sever the link between socio-economic status and human capital outcomes, simply compensating for local effects in each period is not reasonable. An alternative policy objective is to maximize human capital in the poor region over the long run. This would require a deeper understanding of both the importance of local externalities and of the human capital function in general.

An ideal education policy will depend on the relative importance of various objectives, which as we have seen, may be in conflict. This chapter highlights the need to be aware of both why we wish to equalize opportunities, and, if so, to be cognizant of the possible dynamic inconsistency inherent in such policies.

Chapter 4

Education Systems

4.1 Introduction

Economists commonly ascribe “talents” or productivities to the various actors in their models, but rarely consider the fact that nobody begins (or likely ends) life with perfect knowledge of their own abilities. It is therefore prudent to consider the role our education system plays, not only in the training of our youth, but in the discovery of their own unique abilities. This has largely been ignored in the literature on human capital formation thus far.

We consider the relative merits of two institutions we call specialized and general school systems. In the model, individuals are heterogenous with respect to their talents and are also initially uninformed about their particular draw from the talent distribution. We characterize a fundamental trade-off between various educational system designs that weighs the returns to enhancing specific skills with the benefits of a broader education

that reduces risk in skill attainment. In what we refer to as a specialized system, students choose a particular field which is the focus of all their studies. If there were no uncertainty over one's type, this would provide the highest level of human capital possible. When students are forced to choose without knowledge of their type (one only learns their relative talents by trying their hand), then there is the possibility that a poor choice of field brings a low-productivity outcome. Thus a general system, in which students are educated in all fields, albeit to a lesser degree, avoids these "mistakes" and may provide a better outcome on average. This is the basic trade-off considered, but we also note that there are distributional implications that go beyond average returns and may be of interest. A specialized system necessarily creates a division between those who make "mistakes" and those who don't, even though they are ex-ante identical (and remain identical in a general system).

The literature on human capital sees education as a means to enhance productivity. Coupled with efficient labour markets, increased productivity translates directly into higher wages and welfare. The efficiency of labour markets precludes education from affecting any distortions therein.¹ The contribution of this chapter is to analyze the impact of differing educational institutions as mentioned. We do this, however, both when labour markets are efficient (as is common in the literature), as well as when they are characterized by various frictions which exist in more sophisticated views of labour markets.

Specifically, we allow for informational asymmetry in the sense that firms cannot perfectly observe individual productivity. Spence (1973) first posited that labour markets may be inefficient if productivities are not directly observable by firms. This created an additional role for education: as a costly signal for higher ability types to use to distinguish

¹The wage distribution with education first-order stochastically dominates the wage distribution without education. Education is just a technology which increases productivity.

themselves from their lower ability counterparts. In other words, education was seen as resolving inefficiencies in the labour market, the cost of which were borne by higher ability types through excessive and socially wasteful educational investments. While there is seemingly little evidence for this type of signaling in its purest form (where education has no impact on productivity), it is notoriously difficult to estimate these effects. Throughout this chapter we shall assume that education is productive and we allow for the possibility that adverse selection creates real costs to firms and workers alike.² Our model predicts that a specialized system imposes a larger adverse selection cost than a general one.

Beyond the adverse selection problem, we also consider another labour market friction which is especially relevant to the discussion of education systems. This is the existence of coordination failures that can arise when individuals seek employment. Using a simple framework outlined by Montgomery (1991), we consider the impact that different educational institutions have on this labour market friction, and find that a general system leads to a larger coordination failure than a specialized one.

The rest of the chapter is organized as follows. Section 4.2 outlines the basic analytical environment. Sections 4.3 and 4.4 describe the effects of different educational institutions when there are frictionless labour markets. We relax this in sections 4.5 and 4.6 which allow for adverse selection and coordination failure. Finally, section 4.7 concludes.

²For a recent discussion on the empirical evidence supporting job market signaling, see the Handbook of the Economics of Education (2006) Chpt. 8.

4.2 Model

The model consists of three key parts: individuals, education systems and labour markets which are discussed in turn.

4.2.1 Individuals

We consider a two-period model ($t = 1, 2$) in which agents are unaware of their talents at the outset, but learn about them via education. Individuals potentially have talents in one of two fields, labelled A and B. Individual endowments are represented by a vector $\theta = (\theta_A, \theta_B)$ where θ_A is talent in A and θ_B is talent in B.

There are three possible talent pairs: (θ_H, θ_L) , (θ_L, θ_H) , (θ_H, θ_H) where $\theta_H > \theta_L$. Agents with the first pair of talents are labelled A's as they are suited to A. Similarly, the second type we refer to as type B's. Agents with the last pair of talents are equally talented in both fields.³ The proportion of HLs or A types is p_{HL} , LHs or B types in p_{LH} and HH types is p_{HH} .

Agents can earn income in each period by working in the labour market, and in any period in which they choose to acquire education, they earn nothing. Utility in each period is equal to income which is derived from wages. In the following subsection, the effect of education on productivity is described after which the relationship between productivity and wage is discussed.

³We ignore the possibility of types that are low talent in both fields. This allows for cleaner results, but is otherwise of little consequence.

4.2.2 Education systems

We consider two education systems which can be viewed as polar cases: specialized and general. Each offers education in the two fields A and B. Agents are initially unaware of their talent or returns to schooling in each field. Undertaking education in a particular field has two effects. First, it informs agents about their talent in that field (either θ_H or θ_L). Second, it enhances that talent, yielding $r(\theta, e)$ units of productivity for an agent with talent θ undertaking e units of education where $r(\cdot)$ is increasing in both arguments.

The systems differ with respect to when and what education can be acquired. In the specialized system, agents enter a single field and receive e units of education. They learn only about their talent in that field. Therefore, HL types that enter field B will learn that they are low types in field B and have low returns to education. The same will be true of LH types that enter field A. Education nevertheless boosts productivity so while they are learning about their returns to education they are also acquiring skills. For HL types that enter field A (LH types that enter field B) the returns are $r(\theta_H, e)$. If HL types enter field B (LH types enter field A), the returns are $r(\theta_L, e) < r(\theta_H, e)$.

On the other hand, in the general system, agents receive education in both fields simultaneously. Here, agents learn about their talents across both fields. The cost is that they will spend less time in the each field as time has to be split across the two. Hence, relative to the specialized system, they will have smaller investments in human capital spread across the two fields rather than one larger investment in a single field. The exact returns are as follows:

- $(r(\theta_H, e_g), r(\theta_L, e_g))$ for HLs
- $(r(\theta_L, e_g), r(\theta_H, e_g))$ for LHs
- $(r(\theta_H, e_g), r(\theta_H, e_g))$ for HHs

where $e_g < e$.

4.2.3 Labour market

There are two sectors A and B. Labour markets are competitive on the demand side: firms compete for workers in each sector. The wage in each sector equals the marginal product of a worker: $w = r(\theta, e)$ under specialized or $r(\theta, e_g)$ under general. However, two possible frictions may arise in the labour market. First, firms may not be able to observe productivities directly if multiple types are present. Since education can not be used as a signal, firms also cannot infer types from observing education amounts. Second, there may be coordination problems between agents and firms when agents search for jobs.

When firms can not observe types directly, they nevertheless have access to a costly screening technology that they can use to screen types. The cost to a firm of screening a potential employee is κ . Firms announce that they will screen any applicant that applies for a high paying (high productivity) job, and since firms compete, in equilibrium this cost is fully reflected in the wage. The screening cost is high enough so that low types are unwilling to apply to the higher paying jobs (as they will be found out), while low enough that the high types are willing to apply to those jobs rather than stay out of the labour market. Formally,

$$r(\theta_L, e) - \kappa < r(\theta_L, e) < r(\theta_H, e) - \kappa.$$

This corresponds to a reduced form version of a *competitive* screening model⁴ in which agents are free to choose education amounts in the second period and firms offer (wage, education) contract pairs to separate types as in Rothschild-Stiglitz (1976).

Utilities of agents equal their wages which are a function of talent and the education an agent acquires. The timing is described in the next subsection.

4.2.4 Timing

Specialized:

- Period 0: individuals can enter the education system or earn an amount $w_0 = 1$ forever, which implies a lifetime utility of $\frac{1}{1-\beta}$.
- Period 1: if individuals enter the education system, they must choose education in field A (e_A) or B (e_B).
- Period 1 end: nature moves and selects types (returns to education in the field taken) with given probabilities.
- Period 2: individuals enter the labour market, choose which sector to work in, A or B, and work in the chosen field.

General:

- Period 0: individuals can enter the education system or earn an amount $w_0 = 1$ forever, which implies a lifetime utility of $\frac{1}{1-\beta}$.

⁴This relationship is discussed in the appendix to this chapter.

- Period 1: all individuals who enter the education system receive education (e_g) in both A and B.
- Period 1 end: nature moves and selects types (returns to education in *both* fields) with given probabilities.
- Period 2: individuals enter the labour market, choose which sector to work in, A or B, and work in the chosen field.

4.3 Benchmark

It is useful to begin by looking at a welfare benchmark against which we may judge our subsequent results. This benchmark compares the two systems when agents are perfectly informed about their talents at the beginning of period 1 (or before entering the education system). This ensures that agents make efficient human capital investments.

4.3.1 Specialized

We represent the actions, timing and payoffs for the specialized case in the Figure 4.1. Analyzing the subgame for HLs, we see that there is a unique equilibrium: (e_A, A) yielding payoffs of $r(\theta_H, e)$. To see this, first note that $r(\theta_H, e) > r(\theta_L, 0)$, so that, conditional on choosing education in field A, HLs prefer to work in A. Then, conditional on choosing education in field B, HLs prefer to work in B as $r(\theta_L, e) > r(\theta_H, 0)$. Lastly, as $r(\theta_H, e) > r(\theta_L, e)$, HLs prefer to be educated in field A because they have higher returns to education in that field. Similarly, looking at the subgame for LHs, (e_B, B) yielding

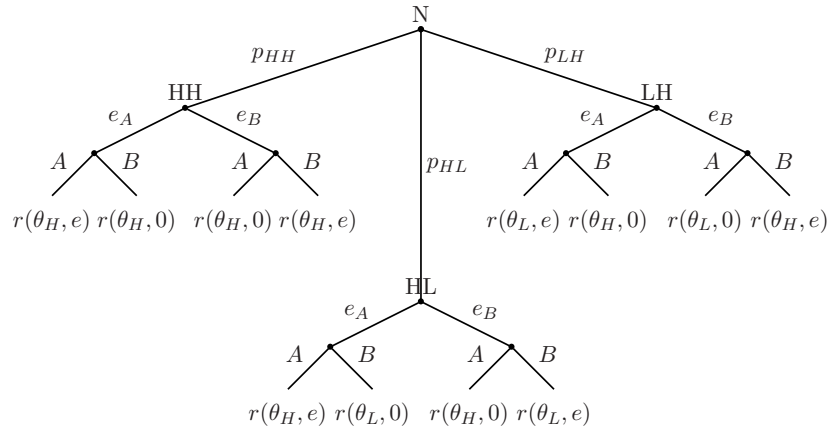


Figure 4.1: Specialized system

payoffs of $r(\theta_H, e)$ is the unique equilibrium. Looking at the subgame for HHs, we see that they are indifferent between choosing education in A or B since each yields a payoff of $r(\theta_H, e)$. We assume that they will randomize so that half will enter A and the other half will enter B.

We can compute the ex-ante welfare associated with a specialized system, U_S^{FB} , as follows:

$$U_S^{FB} = p_{HH}r(\theta_H, e) + p_{HL}r(\theta_H, e) + p_{LH}r(\theta_H, e).$$

4.3.2 General

As with the specialized case, we can represent the actions, timing and payoffs of a general system in the game tree shown in Figure 4.2. Again, working backwards and analyzing the subgame for HLs, we see that working in A is an equilibrium as $r(\theta_H, e_g) > r(\theta_L, e_g)$. Similarly, working in B is an equilibrium for LHs. Moreover, the HHs are indifferent between working in A or B so we assume they randomize with half choosing

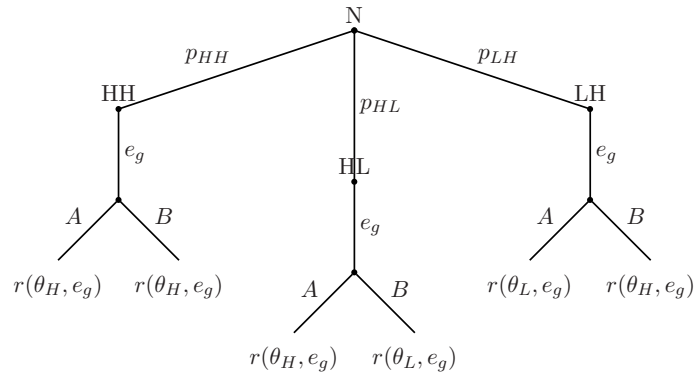


Figure 4.2: General system

A and the other half choosing B.

The payoffs for each type are simply $r(\theta_H, e_g)$ and so the ex-ante welfare in a general system is simply U_G^{FB} :

$$U_G^{FB} = p_{HH}r(\theta_H, e_g) + p_{HL}r(\theta_H, e_g) + p_{LH}r(\theta_H, e_g).$$

Thus we have the following basic result:

Lemma 3 *When agents are perfectly informed about their types, a specialized system dominates a general system, $U_S^{FB} > U_G^{FB}$, as long as $e > e_g$.*

Essentially, as long as agents can make efficient human capital investments, there is no benefit to a general system. This is because in the general system some agents acquire education in the field to which they are not suited. This education is socially wasteful as it does not ultimately translate into higher wages.

4.4 Uncertain returns

When returns to education are uncertain, a specialized system has disadvantages that the general system can overcome. Uncertainty about returns to education in any field can only be overcome by actually taking education in that field. A specialized system restricts agents to learning about only one of their talents leading to potentially inefficient human capital investments. On the other hand, the general system allows agents to learn about both talents simultaneously but, at the cost of reducing their ex-post productivity in any given field.

4.4.1 Specialized

We represent the actions, timing and payoffs for the specialized case with uncertain returns in Figure 4.3. Proceeding backwards, starting with the subgame for the HHs, we can see that conditional on having education in field A, HHs will choose to work in A when $r(\theta_H, e) > r(\theta_H, 0)$. Similarly, HLs will work in field A. Moreover, LHs will also work in field A as $r(\theta_L, e) > r(\theta_H, 0)$ by assumption.⁵

HHs and HLs will receive the benchmark wage, $r(\theta_H, e)$, while LHs will receive the low wage, $r(\theta_L, e)$. The expected payoffs from entering field A are then simply:

$$(p_{HH} + p_{HL})r(\theta_H, e) + p_{LH}r(\theta_L, e). \quad (4.1)$$

The situation is analogous conditional on having acquired education in field B; everyone prefers to work in field B, yet only HHs and LHs get the high wage, $r(\theta_H, e)$, while the

⁵Note that this implies that more than one type will persist in the labour market. This is not a problem here because we assume that firms are able to observe productivities. We relax this assumption in the next section.

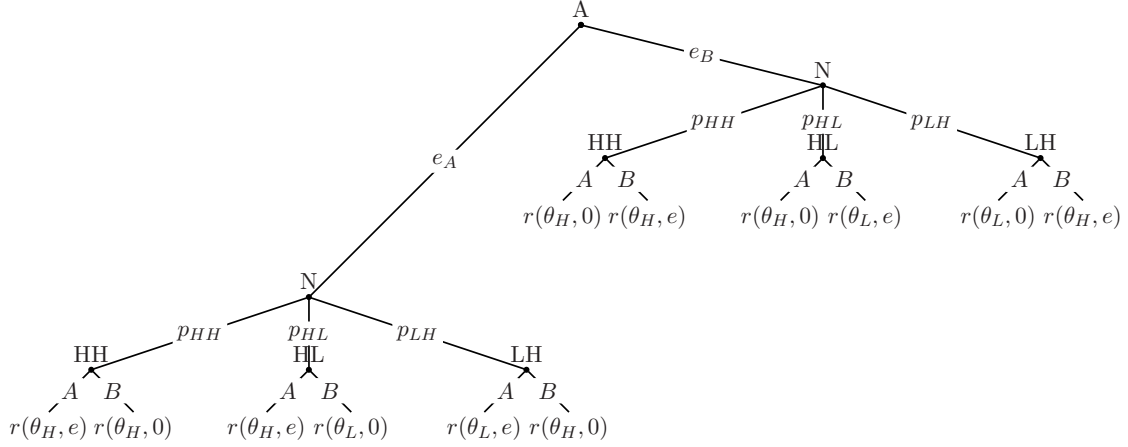


Figure 4.3: Specialized system with uncertainty

HLs get the low wage $r(\theta_L, e)$. The expected payoffs from entering field B are identical to those in field A as long as $p_{HL} = p_{LH} = p_M$.

Hence, agents are ex-ante indifferent between entering field A or B. We assume that they randomize with half entering A and the other half entering B. Here, we see that uncertainty over returns combined with a specialized system leads some to make inefficient investments in education. This is of course different from the situation in the first-best, where everyone made efficient investments.

4.4.2 General

We represent the actions, timing and payoffs for the general case with uncertain returns in Figure 4.4. Solving via backward induction, we see that HL types will choose to work in A and LH types will choose to work in B as $r(\theta_H, e_g) > r(\theta_L, e_g)$. Moreover, HH types are indifferent between working in A or B, so we continue to assume they randomize with half working in A and half working in B. Importantly, the payoffs to each type are exactly as in

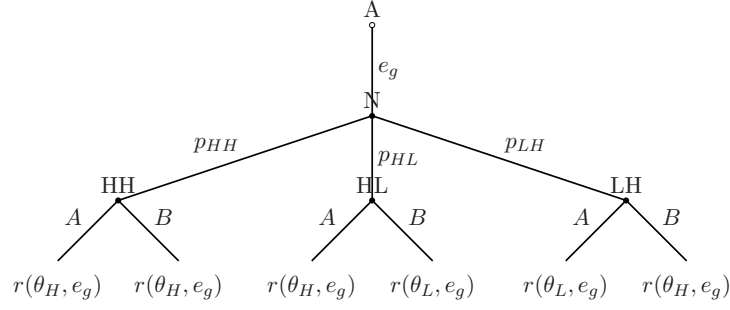


Figure 4.4: General system with uncertainty

the first-best case, namely, $r(\theta_H, e_g)$. As these payoffs are the same as in the case without uncertain returns, we see that the general system provides insurance against investing in the wrong field.

4.4.3 Welfare

The ex-ante welfare under a specialized system is simply:

$$(p_{HH} + p_M)[r(\theta_H, e)] + p_M[r(\theta_L, e)]. \quad (4.2)$$

Under a general system, the ex-ante welfare is $r(\theta_H, e_g)$, and as noted earlier is unchanged from the benchmark case as there are no inefficient education investments. Therefore,

Lemma 4 *A specialized system dominates a general one whenever*

$$(p_{HH} + p_M)[r(\theta_H, e)] + p_M[r(\theta_L, e)] > r(\theta_H, e_g). \quad (4.3)$$

We can rewrite the above expression as follows:

$$[r(\theta_H, e) - r(\theta_H, e_g)] > p_M[r(\theta_H, e) - r(\theta_L, e)]. \quad (4.4)$$

The LHS is the welfare loss resulting from a general system; agents invest in less education in their chosen field as they must acquire some education in both fields. The RHS is the welfare loss from a specialized system: a fraction p_M of agents in aggregate invest incorrectly and are paid $r(\theta_L, e)$ instead of $r(\theta_H, e)$. In other words, a specialized system is better if the expected returns to education, despite the inefficient investments, exceed the corresponding returns under the general.⁶ We have the following result, which will be useful in the next sections.

Lemma 5 *For any fixed level of education e in the specialized system, there exists some $e_g \equiv e_g^* < e$ such that welfare under the two systems is equal.*

Proof. See appendix. ■

The pair of education levels (e, e_g^*) result in equivalent welfare under the two systems when there are no frictions in the labour market. In subsequent sections, we will introduce various frictions to understand the implications for the use of one system versus another.

4.5 Adverse selection

When firms cannot directly observe productivities, a specialized system due to inefficient investments by some HLs and some LHs, generates adverse selection in the labour market. HL types that have invested incorrectly (correspondingly LH types) may mimic the LHs (HLs) that have invested correctly in sector B (sector A). To remedy this situation, firms announce that they will screen any individual who applies for the higher paying

⁶Throughout we ignore the distribution of skills, but obviously this could be of importance as a specialized system results in different outcomes for individuals who are ex-ante identical.

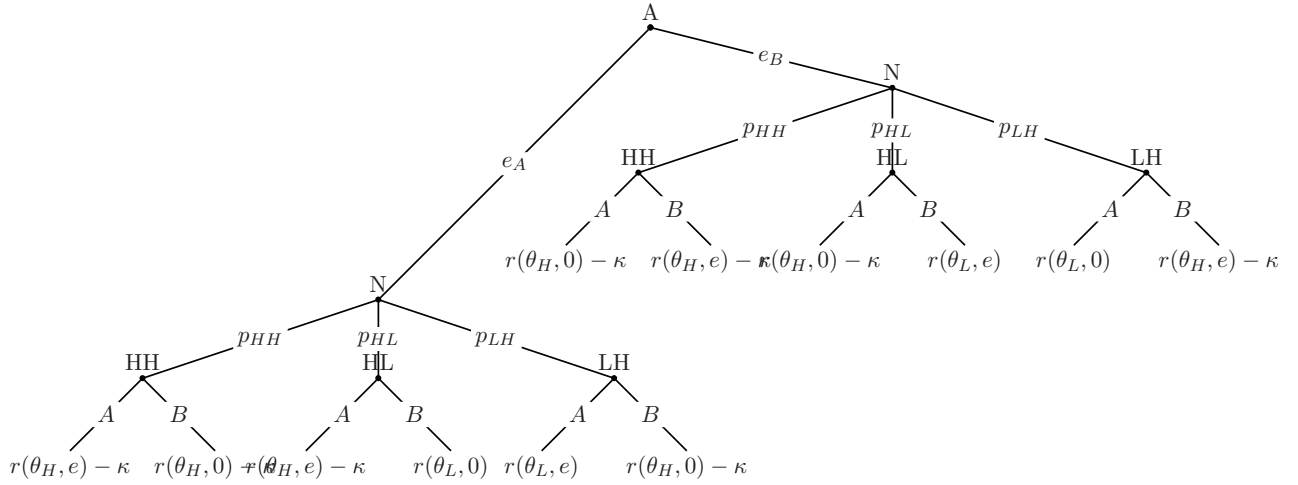


Figure 4.5: Specialized system with uncertainty and screening costs

jobs. Screening a potential employee costs κ and due to competition among firms will be reflected in the wage for higher productivity types. Moreover, given the behaviour of firms, agents that are low types will choose not to mimic high types as they will be found out (get the low wage) and still have to pay the screening cost.

We represent the actions, timing and payoffs in Figure 4.5. Once again, we can solve the game via backward induction. Starting with the subgame for the HHs, conditional on having education in field A, they will choose to work in A as $r(\theta_H, e) - \kappa > r(\theta_H, 0) - \kappa$. Similarly, HLs will work in field A. Moreover, LHs will also choose to work in field A as $r(\theta_L, e) > r(\theta_H, 0) - \kappa$ by assumption. Then, more than one type will persist in the labour market and firms will need to screen workers (i.e., $\kappa > 0$). HHs and HLs will receive the high wage, $r(\theta_H, e) - \kappa$, while LHs will receive the low wage, $r(\theta_L, e)$.

The expected payoffs from entering field A are then simply:

$$(p_{HH} + p_{PHL})[r(\theta_H, e) - \kappa] + p_{PLH}r(\theta_L, e).$$

The situation is analogous conditional on having acquired education in field B; everyone prefers to work in field B, yet only HHs and LHs get the high wage, $r(\theta_H, e) - \kappa$ while the HLs get the low wage $r(\theta_L, e)$. Continuing with the assumption that $p_{HL} = p_{LH} = p_M$, the expected payoffs from entering field B are identical to those in field A. Hence, agents are ex-ante indifferent between entering field A or B. We assume that they randomize with half entering A and the other half entering B.

In the general system, only a single productivity type persists in both sectors so that even if firms cannot observe productivity directly, they may infer it. Hence, no adverse selection arises and the situation is exactly akin to the one described in the previous section.

4.5.1 Welfare

The ex-ante welfare under a specialized system is simply half the value of entering field A and half the value of entering field B:

$$(p_{HH} + p_M)[r(\theta_H, e) - \kappa] + p_M[r(\theta_L, e)]. \quad (4.5)$$

The welfare loss due to adverse selection, is given by (4.2) - (4.5) and is of course just the aggregate screening cost: $(p_{HH} + p_M)\kappa$. Under a general system, the ex-ante welfare is unchanged from the first-best case, as there are no inefficient education investments and therefore only high types in each labour market. The welfare distortion due to adverse selection in the general system is thus zero. To see the aggregate distortions, we can write the analogue of (4.4):

$$[r(\theta_H, e) - r(\theta_H, e_g)] > p_M[r(\theta_H, e) - r(\theta_L, e)] + (p_{HH} + p_M)\kappa. \quad (4.6)$$

The LHS of (4.6) is identical to (4.4), as the distortion due to the general system is unchanged by adverse selection, but RHS of (4.6) now incorporates the adverse selection cost. In other words, a specialized system is better if the expected returns to education in it exceed the corresponding returns under the general plus the screening costs. In summary,

Proposition 9 *The presence of adverse selection in the labour market favours a general system over a specialized one:*

$$(p_{HH} + p_M)r(\theta_H, e) + p_M r(\theta_L, e) < r(\theta_H, e_g^*) + (p_{HH} + p_M)\kappa \quad (4.7)$$

.

Proof. See appendix. ■

4.6 Coordination and matching

This section investigates the importance of coordination in the labour market and subsequently the relative merits of different educational institutions. Firms will post wage offers, and individuals respond to those offers and apply. We assume that the number of firms equals the number of job seekers and that each firm posts exactly one vacancy. Further, individuals may apply to one job only.⁷

Consider a simple case with two job postings and two seekers. If both apply to different jobs, then they are both hired and produce. If, however, both apply to the same job, then one is out of luck and the other job is left unfilled (we assume the lucky one is chosen randomly). One may wonder why seekers don't agree to apply to different firms

⁷The coordination problem is described in Montgomery (1991), in which he shows that these assumptions can be relaxed without changing the results in any substantial way.

and this is the essence of the coordination problem, as large labour markets do not allow seekers to interact. As we have an equal number of jobs and seekers, in the absence of any coordination failure, there will be no unemployment and any analysis of education systems will be as in the first-best.

We assume in this section that individual productivity is observable, and that if an applicant shows up for the position they are hired (if more than one shows up they're randomly selected with equal probability). As individual type is observable, different productivity types are in different labour markets and respond to the postings directed towards them. The value of an employee to a firm in either field is given by individual productivity less the wage paid

$$\pi = r(\theta, e) - w \quad (4.8)$$

where $r(\theta, 0) = 0$.

Posting higher wages attracts more workers but reduces the surplus to the firm if a worker is hired. A firm i with an opening solves

$$\max_{w_i} (r_i - w_i) \cdot \text{prob}(\text{attracting at least one worker}).$$

The objective function can be written

$$(r_i - w_i) \cdot [1 - (1 - p_i(w))^n] \quad (4.9)$$

where n is the number of qualified applicants in the labour market (and postings here as well as there is no unemployment without a friction). The vector of wage postings is given by $w = (w_1, \dots, w_n)$, and $p_i(w)$ is the probability that a worker will apply to position i . The mixed-strategy application probabilities (the p_i 's) are determined by the following two

conditions:⁸

$$w_i \cdot \text{prob}(\text{getting a job at firm } i.) = w_j \cdot \text{prob}(\text{getting a job at firm } j.) \quad \forall i, j \quad (4.10)$$

and

$$\sum_i p_i = 1. \quad (4.11)$$

As shown in Montgomery (1991), $\text{prob}(\text{getting a job at firm } i.) = [1 - (1 - p_i(w))^n]/p_i n$, so that we can rewrite (4.10) as

$$w_i \cdot [1 - (1 - p_i(w))^n]/p_i n = w_j \cdot [1 - (1 - p_j(w))^n]/p_j n. \quad (4.12)$$

This system of $n - 1$ equations along with (4.11) could be used to solve for the application probabilities as a function of w which could then be used to derive reaction functions for each firm. As the market gets large, however, firms become price takers in the labour market. Thus firms maximize profits subject to

$$w_i \cdot [1 - (1 - p_i(w))^n]/p_i n = K \quad (4.13)$$

where K represents market tightness. Substituting (4.13) into the firm's objective function we have the unconstrained problem

$$\max_{p_i} r[1 - (1 - p_i(w))^n] - p_i(w)nK.$$

The first-order condition is

$$rn(1 - p_i(w))^{n-1} = nK.$$

⁸We focus on mixed strategies because pure strategies require coordination. See Montgomery (1991) for details.

Rearranging gives the equilibrium application probability for firm i

$$p_i = 1 - (K/r)^{1/(n-1)}.$$

As firms have the same valuation of a worker $r(\cdot)$, $p_i = p_j \forall i, j$. Further, as $\sum p = 1$, we have $p_i = p_j = p = 1/n$.

Upon finishing schooling, individuals enter the labour market and apply for a job in one of the fields A or B . The education system enhances productivity as described earlier. Under a specialized education system, as only education in a single field is undertaken, the choice of field is trivial, as $r(\theta_L, e) > r(\theta_H, 0)$. Under a general system, however, those not particularly suited to either field (the (θ_h, θ_h) types) are indifferent over the two (which will exacerbate the coordination problem as we shall see).

4.6.1 Welfare

We ignore the division of output between firms and workers (all surplus should go to workers with large markets). Welfare or surplus is simply the value of production weighted by the probability that a match occurs. For each productivity type i (group size n), the contribution to total surplus is

$$\sum_{i=1}^n r_i \Delta(n) = n \cdot r_i \Delta(n),$$

where $\Delta(n) = (1 - (1 - p_i)^n) = (1 - (1 - \frac{1}{n})^n) < 1$ represents the distortion caused by coordination failure ($\Delta = 1$ with perfect coordination as there are an equal number of jobs and job seekers). The larger the market, the worse the coordination failure as $\partial \Delta / \partial n < 0$. Our interest is in the comparison of educational systems which influence both the productivity and size of each group of employees. We consider each in turn.

Specialized

Under a specialized system, individuals are only productive in one industry and thus apply for work only in that industry. As individuals have no information on type, they enter each field with equal probability. Thus there will be two types of job seeker, high and low-productivity workers. The high types are those who received education in the field to which they were suited or were high productivity types in either field. The low-productivity types are those who received training in the field in which they were not inclined (i.e., a specialized system forced them to choose a field and they choose incorrectly). Formally, the number of high types is $n_h = p_{HH} + \frac{1}{2}p_{HL} + \frac{1}{2}p_{LH}$ and low-types $n_l = \frac{1}{2}p_{HL} + \frac{1}{2}p_{LH}$. The productivity of the former is $r_h = r(\theta_H, e)$ and the latter $r_l = r(\theta_l, e^s)$. Thus total surplus is given by

$$S_s = \sum_{i=1}^{n_h} r_h \Delta(n_h) + \sum_{i=1}^{n_l} r_l \Delta(n_l) = n_h \cdot r_h \Delta(n_h) + n_l \cdot r_l \Delta(n_l).$$

General

Under a general system there are no “mistakes” and all types are high-types. The cost of course is that $e^g < e$. Thus surplus in the general case is simply

$$S_g = \sum_{i=1}^{n_g} r_g \Delta(n_g) = n_g \cdot r_g \Delta(n_g),$$

where $n_g = p_{HH} + p_{HL} + p_{LH}$ and $r_g = r(\theta_H, e_g)$.

With these expressions, we can compare the second-best surplus under the two systems. It is most interesting to assume that both systems provide equal surplus in the first-best (this is for e_g^* as defined in lemma 5). Thus we can analyze the relative impact of each education system on the coordination problem. Again, if the two provide equal surplus in

the first-best, we have

$$(p_{HH} + p_{LH} + p_{HL})r_g = \left[p_{HH} + \frac{1}{2}(p_{LH} + p_{HL}) \right] r_h + \frac{1}{2}(p_{LH} + p_{HL})r_l. \quad (4.14)$$

Further, we shall make use of the following Lemma which is easily shown to be true.

Lemma 6 *The distortion caused by the coordination failure is increasing in the size of the labour market ($\Delta'(\cdot) < 0$).*

Lemma 6 implies that $\Delta_g < \Delta_h < \Delta_l$. Given this and (4.14), we have the following result.

Proposition 10 *A general education system provides a lower surplus than a specialized one as it leads to a larger coordination problem.*

Proof. See appendix. ■

4.7 Conclusions

We have analyzed the impacts of two different educational institutions and discussed the relative merits of each. A specialized system offers the opportunity to gain a larger amount of human capital in a given field, but comes with the risk of choosing a field to which one is not suited, which leads to a low outcome. Moreover, we have seen that under such a system, an adverse selection problem will arise that further reduces the returns to education. This is because there will be different productivity types in each labour market that are costly to screen.

A general system on the other hand provides insurance against a low-productivity outcome, but comes at the cost of reduced productivity in any given field. We have also seen

that, if there are coordination problems in the labour market, then this type of educational institution will exacerbate this friction relative to a specialized system. This is because there will be individuals who are indifferent over which field they wish to pursue a career.

Chapter 5

Leviathan Governments and Public Debt

5.1 Introduction

The sizeable debts accumulated by most countries have likely not come about solely as a result of society's needs.¹ Are governments as willing to save as they are to spend? Is it not strange to think of a government as a creditor?

If incumbents are indeed inclined to spend too much, a natural question to ask is whether we should constrain them through some sort of constitutional restriction? This chapter outlines a simple framework in which this question is addressed. The model characterizes the trade off between the benefits of better politician control and the cost that such restrictions will have in times of scarcity. The value of a limit depends both on the

¹The U.S national "debt clock" in New York recently topped \$10,000,000,000,000 [*The Economist*, Oct.18th, 2008].

extent to which politicians' goals deviate from their constituencies' and how effectively the electoral process disciplines them for misbehaving. Further, the results suggest that such a constraint will have a positive/negative expected benefit depending on the fiscal position at the time in which it is imposed. In particular, the value of constraining politicians depends positively on the size of *current* public asset holdings.

The basic intuition of the model is as follows. Suppose there is a random process which determines the government budget. Further suppose that politicians are free to borrow and save, but are self-interested and have the option of running up a large debt (up to some fixed limit). Perhaps the public enjoys better public services in the first period, but are then faced with a debt that must be repaid tomorrow. Not only that, if the debt is up to the limit there is no longer the option of running debt if there is a negative shock to the economy tomorrow. Imagine now that a rule is imposed which limits borrowing. If the government is somewhat benevolent and values the returns to good behaviour (e.g., re-election), then a debt control reduces his/her outside option of running up the debt, and the voters may be better off. It is not necessarily beneficial, however, as politicians may be reasonably well controlled through the electoral process. Thus it may not be worth the cost of debt controls, which reduce the ability to smooth consumption over time as well as increase any distortions to the economy caused by tax financing.

The remainder of the chapter is organized as follows. Section 5.2 outlines two versions of the basic finite period model, one with taxes and one without. Section 5.3 considers optimal debt limits. Section 5.4 extends the analysis to include an infinite horizon and section 5.5 concludes. All proofs can be found in the appendix.

5.2 The model

Section 5.2 characterizes both the voters' and politicians' problems, the environment in which they interact, and finally it describes tax and spending behaviour.

5.2.1 Voters

As is the norm when using this analytical framework, we assume society is represented by a single voter. Each period, the voter receives an endowment $y_t \in [\underline{y}, \bar{y}]$, which follows an i.i.d process characterized by distribution function $F(y)$.² Utility in period t is derived from both private consumption c_t and government services g_t and defined as

$$U(c_t, g_t) = c_t + h(g_t). \quad (5.1)$$

Where $h()$ is concave and monotonically increasing.³ Private consumption and public services are defined by

$$c_t = (1 - \tau_t)y_t$$

$$g_t = \delta(\tau_t)y_t - (1 + r)b_{t-1} + b_t.$$

Government borrowing at time t is b_t , the tax rate is τ_t and the interest rate is fixed at r .⁴

The function $\delta()$ represents the distortionary costs associated with taxation (as in Barro, 1979) that are not made explicit in the model and satisfies $1 > \delta'() > 0$ and $\delta''() < 0$.

Debt taken in any period is to be repaid in full the following period.

²A persistent process is more realistic but it is of little consequence to the analysis here.

³Utility is linear in private consumption to avoid the question of private saving which adds complication and little else.

⁴Assuming a small open economy.

It is generally agreed upon that optimal government policy should be countercyclical, and in fact this is what we see in the majority of the developed world.⁵ This is justified through a variety of arguments including tax-smoothing or simply smoothing in that discretionary spending should remain constant in the face of temporary fluctuations. Given the framework here, countercyclical policy is optimal as voters wish to smooth public spending in the face of uncertainty, both due to inefficiencies in the tax system and concavity in the utility of public goods. The extent to which this can be achieved will, however, be constrained by the political environment as discussed below.

Government

The government consists of an elected party whose actions each period are taken by individual politicians. Politicians themselves are transitory and last only one period, but are not completely irresponsible and care about the party and its re-election.⁶

There is no shortage of explanations as to why governments may overspend.⁷ Here we will abstract away from these and assume simply that governments gain utility from spending your money. There are, however, no “rents”, as in many political agency models, and funds are never completely wasted. This is an attempt to capture the pressures on governments from bureaucracies, interest groups, etc., as well as the benefits politicians receive personally from handing out money. Aside from spending, the politician’s utility depends

⁵For discussions of these and related issues see Lane (2003), Talvi and Vegh (2005) or Alesina et. al (2007).

⁶If politicians last longer than a single period then there is the question of re-electing an incumbent vs. a challenger. There are many papers that consider this issue, but it adds little to the analysis here, so we avoid the complication.

⁷See Persson and Tabellini (2000) or Besley (2006) for a description of the literature.

on a sense of party loyalty, responsibility and/or legacy concerns. Formally, politician utility will be a function of current spending g_t and a constant term L , which captures loyalty or legacy concerns. Let the function be separable and linear so that politician utility is given by

$$G(g_t, L) = g_t + L = \delta(\tau_t)y_t - (1 + r)b_{t-1} + b_t + L, \quad (5.2)$$

where the politician does not receive L if the party is not re-elected. Party loyalty L is an attempt to capture the idea that the incumbent is a member of a party whose horizon is longer term. This can also or partly be viewed as a legacy or ego term in that a politician may get utility from being viewed as responsible and competent, not to mention ethical obligations and employment prospects after politics. Thus, not only do they owe the party and care about re-election, but they wished to be viewed in a good light once they are out of power.

Politics

The approach to politics is adapted from Barro (1973), Ferejohn (1986) and Persson, Roland and Tabellini (1997). In this set-up, elections are held at the end of each period, after government policy is chosen. Promises are meaningless and politicians can only be held accountable for misbehaving after the fact. This is achieved simply by assuming that, in such a case, the politician is ousted and an identical challenger is brought in. This amounts to a simple agency problem, in which voters are assumed to coordinate on an incentive compatible voting rule (one that the politician would rather follow than not), which is observed by the politician before policy is undertaken. Alesina and Tabellini

(2007) extend this approach to allow for public debt and use this theoretical framework to explain the existence of procyclical fiscal policy observed in many developing nations.

Generally, the government is assumed to have some informational advantage that is exploited to obtain rents. However, due to the nature of the relationship between the government and its public, this informational asymmetry is not necessary to obtain a sub-optimal outcome.⁸ We will assume here that voters are fully informed but can still only punish politicians in the next election. The very fact that they are in power allows incumbents to overspend regardless of whether voters observe it. This must be accounted for in the voters' calculus, and thus saving will not be optimal in general. Allowing for an informational asymmetry here would add some realism and exacerbate any inefficiencies, but is cumbersome and is not necessary for the analysis.

At the beginning of period t , income is realized and the state is represented by (b_{t-1}, y_t) . A voting rule is a tax rate τ and saving/borrowing level b which implies spending g . Given the rule and the state, an incumbent chooses taxes and public spending. If the government follows the rule, the party is re-elected; and if not, they are ousted.⁹ The timing is as follows:

1. The period begins with debt/saving level $(1 + r)b_{t-1}$.
2. Voters observe current income y_t and set a voting rule, which is a tax rate τ_t and borrowing b_t (implying a current spending level g_t).
3. The government chooses τ_t, b_t and thus g_t .

⁸Persson, Roland and Tabellini (1997) make this distinction explicitly when they distinguish between what they term information rents from those attained from power.

⁹The government would always tax and spend as much as possible so there is no concern of too little public spending.

Voters wish to set a re-election rule that optimizes both private consumption and public spending to the extent possible (given the behaviour of the incumbent government). Before deriving an incentive compatible voting rule, we must consider the behaviour of an incumbent politician. When not seeking re-election, utility is given by the highest possible value of current spending given as

$$G(g_t, 0) = \bar{g}_t = \delta(\bar{\tau})y_t - (1 + r)b_{t-1} + \bar{b},$$

where $\bar{\tau}$ is the highest that taxes can be set (exogenous) and \bar{b} is the limit on borrowing.¹⁰ The debt limit could be a rule (as will be discussed later), but for now can be interpreted as the most that can be repaid with certainty. Politicians will abide by the voting rule if the following inequality holds

$$\delta(\tau_t)y_t + b_t \geq \delta(\bar{\tau})y_t + \bar{b} - L,$$

where τ_t, b_t are the tax rate and borrowing/saving level under the rule. The next section considers the behaviour of taxes and debt in a two-period setting. Once a framework is established, we can consider the possible value in constraining the government.

5.2.2 Optimal spending/voting

There are two periods, 1 and 2. In the second period, the voter chooses an optimal tax rate given income y_2 ,¹¹ the current debt $(1 + r)b_1$ and the incentives of the politician (there is no borrowing in the last period). The problem is solved recursively, so that in period

¹⁰The maximum tax rate could be considered the highest taxes that can be set without revolt. This is likely related to current public opinion (i.e., $\bar{\tau} = f(\tau_t)$), but for simplicity is fixed here.

¹¹There is no uncertainty here. We could allow for a random draw in period two and thus maximize expected utility, but this adds little.

1 the voter chooses both taxes and borrowing/saving given optimal behaviour in period 2, which is determined by

$$\max_{\tau_2} (1 - \tau_2)y_2 + h(\delta(\tau_2)y_2 - (1 + r)b_1) \quad (5.3)$$

$$\text{s.t. } \delta(\tau_2)y_2 \geq \delta(\bar{\tau})y_2 - L. \quad (5.4)$$

Let τ_2^* be the solution to this problem which will either equalize the marginal value of public and private spending (when unconstrained) or satisfy the political constraint with equality (when constrained). Further, denote period 2 utility at the optimum τ_2^* as the value function $V_2(b_1)$. If initial debt is d (exogenous) and β is the discount factor, the period 1 problem is

$$\max_{\tau_1, b_1} (1 - \tau_1)y_1 + h(\delta(\tau_1)y_1 + b_1 - d) + \beta V_2(b_1) \quad (5.5)$$

$$\text{s.t. } \delta(\tau_1)y_1 + b_1 \geq \delta(\bar{\tau})y_1 + \bar{b} - L \quad (5.6)$$

$$b_1 \leq \bar{b}. \quad (5.7)$$

The corresponding Lagrangian expression is

$$\mathcal{L} = (1 - \tau_1)y_1 + h(\delta(\tau_1)y_1 + b_1 - d) + \beta V_2(b_1) + \lambda_1(\delta(\tau_1)y_1 + b_1 \quad (5.8)$$

$$- \delta(\bar{\tau})y_1 - \bar{b} + L) + \lambda_2(\bar{b} - b_1). \quad (5.9)$$

The first-order conditions are,

$$[h'(\delta(\tau_1)y_1 + b_1 - d) + \lambda_1]\delta'(\tau_1) = 1 \quad (5.10)$$

$$h'(\delta(\tau_1)y_1 + b_1 - d) + \beta(1 + r)h'(\delta(\tau_2)y_2 - (1 + r)b_1) + \lambda_1 - \lambda_2 = 0. \quad (5.11)$$

Equation (5.10) is the equalization of marginal returns between private and public consumption. Condition (5.11) is a familiar Euler equation that relates the value of public consumption between the two periods. The optimal voting rule obviously depends on the importance of the two constraints, which create 4 possible outcomes considered in turn.

Case 1: $\lambda_1 = \lambda_2 = 0$

In the unconstrained case, the voter would like to use public debt to smooth the distortions brought about through tax finance (as well as smooth public consumption over periods). If income is low and/or debt is high today relative to tomorrow, then it is optimal to borrow to cover shortfalls rather than use inefficiently high taxes. Denoting optimal policies with a *, we have

Lemma 7 *Borrowing is increasing in initial debt d and decreasing in y_1 . Taxes are increasing in both y_1 and d so that*

$$\frac{\partial b_1^*}{\partial d} > 0, \quad \frac{\partial b_1^*}{\partial y_1} < 0, \quad \frac{\partial \tau_1^*}{\partial d} > 0, \quad \frac{\partial \tau_1^*}{\partial y_1} > 0.$$

We see that tax and borrowing behaviour are quite simple, but may be distorted by either politics or borrowing constraints. With the former, there will be too much overall spending. With the latter, there will be too much tax financing. If both constraints bind, there will be too much spending *and* taxes will be too high relative to debt.

Case 2: $\lambda_1 > 0, \lambda_2 = 0$

When the political constraint binds, then there is too much public spending. As $\lambda_2 = 0$, however, the mix of financing between debt and taxes given that level of spending remains

optimal.

Case 3: $\lambda_1 = 0, \lambda_2 > 0$

Here the ability to use debt financing is hampered by the constraint, and thus public spending will be lower than optimal as it is more costly.

Case 4: $\lambda_1 > 0, \lambda_2 > 0$

When both constraints bind, there is no choice in that debt is at the limit and taxes are defined by the participation constraint. It is unclear whether there is too much or too little spending in this case.

With this characterization of optimal debt and tax behaviour, we can consider the value of limiting governments through \bar{b} , which up to this point has been considered given. Denote total utility by $W(\bar{b})$ which is utility under optimal tax and spending behaviour characterized by (5.5). Changes to \bar{b} would only have an impact in cases 2-4. In particular, the envelope theorem implies

$$\frac{\partial W}{\partial \bar{b}} = -\lambda_1 + \lambda_2. \quad (5.12)$$

Of course, these multipliers are both non-negative, so the impact is generally ambiguous. If we are in case 2, the debt limit is not binding but the political constraint is; then tightening it will increase utility as it reduces the politicians' leverage. In case 3, the opposite is true and reducing \bar{b} limits the ability to borrow and has no benefit. In case 4 it is not clear whether changes in the debt limit will be harmful or beneficial. A tighter debt limit would reduce the politicians' leverage, but if politics just lead to higher tax financing it may do more harm than good. As both constraints bind in this case, we can characterize the impact

of changes in \bar{b} quite simply as

$$\frac{\partial W}{\partial \bar{b}} = h'(\delta(\bar{\tau})y_1 - \frac{L}{y_1} + \bar{b} - d) + \beta(1+r)h'(\delta(\tau_2)y_2 - (1+r)\bar{b}). \quad (5.13)$$

Define \bar{b}^* as the debt limit which satisfies $\partial W(\bar{b}^*)/\partial \bar{b} = 0$. For a debt limit greater than this, further increases have a negative impact on utility. For a debt limit less than this, increases have a positive effect.¹² Assume that $\beta(1+r) = 1$,¹³ then \bar{b}^* is defined by

$$\bar{b}^* = \frac{1}{2+r} \left(\delta(\tau_2)y_2 - \delta(\bar{\tau}) + \frac{L}{y_1} + d \right). \quad (5.14)$$

We could do some comparative statics calculations to further our intuition on such changes; and while this may shed some light on the issue, the important question that remains is when we expect these constraints to bind. Is it worth the possible inefficiency of being up against a debt limit if it helps control politician overspending? This problem is quite complex, so we shall simplify. Consider the following which is assumed to hold throughout.

Assumption 2 $L > [\delta(\bar{\tau}) - \delta(\tau_1)]y_1$

Effectively, this says that an incumbent is limited in the amount it can overspend using taxes. This will hold if $\bar{\tau}$ is not too large, if there is sufficient curvature in $\delta(\cdot)$ or the politician is “benevolent enough”. For instance, we could define $\bar{\tau} = \tau + \epsilon$ where ϵ is small so that governments can only raise taxes by some “reasonable” amount above the optimum without being thrown out of office or not being able to pass the legislation.¹⁴ This seems reasonable and there are a variety of papers that make similar assumptions. For instance, one could assume that taxes are generally more easily observed than debt

¹²We ensure that this is indeed the case in the next section. In particular, the proof for Proposition 11 shows that there is a unique debt limit that optimizes indirect utility.

¹³This assumption is made purely for simplicity and can be relaxed without difficulty.

¹⁴Of course, then we would have to bound the choice set of τ appropriately.

financing when there is a lack of transparency over government actions, due perhaps to creative accounting practices.¹⁵ When rewriting the political constraint, we note that if assumption 1 holds, both constraints cannot bind at the same time.

$$b_1 \geq (\bar{b}) - L + [\delta(\bar{\tau}) - \delta(\tau_1)]y_1. \quad (5.15)$$

We are now concerned only with cases 2 and 3. Either the political constraint is binding *or* the debt constraint is binding. Changes to \bar{b} have opposite impacts on both of these constraints so the question of optimal debt limits becomes one of when we expect these constraints to bind.

The value in government debt in this model is two-fold: to smooth tax distortions and optimize the timing of public consumption under a varying income process. In fact, if we fix the tax rate, the timing of funds is still an issue and can be looked at as a simple one-dimensional constrained spending/savings problem. When assumption 1 holds, the two problems are essentially equivalent when the focus is on debt controls (although the interpretation is slightly different). The impacts are the same in that a debt control can be welfare enhancing (through better behaved governments), or welfare reducing (if the constraint binds and more borrowing is constrained). This simpler version is outlined below and will be used to consider optimal restrictions on \bar{b} . A note before moving on is that, if assumption 1 fails to hold, then the two problems are generally not equivalent and the question of constraining governments must consider limiting tax behaviour as well. This is left to further work.

¹⁵For example, Alesina and Tabellini (2007) make an informational assumption of this type.

Fixed taxes

This section outlines the two-period problem with a fixed tax rate. Voter utility is simply given by public good consumption $h(g)$ which is ideally smoothed over each period. Public good spending in period t is now simply

$$g_t = y_t - (1 + r)b_{t-1} + b_t,$$

where y_t is now considered government revenue in period t (as with a fixed tax rate and random income above). The voter gives the politician a voting rule which governs spending in that period. A politician either follows the rule and gets the party re-elected or spends as much as possible and is ousted. The political constraint is

$$b_t \geq \bar{b} - L.$$

The voters' problem is the same as above; given an income draw and a current debt, they choose a level of borrowing/saving that is optimal given that the political constraint is not violated. As above, there are two periods and the problem will be solved recursively. In the last period there is no uncertainty and no borrowing, so that all current wealth is just consumed. In period 1, borrowing is chosen to optimize consumption between the last two periods to the extent possible given politician incentives.

Period 2:

Simply consume all wealth, so that utility is given by

$$h(y_2 - (1 + r)b_1).$$

Period 1: The voter sets the rule to optimize borrowing subject to the constraints, which amounts to solving

$$\begin{aligned} \max_{b_1} \quad & h(y_1 - (1+r)b_0 + b_1) + \beta h(y_2 - (1+r)b_1) \\ \text{s.t.} \quad & \bar{b} \geq b_1 \geq \bar{b} - L. \end{aligned}$$

Again assuming $\beta(1+r) = 1$, in the interior, optimal borrowing is given by

$$b_1^*(y_1, b_0) = \frac{y_2 - (y_1 - (1+r)b_0)}{2+r}. \quad (5.16)$$

We can see that borrowing when unconstrained is strictly increasing (decreasing) in future (current) wealth. For the problem to be interesting, there must be some period 1 wealth levels at which the constraints will bind. Let \underline{w} be that level at which borrowing is constrained from above (would like to borrow more), and let \bar{w} be the wealth level such that they would like to save more, but can not due to politics. These cut-offs are defined formally by

$$b_1^*(\underline{w}) = \bar{b} \Rightarrow \underline{w} = y_2 - \bar{b}(2+r)$$

and

$$b_1^*(\bar{w}) = \bar{b} - L \Rightarrow \bar{w} = y_2 - (\bar{b} - L)(2+r).$$

Figure 5.1 depicts optimal borrowing behaviour in period 1.

5.3 Optimal debt limits

It is useful to consider another period before discussing optimal \bar{b} . Period 1 behaviour is described completely by $b_1^*(y_1, b_0)$ as described in (5.16) and the wealth cutoffs \bar{w} and

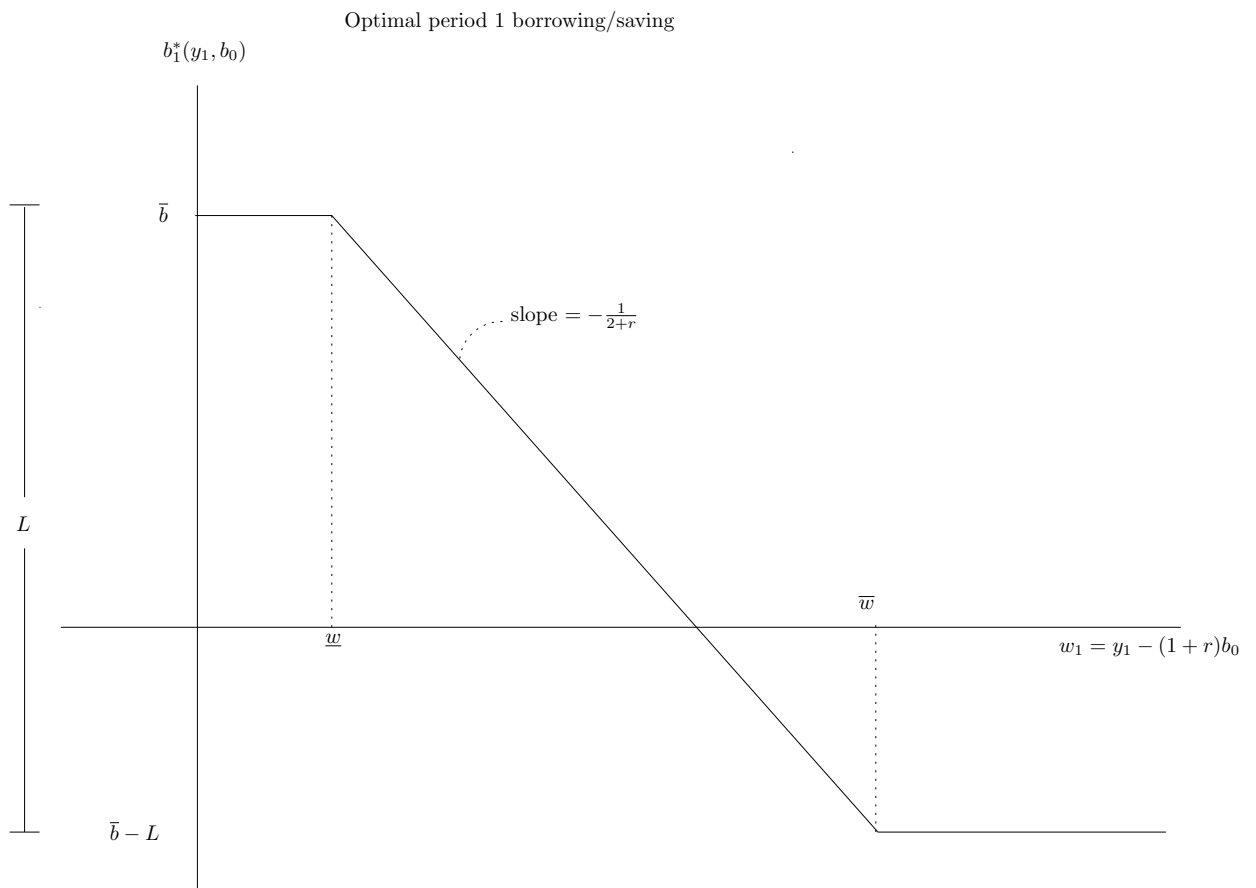


Figure 5.1: Saving behaviour

w . Consider the period preceding period 1 which we label period 0. Assume the voter is unconstrained in this initial period (otherwise there is no decision to make). The value function of the entire problem from a period 0 perspective is then (where w_0 is initial wealth)

$$\begin{aligned}
V(w_0) = \max_{b_0} & h(w_0 + b_0) + \\
& \beta \int_{\underline{y}}^{\overline{w} + (1+r)b_0} [h(y_1 - (1+r)b_0 + \bar{b}) + \beta h(y_2 - (1+r)\bar{b})] f(y_1) dy_1 \\
& + \beta \int_{\underline{w} + (1+r)b_0}^{\overline{w} + (1+r)b_0} [h(y_1 - (1+r)b_0 + b_1^*) + \beta h(y_2 - (1+r)b_1^*)] f(y_1) dy_1 \\
& \beta \int_{\overline{w} + (1+r)b_0}^{\overline{y}} [h(y_1 - (1+r)b_0 + \bar{b} - L) + \beta h(y_2 - (1+r)(\bar{b} - L))] f(y_1) dy_1.
\end{aligned}$$

Let $b_0^* = b_0^*(w_0)$ be optimal borrowing in period 0 which satisfies the expression above. Borrowing increases consumption of public goods today and increases debt tomorrow. In the two-period case, we are concerned with the chance of ending up constrained which is exogenous. Considering the period 0 problem highlights the fact that a longer time horizon means more sophisticated voting behaviour can occur. From the period 0 perspective the optimal choice now influences the chance of ending up constrained in period 1.

With this in mind, we return to our focus of characterizing optimal debt limits. To gain some intuition, consider that changes in the debt limit have the following effects; when the constraints are binding from above and below (which relate directly to cases 2 and 3 above).¹⁶

¹⁶Leibniz integral rule and the envelope theorem are both used to attain the derivative.

$$\frac{\partial V(w_0)}{\partial \bar{b}} = \underbrace{\beta \int_{\underline{y}}^{w+(1+r)b_0} [h'(y_1 - (1+r)b_0^* + \bar{b}) - h'(y_2 - (1+r)\bar{b})] f(y_1) dy_1}_{> 0, \text{ effect of easing debt constraint when borrowing is constrained}} +$$

$$\underbrace{\beta \int_{\bar{w}+(1+r)b_0}^{\bar{y}} [h'(y_1 - (1+r)b_0^* + \bar{b} - L) - h'(y_2 - (1+r)(\bar{b} - L))] f(y_1) dy_1}_{< 0, \text{ increasing the borrowing constraint gives the politician more leverage}}.$$

Proposition 11 *There exists a unique optimal debt limit \bar{b}^* which is increasing in w_0 .*

The optimal \bar{b} depends on initial wealth. With higher initial wealth (and thus less borrowing/more saving in period 0), there is less chance of being constrained by the debt limit and more incentive to save in period 1. In fact, if initial wealth is high enough there may be no chance of hitting the upper bound. To gain some intuition, consider the following simple example where we can solve explicitly for the optimal debt limit.

5.3.1 Numerical Example

Let the function $h()$ be given by

$$h(g) = g - \frac{g^2}{2M},$$

where M is large enough to ensure that $h'()$ is always non-negative. Furthermore, let government resources or wealth take on three values $y_L \leq \underline{w}$, $y_M \in (\underline{w}, \bar{w})$ and $y_H \geq \bar{w}$ which occur with probabilities p_L , p_M and p_H , respectively. Let $y_2 = y_M$ and b_1^* be the optimal borrowing when unconstrained.¹⁷ From above we have expected welfare over the

¹⁷There is no need to solve this explicitly as this term will drop out due to the envelope theorem.

two periods as

$$\begin{aligned}
 W = & p_L \left[y_L + \bar{b} - \frac{(y_L + \bar{b})^2}{2M} + \beta \left(y_M - (1+r)\bar{b} - \frac{(y_M - (1+r)\bar{b})^2}{2M} \right) \right] + \\
 & p_M \left[y_M + b_1^* - \frac{(y_M + b_1^*)^2}{2M} + \beta \left(y_M - (1+r)b_1^* - \frac{(y_M - (1+r)b_1^*)^2}{2M} \right) \right] + \\
 & p_H \left[y_H + \bar{b} - L - \frac{(y_H + \bar{b} - L)^2}{2M} + \beta \left(y_M - (1+r)(\bar{b} - L) - \frac{(y_M - (1+r)(\bar{b} - L))^2}{2M} \right) \right].
 \end{aligned}$$

Optimizing with respect to \bar{b} yields the following optimal debt limit

$$\bar{b}^* = \frac{p_L(y_M - y_L) - p_H(y_H - y_M) + p_H(2+r)L}{(2+r)(p_L + p_H)}.$$

We can see that the optimal debt limit is

- increasing in L
- increasing in the probability and “severity” of the low state
- decreasing in the probability and “size” of the good state
- positive or negative

The optimal rule depends on the chances of ending up debt or savings constrained as expected. The more well behaved politicians are (the larger is L), the larger the optimal debt limit to allow for debt financing should it be necessary.

5.4 Infinite horizon

This section considers the issue in an infinite horizon setting. We will analyze the simpler case where tax rates are fixed for continuity.¹⁸ As above, private consumption is fixed (and thus ignored) and government revenues are an i.i.d process. Denote current debt by d , current borrowing b and let a ' represent next period values. The voter is now optimizing over

$$E_0 \sum_{t=0}^{\infty} \beta^t h(g_t).$$

Government incentives and political structure are the same as above. The problem is characterized by the following Bellman equation.

$$P(d, y) = \max_b [h(y - (1 + r)d + b) + \beta EP(b, y')], \quad (5.17)$$

where the maximization is again subject to

$$b \geq \bar{b} - L \quad (5.18)$$

$$b \leq \bar{b}. \quad (5.19)$$

The corresponding Lagrangian expression is

$$\mathcal{L} = h(g) + \beta EP(b, y') + \lambda_1 [b - (\bar{b} + L)] + \lambda_2 [\bar{b} - b],$$

yielding the first-order condition

$$h'(g) = -\beta E (P_b(b, y')) + \lambda_2 - \lambda_1.$$

¹⁸We can easily extend this to the more general case with tax smoothing as this section does not attempt to derive any analytic results.

This and the envelope condition give the following stochastic Euler equation

$$h'(g) = \beta(1 + r)E(h'(g')) + \underbrace{\lambda_2 - \lambda_1}_{\text{inter-temporal wedge}} .$$

A social planner would smooth consumption and set government spending such that there is a random walk of marginal utility.¹⁹ The political environment generally prevents this and creates an “inter-temporal wedge”, which as in the two-period case depends on which constraint binds. There is an important difference in the infinite horizon problem, however. The initial wealth, which was fixed in the finite period problem, now evolves due in part to any debt limit. Thus we can capture the behaviour of public monies over the short and long run. We expect that the imposition of a debt control will have short-run costs that will be higher if the initial wealth is low. These costs are offset by the ability to control politicians in the long run. For future work, as a complement to the two period model, it would be interesting to characterize the effect of debt limits on the stationary distribution of wealth at least for some specified function and parameter values.

5.5 Conclusions

Politicians hinder the ability of the state to accumulate wealth. Incumbents are less likely to behave during good economic times as they can not resist spending money when it is around. The process of elections mitigates the problem as it imposes some degree of

¹⁹In the unconstrained case the planner will in fact save an infinite amount to finance an infinite consumption as the problem is not stationary. This can be avoided by assuming that at some very large level of consumption, marginal utility falls to zero. Otherwise, we could consider a model where the politician gets some extra utility from spending when not seeking re-election and that could serve to limit asset accumulation.

discipline on incumbents. That this is insufficient is the reason for this chapter and many like it in the vast literature on political economy.

Imposing restrictions on borrowing/saving seems a natural solution to the problem, but the impacts of such a restriction are generally ambiguous. Allowing governments flexibility has value and the gains of a tighter limit due to control of politicians is offset by a lessened ability to smooth consumption (and tax distortions in the more general case). The value of imposing borrowing limits depends on how irresponsible governments are, which is captured simply in the model by the parameter L . The larger is L , the more likely a limit will do more harm than good.

The results also suggest that the current fiscal position is important when considering a debt constraint. If one can be imposed during good economic times, it is more likely to be beneficial. If governments spend and borrow too much, however, we have a chicken and egg problem, as imposing debt controls may not be ideal if the government is always in debt. To address this we must be able to describe the evolution of public debt and specifically the steady state distribution of wealth. Doing so requires a fully dynamic framework, as is laid out in section 5.4 and is left for future work.

Chapter 6

Summary and Conclusions

We have analyzed a variety of economic questions throughout this thesis. Using simple models, we have discussed the division of education budgets, educational institutions and the control of public debt.

The results of chapter three suggest that redistribution through education may or may not be optimal when policy objectives are defined by a standard consumption/leisure model. In particular, when the optimal first-best (perfect information) policy is non-zero, the second-best policy will be more progressive than the first-best. Implying that education spending on low-productivity types will mitigate distortions generated by the tax system.

For a large class of functions however, the optimal first-best policy will indeed be to spend nothing on the low-productivity types. In this case it is ideal to spend all educational resources on high-productivity students and make up for inequities through money transfers even in the second-best. There are a variety of reasons why this is the case.

Spending in the low-productivity location results in a reduction in total output (and consequently transfers), as that of high-types is decreased, while that of the low-types increases relatively less. It is obvious that spending on less productive students can reduce output; less obvious is the role played by the tax system. For a fixed set of skills, low-types are discouraged from working by the tax scheme, so any benefit from increasing their human capital is not fully realized. Further, increasing the productivity of the low-types tightens the incentive constraint in the tax problem, which creates an even greater distortion in the labour market.

Finally, we note that regardless of functional forms, there are no social preferences under which an equal opportunity policy is called for when there is individual heterogeneity amongst groups.

A further important caveat is that the *talented* poor, who will possibly achieve a reasonable outcome regardless, may be the major beneficiaries of these policies. This will be true when the return to education funding is small for those on the low end of the distribution of individual endowments. We have seen that these types are hurt the most by equal opportunity policies, as they will see little but a reduction in money transfers through the tax system. If our redistributive goals are to help the least fortunate (which is implied by the standard social welfare analysis), policies that attempt to equalize opportunities will certainly not be optimal in this case.

The point of this chapter is not to argue against policies designed to bring about more equality in productivity, but rather to bring to light the conflict between various objectives. The merits of equal opportunity policies clearly extend beyond the standard limited concepts of utility maximization. The value of education undoubtedly transcends its impact on

consumption outcomes both individually and in the aggregate. The impacts of educational spending on self esteem, crime, voter savvy, general social cohesion, etc., are central to the policy debate. Although sometimes difficult to model in an economic framework (at least in an interesting way), these concerns are of great importance nonetheless.

We have shown that, if we accept the validity of such arguments and choose policies which create greater equality in productivities amongst socio-economic groups, then it is imperative that spending decisions take into account dynamic implications. In particular, when individuals are mobile and there are local externalities in the production of human capital, then a complete equalization of opportunities will be too extreme. If the objective is to sever the link between socio-economic status and human capital outcomes, simply compensating for local effects in each period is not reasonable. An alternative policy objective is to maximize human capital in the poor region over the long run. This would require a deeper understanding of both the importance of local externalities and of the human capital function in general. An ideal education policy will depend on the relative importance of various objectives, which as we have seen, are generally in conflict. Chapter three highlights the need to be aware of both why we wish to equalize opportunities, and, if so, to be cognizant of the possible dynamic inconsistency inherent in such policies.

Chapter four considers the impacts of two different educational institutions and analyzes the relative merits of each. A specialized system offers the opportunity to gain a larger amount of human capital in a given field, but comes with the risk of choosing a field to which one is not suited, which leads to a low outcome.

Moreover, we have seen that under a specialized system, an adverse selection problem

will arise that further reduces the returns to education. This is because there will be different productivity types in each labour market that are costly to screen. This arises here and not in the general system because “mistakes” in education acquisition result in ex-post differences between people who were identical ex ante (identical in their endowments before undertaking education).

A general system, on the other hand, provides insurance against a low-productivity outcome, but comes at the cost of reduced productivity in any given field. Also, under this type of system, individuals who are not particularly suited to a specific field will be indifferent over which field they wish to pursue a career. As a result, if there are coordination problems which exist in the labour market, a general system will exacerbate this friction relative to a specialized one.

In the final chapter, we consider the importance of politics and the potential benefit of debt controls. In the model, politicians hinder the ability of the state to accumulate wealth. Incumbents are less likely to behave during good economic times as they can not resist spending money when it is around. The process of elections mitigates the problem as it imposes some degree of discipline on incumbents. This discipline may be improved upon through some explicit set of restrictions.

Imposing restrictions on borrowing/saving seems a natural solution to the problem, but the impacts of such a restriction are generally ambiguous. Allowing governments flexibility has value and the gains of a tighter limit due to control of politicians is offset by a lessened ability to smooth consumption and distortions brought about by the tax system. The value of imposing borrowing limits depends on how irresponsible governments are, as well as how good a job the electoral process does in policing incumbents in the first place.

The results also suggest that the current fiscal position is important when considering a debt constraint. If one can be imposed during good economic times, it is more likely to be beneficial. If governments spend and borrow too much, however, we have a chicken-and-egg problem as imposing debt controls may not be ideal if the government is always in debt. To address this, we must be able to describe the evolution of public debt and specifically the steady state distribution of wealth. Doing so requires a fully dynamic framework as is laid out in section 5.4 and is left for future work.

6.1 References

Alesina, A. and G. Tabellini (2008), “Why is Fiscal Policy often Procyclical?” *Journal of the European Economic Association*, 6(5), 1006-1036.

Arrow, K. (1971), “A utilitarian approach to the concept of equality in public expenditures” *Quarterly Journal of Economics*, 85, 409-415.

Barro, R. (1973), “The control of politicians: An economic model”, *Public Choice*, 14, 19-42.

Barro, R. (1979), “On the determination of the public debt”, *Journal of Political Economy*, 87, 940-971.

Bassetto, M. and T. Sargent (2006), “Politics and Efficiency of Separating Capital and Ordinary Government Budgets”, *Quarterly Journal of Economics*, 121(4), 1167-1210.

Battaglini, M. and S. Coate (2008), “A dynamic theory of public spending, taxation and debt”, *American Economic Review*, 98:1, 201-236.

Bayer, P., R. McMillan and K. Reuben (2005), “An equilibrium model of sorting in an urban housing market”, Working Paper.

Bayer, P., F. Ferreira and R. McMillan (2007), “A unified framework for measuring preferences for schools and neighbourhoods”, *Journal of Political Economy*, 115(4), 588-638.

Benabou R. (1996a), "Equity and efficiency in human capital investment: The local connection", *Review of Economic Studies*, 63,237-264.

Benabou R. (1996b), "Heterogeneity, stratification, and growth: Macroeconomic implications of community structure and school finance", *American Economic Review*, 86, 584-609.

Benabou R. (1993), "The workings of a city: Location, education, and production", *Quarterly Journal of Economics*, 108(3),619-652.

Berne R. (1988), "Equity issues in school finance", *Journal of Education Finance*, 14, 159-180.

Bertocchi, C. and M. Spagat (2004) "The Evolution of Modern Educational Systems: Technical Vs. General Education", *Journal of Development Economics*, Volume 73, Issue 2, April 2004, 559-582.

Besley, T. (2006), "Principled Agents? The political economy of good government, the Lindahl lectures", *Oxford University Press*.

Besley, T. and M. Smart, (2007), "Fiscal restraints and voter welfare", *Journal of Public Economics*, 91, 755-773.

Boadway R. and P. Pestieau (2006), "Tagging and redistributive taxation", *Annales D'conomie et de statistique*, 83/84, 123-147.

Boadway R., Marceau, N. and M. Marchand (1996), "Investment in education and the time

inconsistency of redistributive tax policy”, *Economica*, 63, 171-189.

Bovenberg, L. and B. Jacobs (2005), “Redistribution and education subsidies are siamese twins”, *Journal of Public Economics*, 89, 2005-2035.

Brett, C. and J. Weymark (2008), “The impact of changing skill levels on optimal nonlinear income taxes”, *Journal of Public Economics*, 92, 1765-1771.

Brunello, G. (2004), “Stratified or Comprehensive? Some Economic Considerations on the Design of Secondary Education”, CESifo DICE Report, 2004, 7-10

Bruno (1976), “Equality, complementarity and the incidence of public expenditures”, *Journal of Public Economics* 6, 395-407.

Cremer, H., Pestieau, P. and M. Racionero (2008), “Unequal wages for equal utilities”, Working Paper.

de Bartolome, C. (1990), “Equilibrium and inefficiency in a community model with peer group effects”, *Journal of Political Economy*, 98(1), 110-132.

Ding, W. and S. Lehrer (2007), “Do peers affect student achievement in China’s secondary school?”, *Review of Economics and Statistics*, 89(2), 300-312.

Dworkin, R (1981), “What is equality? Part 2: Equality of resources”, *Philos Publ Affairs*, 10, 283-345.

Epple, Dennis, Elizabeth Newlon and Richard Romano (2002), “Ability tracking, school competition, and the distribution of educational benefits,” *Journal of Public Economics*,

83:1, 1-48.

Ferejohn, J. (1986) "Incumbent performance and electoral control", *Public Choice*, 50, 5-25.

Fernandez, R. and R. Rogerson (1996), "Income distribution, communities and the quality of public education", *Quarterly Journal of Economics*, 111(1), 135-164.

Fleurbaey M. and F. Maniquet (2005), "Compensation and responsibility", in K. J. Arrow, A. K. Sen, K. Suzumura eds., *Handbook of Social Choice and Welfare*, vol. 2, North-Holland.

Fleurbaey M., Gary-Bobo R. and D. Maguain (2002), "Education, distributive justice, and adverse selection", *Journal of Public Economics*, 84, 113-150.

Goux, D. and E. Maurin (2007), "Close neighbours matter: Neighbourhood effects on early performance at school", *Economic Journal*, 117, 1193-1215.

Gradstein, Mark, Moshe Justman and Volker Meier (2005), *The Political Economy of Education: Implications for Growth and Inequality*, MIT Press.

Guesnerie R. and J. Seade (1982), "Nonlinear Pricing in a Finite Economy", *Journal of Public Economics*, 17, 157-180.

Hanushek, Eric and Finis Welch (2006), *Handbook of the Economics of Education*, Vol 1, North Holland Press.

Hare P., and D. Ulph (1979), "Education and income distribution", *Journal of Political Economy*, 87, S193-S212.

Heckman, James and Alan B. Krueger, (2003), *Inequality in America: What role for human capital policies?*, MIT Press.

Hoxby, C. (2000), "Peer effects in the classroom: learning from gender and race variation" NBER Working Paper no. 7867.

Krause, A. (2006), "Redistributive taxation and public education", *Journal of Public Economic Theory*, 8(5), 807-819.

Krueger, D., and K. Kumar(2004), "Skill-specific rather than General Education: A Reason for US-Europe Growth Differences?" *Journal of Economic Growth*, Vol. 9, No.2, pp.167-207.

Lane, P. (2003) "The cyclical behaviour of fiscal policy: evidence from the OECD", *Journal of Public Economics*, 87, 2661-2675.

MacDonald, G (1980), "Person-Specific Information in the Labor Market", *Journal of Political Economy*, 88:3, 578-597.

Mirrlees, J.A. (1971), "An exploration in the theory of optimal income taxation", *Review of Economic Studies*, 38, 175-208.

Montgomery, J. (1991), "Equilibrium Wage Dispersion and Interindustry Wage Differentials", *Quarterly Journal of Economics*, vol. 106, No. 1 (Feb., 1991) 163-179.

Persson, T. and G. Tabellini (2000), *Political economics: Explaining economic policy*, MIT Press.

Riley, J. (1979), "Informational Equilibrium," *Econometrica*, 47, 331-59.

Riley, J. (2001), "Silver signals: twenty-five years of screening and signaling", *Journal of Economic Literature*, 39, 432-478.

Roemer J. (1998), *Equality of Opportunity*, Harvard University Press.

Rogerson, Shimer and Wright (2005), "Search-theoretic models of the labor market: a survey", *Journal of Economic Literature*, XLIII, 959-988.

Rothschild, M. and J. Stiglitz (1976), "Equilibrium in competitive insurance markets: An essay on the economics of imperfect information", *Quarterly Journal of Economics*, 90:4, 629-649.

Sen, A. (1985), *Commodities and Capabilities*, Elsevier, Amsterdam.

Sen, A. (1995), *Inequality Reexamined*, Harvard University Press.

Simula, L. (2007), "Optimality conditions and comparative static properties of non-linear income taxes revisited", Working Paper.

Spence, M. (1973), "Job Market Signaling", *Quarterly Journal of Economics*, 87, 355-379.

Sugden, R. (2004), "Living with unfairness: The limits of equal opportunity in a market

economy”, *Social Choice and Welfare*, 22, 211-236.

Talvi, E., and Vegh, C. (2005). “Tax base variability and procyclical fiscal policy in developing countries”, *Journal of Development Economics*, vol. 78(1), pages 156-190.

Ulph, D. (1977), “On the optimal distribution of income and educational expenditure”, *Journal of Public Economics*, 8, 341-356.

Weymark, J. (1986), “A reduced-form optimal nonlinear income tax problem”, *Journal of Public Economics*, 30, 199-217.

Appendix A

Appendix for Chapter 3

A.1 Tagging and welfare

Throughout the analysis of section 3.3 we assumed that the planner could target education funds to neighbourhoods/regions, but not implement separate tax policies due to limited information ex-post. Do the results of section 3.3 hold if we relax this?

Allowing for separate tax treatments between types is equivalent to assuming the planner can observe one's socio-economic status ex post. If there is no individual heterogeneity, this amounts to perfect information and the planner simply solves the first-best problem laid out in section 3.3.1.

In the more general case, the impact is less obvious. As with the tax problem laid out in section 3.3.3, it is difficult to attain an analytic solution to the education problem in this case. Technically, the only difference between the two problems is that there are only two incentive constraints as opposed to three. These incentive constraints apply to

the high and low θ 's *within* each neighbourhood/region (as θ is still unobservable). It turns out that both Propositions 6 and 7 continue to hold when we allow tagging in the tax problem. Proposition 6 continues to hold because, at the equal opportunity policy, the only heterogeneity that exists is over θ , so the problems are equivalent. Further at $\delta = 0$, the impacts of a change in spending on the poor, although different in these two cases, can be shown to be unambiguously negative without much difficulty.

As with the general model above with no tagging, we can solve this for specific functional forms and parameters. Figures A.1 and A.2 map out welfare as the value function of the tax problem when we allow for differing tax treatments across socio-economic groups. The curves are formulated with the same parameters used in section 3.3.3. Again, the solid curve represents welfare under a utilitarian planner and the dashed curve welfare under a maxmin planner. We see that the results are similar and the optimal policy remains $\delta = 0$ for both.

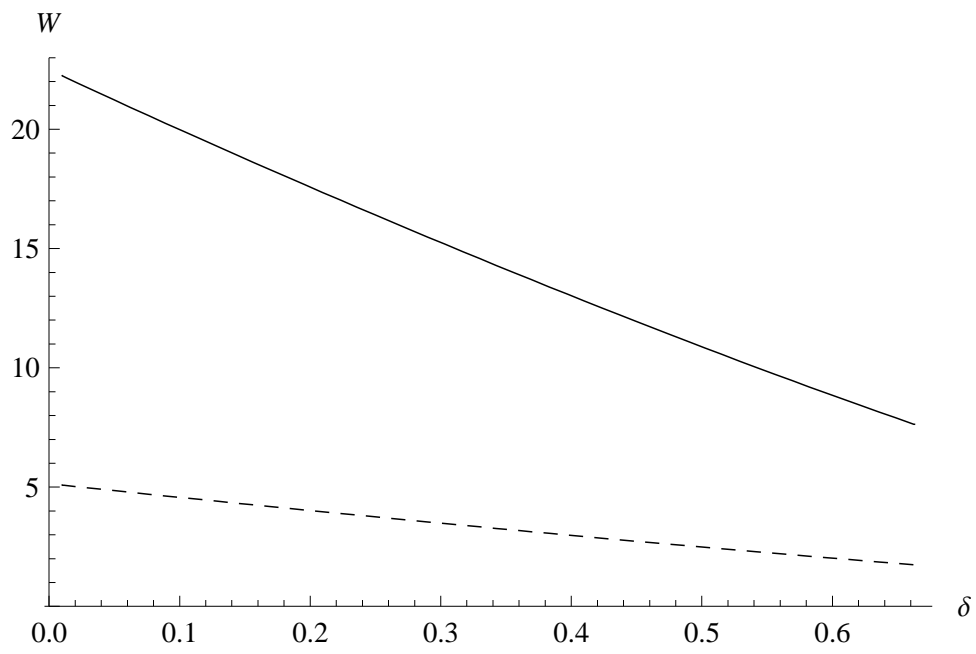


Figure A.1: Tagging case 1

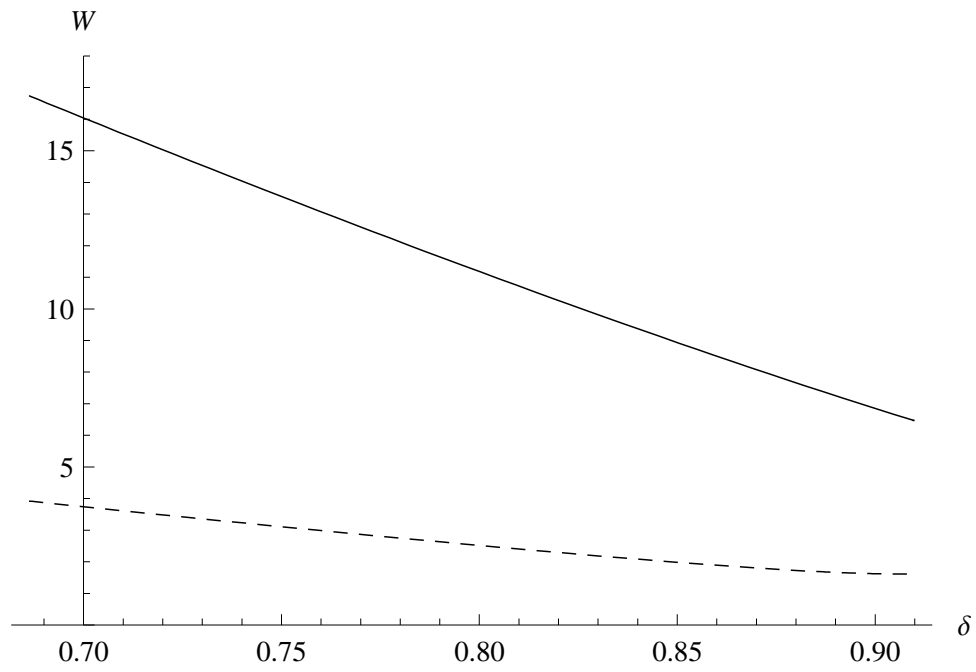


Figure A.2: Tagging case 2

A.2 Proofs

Proof of Proposition 1. After some manipulation we can write

$$\frac{\partial \mathcal{L}}{\partial \delta} = \frac{\epsilon_w}{\delta(1-\delta)} \left[\frac{f'(\frac{y_p}{w_p})}{w_p} y_p - \delta \lambda (y_p + y_r) \right] \quad (\text{A.1})$$

where $\epsilon_w = e^{\frac{w'(e)}{w(e)}}$ is the elasticity of productivity with respect to education funding. When $y_p = 0$, it is obvious that $\partial \mathcal{L} / \partial \delta$ is negative. Thus, if $w(0) = 0$, we see that increasing spending on the poor from a first best of $\delta = 0$ is strictly welfare reducing which is the first part of the proposition. This corresponds to the productivity technology considered in Cremer et.al (2008).

If the solution to the first-best (surplus maximization) problem is characterized by some spending in both regions, then the first-order conditions given by (3.5) imply

$$y_p(1 - \delta^{fb}) = y_r \delta^{fb}.$$

Inserting this into (A.1) yields

$$\frac{\partial \mathcal{L}}{\partial \delta} = y_p \left[\frac{f'(\frac{y_p}{w_p})}{w_p} - \lambda \right].$$

Which can be rewritten, using the first-order conditions on the tax problem (specifically (3.13), and (3.14)), as

$$\frac{\partial \mathcal{L}}{\partial \delta} = y_p \left[\gamma \frac{f'(\frac{y_p}{w_r})}{w_r} \right] \geq 0.$$

This is positive, so that there is an increase in social welfare from an increase in δ , whenever the incentive constraint in the tax problem is binding (i.e. the second-best is different from the first-best). ■

Proof of Proposition 2. After some manipulation we can write

$$\frac{\partial \mathcal{L}}{\partial \delta} = \frac{\epsilon_w}{\delta(1-\delta)} \left[\frac{f'(\frac{y_p}{w_p})}{w_p} y_p - \delta \lambda (y_p + y_r) \right]$$

where $\epsilon_w = \frac{e}{w(e)} \frac{\partial w}{\partial e}$ is the elasticity of productivity with respect to education funding (which equals $\frac{1}{2}$ in this case). We can immediately see that this is negative when $y_p = 0$. We've assumed that $f(l) = \frac{1}{2}l^2$, so that we can solve for y_p and y_r explicitly using the first-order conditions on the tax problem.

$$y_p = \frac{\lambda w_p^2 w_r^2}{w_r^2 - (1-\lambda)w_p^2}, \quad y_r = w_r^2. \quad (\text{A.2})$$

Using these we can see that $\frac{\partial \mathcal{L}}{\partial \delta}$ is negative when

$$w_r^2 + (2\lambda - 1)w_p^2 \geq \frac{\lambda w_p^2 w_r^2}{\delta(w_r^2 - (1-\lambda)w_p^2)}. \quad (\text{A.3})$$

For a utilitarian objective, $\lambda = 1$, so that (A.3) reduces to

$$\frac{[w_r(1-\delta)]^2}{1-\delta} \geq \frac{[w_p(\delta)]^2}{\delta}. \quad (\text{A.4})$$

By assumption, the left (right)-hand side is decreasing (increasing) in δ . Since $\delta \leq \delta_E$, showing this holds for δ_E shows that the derivative is always negative and $\delta = 0$ is optimal.

At δ_E , wages are equal, so this inequality can be written as

$$\delta_E \geq 1 - \delta_E \quad (\text{A.5})$$

which is true by definition. To see this, we know that $w_r(1 - \delta_E) = w_p(\delta_E)$ and that $w_r(e) > w_p(e) \forall e \Rightarrow \delta_E > 1 - \delta_E$. Now turn to the maxmin problem.

For large enough δ , an increase in spending in the poor region can have a positive impact on welfare, depending on the planner's aversion to inequality. For maxmin preferences, $\lambda = \frac{1}{2}$, so that equation (A.3) implies that welfare is decreasing whenever

$$\frac{2}{1+\delta}[w_r(1-\delta)]^2 \geq \frac{[w_p(\delta)]^2}{\delta}.$$

As above, the left (right)-hand side is decreasing (increasing) in δ . However in this case, at δ_E this is unambiguously positive. Since $\frac{2}{1+\delta}[w_r(1-\delta)]^2 - \frac{[w_p(\delta)]^2}{\delta}$ is strictly increasing in δ , we know that if welfare is larger at $\delta = 0$ than δ_E then this proves the result.

First, show that welfare under the policy $\delta = 0$ is greater than welfare under the equal opportunity policy. With the assumptions given, we can derive analytic solutions for consumption and income by using the first-order conditions and the incentive and budget constraints (both of which bind for the maxmin problem). Welfare when $\delta = 0$ is simply $\frac{[w_r(1)]^2}{4}$. Under the equal opportunity policy it is $\frac{[w_r(1-\delta_E)]^2}{2}$, so that welfare is higher under $\delta = 0$ whenever $\frac{w_r(1)}{w_r(1-\delta_E)} \geq \sqrt{2}$, which is exactly the condition in (3.21).

At δ_E , there is no heterogeneity and thus there is no redistribution. It is reasonable to think that welfare is discontinuous at the point where there ceases to be a transfer between the two types. We now show that welfare is left continuous at δ_E . Note that $y_p \rightarrow w_p^2$ and $w_p \rightarrow w_r$ as $\delta \rightarrow \delta_E$. Thus

$$\lim_{\delta \rightarrow \delta_E} W = \lim_{\delta \rightarrow \delta_E} \frac{1}{2} \left(y_p + \frac{1}{2} \left(\frac{y_p}{w_r} \right)^2 - \left(\frac{y_p}{w_p} \right)^2 \right) + \frac{1}{4} w_r^2 = \frac{[w_r(1-\delta_E)]^2}{2}$$

which is exactly the welfare under the equal opportunity policy (again where there is no redistribution). ■

More generally, we know that $2\lambda - 1 \geq 0$, so that a sufficient condition for (A.3) to hold is

$$w_r^2 \geq \frac{w_p^2}{\delta}$$

so that for any degree of inequality aversion, the second-best problem is similar in that welfare is decreasing up to some threshold (which is increasing in the advantage of the rich) and *possibly* increasing for higher δ (this is only a sufficient condition). Although the problem is similar in general, showing that welfare is monotonically increasing after this threshold is difficult and we can not use the analysis as above with the maxmin. Generalizing this result is left to further work.

Proof of Proposition 3. We know that $\partial\mathcal{L}/\partial\delta$ is increasing for a maxmin planner whenever

$$\frac{2}{1+\delta}w_r(1-\delta)^2 \leq \frac{w_p(\delta)^2}{\delta}.$$

This has been shown to hold generally above at δ^{fb} and can easily be seen to be true at the equal opportunity policy where wages are equalized as $1 + \delta_E \geq 2\delta_E$. Thus, if $(1 + \delta)w_p(\delta)^2 - 2\delta w_r(1 - \delta)^2$ is everywhere positive over $[\delta^{fb}, \delta_E]$ so that welfare is monotone in δ , we have the result. For our form of the productivity function, $w_p(\delta) = \delta^\alpha$ and $w_r(1 - \delta) = r(1 - \delta)^\alpha$, this is true when

$$r^2 \leq \frac{1 + \delta}{\delta^{1-2\alpha}(1 - \delta)^{2\alpha}}.$$

Note that for the first-best to be non-zero we require $\alpha < \frac{1}{2}$ so that the term on the right hand side is greater than one. Further we see this holds strictly when $r = 1$, which is the case when there are no differences between types initially.

■

Proof of Proposition 4. A change in education policy is captured by

$$\frac{\partial \mathcal{L}}{\partial \delta} = \frac{\partial \pi}{\partial \delta} \lambda \underbrace{(y_2 - c_2 - y_1 + c_1)}_{\Delta} = \frac{\partial \pi}{\partial \delta} \lambda \underbrace{(t_2 - t_1)}_{\Delta} \quad (\text{A.6})$$

where t_2 and t_1 are the total taxes paid. As $t_2 > t_1$ we see that the term labelled Δ is always positive. Thus an increase in δ is positive (negative) when $\frac{\partial \pi}{\partial \delta} > 0$ (< 0). Thus the first-best policy is optimal. Further, we know that by definition the equal opportunity policy $\delta_E > \frac{1}{2} > \delta^{fb}$ ■

Proof of Proposition 5. Any deviation from the optimum δ^{fb} such as equal funding or the equal opportunity policy results in an increase in the fraction of unskilled workers. Boadway and Pestieau (2006) show that an increase in the fraction of unskilled workers has a negative impact on the welfare of both high and low types as would be expected. More importantly, they show that the negative effect is larger for low types and that this is true for any positive aversion to inequality. ■

Proof of Proposition 6. Under an equal opportunity policy there are only two wage/productivity outcomes:

$$w_l = \theta_l h_p(\delta_E) = \theta_l h_r(1 - \delta_E)$$

$$w_h = \theta_h h_p(\delta_E) = \theta_h h_r(1 - \delta_E)$$

Denote those with high and low productivity (those with high and low θ) with the subscripts h, l , respectively. The optimal tax problem now has only one incentive constraint; thus the lagrangian for this problem can be written

$$\mathcal{L} = \sum_{k \in \{l, h\}} 2\Psi \left(c_k - f\left(\frac{y_k}{w_k}\right) \right) + 2\lambda \left(\sum_k y_k - \sum_k c_k \right) + \gamma \left(c_h - f\left(\frac{y_h}{w_h}\right) - c_l + f\left(\frac{y_l}{w_h}\right) \right)$$

Note that, although rich and poor students have the same productivity, they are affected differently by spending changes at the optimum. In particular, the impact on rich students of a change in δ is greater and of opposite sign than on poor students. After some manipulation, we can write this effect as

$$\frac{\partial \mathcal{L}}{\partial \delta} = \left(2\Psi'(v_l) f'\left(\frac{y_l}{w_l}\right) \frac{y_l}{w_l^2} \theta_l + 2\Psi'(v_h) f'\left(\frac{y_h}{w_h}\right) \frac{y_h}{w_h^2} \theta_h + \gamma \theta_h \left[f'\left(\frac{y_h}{w_h}\right) \frac{y_h}{w_h^2} - f'\left(\frac{y_l}{w_h}\right) \frac{y_l}{w_h^2} \right] \right) \left[\frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} \right]$$

This is negative whenever $\left[\frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} \right] < 0$.

By assumption $h(e)$ is concave and satisfies $h_r(e) > h_p(e)$, $h'_r(e) > h'_p(e) \forall e$ so that $\delta_E > 1 - \delta_E$. Therefore $\left[\frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} \right] = h'_p(\delta_E) - h'_r(1 - \delta_E)$ is negative. ■

Proof Corollary 1. We saw in the proof of Proposition (6) that, at the equal opportunity policy, we have

$$\frac{\partial \mathcal{L}}{\partial \delta} = \left(2\Psi'(v_l) f'\left(\frac{y_l}{w_l}\right) \frac{y_l}{w_l^2} \theta_l + 2\Psi'(v_h) f'\left(\frac{y_h}{w_h}\right) \frac{y_h}{w_h^2} \theta_h + \gamma \theta_h \left[f'\left(\frac{y_h}{w_h}\right) \frac{y_h}{w_h^2} - f'\left(\frac{y_l}{w_h}\right) \frac{y_l}{w_h^2} \right] \right) \left[\frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} \right]$$

As $h(L, e) = h(L + ke)$ and is concave, $L^r > L^p$ implies that for all $\delta \leq \delta_E$ we have

$$\left[\frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} \right] \geq 0 \quad \blacksquare$$

Proof of Proposition 7. Differentiate the lagrangian (??) and set $y_1 = y_2 = 0$ and note there is no longer an incentive constraint corresponding to λ_2 and that the incentive constraint corresponding to λ_3 becomes simpler. It is simple to show that the derivative is unambiguously negative. ■

Proof of Lemma 2. The result requires that human capital in both neighbourhoods be decreasing in δ at the policy δ_E . We can solve for

$$\frac{\partial L_1^p}{\partial \delta} = \frac{\bar{\theta} (h_p^2 \frac{\partial h_p}{\partial \delta} + h_r^2 \frac{\partial h_r}{\partial \delta})}{8(h_p + h_r)^2}$$

At the policy δ_E we have $h_p = h_r$ by definition so that this term is negative whenever $\frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} < 0$ which it is at δ_E . In fact as noted earlier, this is the impact on total human

capital, which is negative for any $\delta \geq \delta^*$.

Similarly, we can derive the change in the externality in the rich neighbourhood evaluated at the equal opportunity policy

$$\frac{\partial L_1^p(\delta_E)}{\partial \delta} = \frac{3\bar{\theta}}{8} \left(\frac{\partial h_p}{\partial \delta} + \frac{\partial h_r}{\partial \delta} \right)$$

which is of course negative as shown above. ■

Proof of Proposition 8.

Total welfare consists of welfare in periods 1 and 2 where $\hat{\delta}_1(\delta_0)$ is the optimal policy in period 1 that maximizes W_2^* .

$$W = \max_{\delta_0} \underbrace{\int_{\theta} ([\theta h(L_0^p, \delta_0)]^{\sigma} + [\theta h(L_0^r, 1 - \delta_0)]^{\sigma}) dF(\theta)}_{W_1(\delta_0)} + \underbrace{\int_{\theta} ([\theta h(L_1^p, \hat{\delta}_1)]^{\sigma} + [\theta h(L_1^r, 1 - \hat{\delta}_1)]^{\sigma}) dF(\theta)}_{W_2^*(\delta_0)} \quad (\text{A.7})$$

First, we note that both $W_1(\delta_0)$ and $W_2^*(\delta_0)$ are concave in δ_0 . We don't show this here as the latter requires significant manipulation, but it is, however, straightforward. Define the optimal "static" policy in period 0 (choice of δ_0) as $\hat{\delta}_0$, which optimizes W_1 and ignores the impact on the future generation (through the impact on neighbourhood composition).

Thus $\hat{\delta}_0$ is defined analogously to $\hat{\delta}_1$. Consider the impact of δ_0 on period 2 welfare

$$\frac{\partial W_2^*}{\partial \delta_0} = \frac{\partial L_1^p}{\partial \delta_0} \left(h(L_1^p, \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^p, \hat{\delta}_1)}{\partial L} \right) + \frac{\partial L_1^r}{\partial \delta_0} \left(h(L_1^r, 1 - \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^r, 1 - \hat{\delta}_1)}{\partial L} \right)$$

The first (second) term is the impact of a change in δ_0 period 2 welfare in the poor (rich) region. This effect is the result of education policy on the composition of each neighbourhood.

By definition, $L_1 = L_1^p + L_1^r$ so that $\frac{\partial L_1}{\partial \delta_0} = \frac{\partial L_1^p}{\partial \delta_0} + \frac{\partial L_1^r}{\partial \delta_0}$. Use this to substitute for L_1^r and rewrite as

$$\frac{\partial W_2^*}{\partial \delta_0} = \frac{\partial L_1^p}{\partial \delta_0} \left[h(L_1^p, \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^p, \hat{\delta}_1)}{\partial L} - h(L_1^r, 1 - \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^r, 1 - \hat{\delta}_1)}{\partial L} \right] + \frac{\partial L_1}{\partial \delta_0} \left[h(L_1^r, 1 - \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^r, 1 - \hat{\delta}_1)}{\partial L} \right]$$

As $\frac{\partial L_1}{\partial \delta_0} \leq 0$ for any $\delta \geq \delta^*$, the term on the right-hand side is less than or equal to zero. Further, the term in brackets on the left-hand side of the above expression is positive by assumption 1. Thus a sufficient condition for part 2 of the proposition to hold is $\frac{\partial L_1^p(\hat{\delta}_0)}{\partial \delta_0} \leq 0$. Thus, whenever the impact on the poor neighbourhood is negative, the static optimal δ is too large. This is quite weak in that even when this term is positive it may still be outweighed by the negative impact on the rich.

Consider the following

$$\frac{\partial L_1^p}{\partial \delta} = \frac{\bar{\theta}(h_p^2 \frac{\partial h_p}{\partial \delta} + h_r^2 \frac{\partial h_r}{\partial \delta})}{8(h_p + h_r)^2}$$

so that $\frac{\partial L_1^p}{\partial \delta}$ is negative when

$$\left(\frac{h_r}{h_p} \right)^2 < \frac{\frac{\partial h_r}{\partial \delta}}{\frac{\partial h_p}{\partial \delta}}$$

The first-order condition from the static problem (which determines $\hat{\delta}_0$) can be written

$$\left(\frac{h_p}{h_r} \right)^{\sigma-1} = \frac{\frac{\partial h_r}{\partial \delta}}{\frac{\partial h_p}{\partial \delta}}$$

The condition is

$$\left(\frac{h_p}{h_r} \right)^{\sigma-1} > \left(\frac{h_r}{h_p} \right)^2$$

which is true whenever $\sigma < -1$ or $\rho > 2$. This proves the second part of the proposition. To show part 1, note that part 2 implies that the left-hand side of (??) is positive. Furthermore, $\frac{\partial L_1}{\partial \delta_0} = 0$ when $\rho = 0$ from the first-order conditions on the static problem. Thus $\frac{\partial W_2^*}{\partial \delta_0}$ is positive when $\rho = 0$. ■

Proof of Corollary 2. As above, the impact on welfare of a change in δ_0 is

$$\begin{aligned} \frac{\partial W_2^*}{\partial \delta_0} = \frac{\partial L_1^p}{\partial \delta_0} \left[h(L_1^p, \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^p, \hat{\delta}_1)}{\partial L} - h(L_1^r, 1 - \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^r, 1 - \hat{\delta}_1)}{\partial L} \right] \\ + \frac{\partial L_1}{\partial \delta_0} \left[h(L_1^r, 1 - \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^r, 1 - \hat{\delta}_1)}{\partial L} \right] \end{aligned}$$

Substituting $\frac{\partial h}{\partial L} = \frac{\partial h}{\partial e} K$, the first term on the left-hand side becomes

$$\frac{\partial L_1^p}{\partial \delta_0} \left[h(L_1^p, \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^p, \hat{\delta}_1)}{\partial \delta_1} - h(L_1^r, 1 - \hat{\delta}_1)^{\sigma-1} \frac{\partial h(L_1^r, 1 - \hat{\delta}_1)}{\partial \delta_1} \right] K$$

which equals 0 by the f.o.c that defines $\hat{\delta}_1$. The sign of the second term depends solely on $\frac{\partial L_1}{\partial \delta_0}$. The optimal policy increases δ_0 until the benefit from doing so equals the cost (to the poor neighbourhood). The f.o.c determining $\hat{\delta}_0$ can be written (using $\frac{\partial h}{\partial L} = \frac{\partial h}{\partial e} K$)

$$\left(\frac{h_p}{h_r} \right)^{\sigma-1} = \frac{\partial h_r}{\partial L} \frac{\partial L}{\partial h_p}$$

Further, for any $\delta \in [0, \delta_E]$ we know that

$$1 \leq \left(\frac{h_p}{h_r} \right)^{\sigma-1} = \frac{\partial h_r}{\partial L} \frac{\partial L}{\partial h_p} \leq 1$$

where the second inequality comes from the fact that $h(L, e)$ is concave in L and $L^r \geq L^p$.

Thus $h_p = h_r$ and the optimal policy is δ_E . Also, we have $\frac{\partial h_r}{\partial L} = \frac{\partial h_p}{\partial L}$ which implies that $\frac{\partial L_1}{\partial \delta_0} = 0$ ■

Appendix B

Appendix for Chapter 4

B.1 Screening and the reduced form

The qualitative conclusions from the basic screening model (BSM) can be summarized as follows:

- firms make no profits due to competition and individuals capture all the surplus
- overall welfare reduced because of wasteful education acquisition by high types trying to differentiate themselves from low types (socially wasteful)
- low type's ability to mimic and competition on the firm side guarantees him his first-best outcome
- if there are not enough high types, an equilibrium may not exist

We can model costly screening (CSM) in a reduced form manner as follows:

- firms face a fixed cost of screening an applicant: cost is socially wasteful so is akin to high type's education investment
- because firms compete and already make zero profits, screening cost is fully reflected in reduced wages
- if the returns from screening are not sufficiently large (i.e., screening costs are too high), then screening is not worth it and is not carried out in equilibrium

The basic screening model (BSM) can be mapped into the costly (CSM) model using the following parameters: $k_H = r(\theta_H, e^{SB})$, and $k_L = 0$.

B.2 Proofs

Proof of Lemma 5. We can rewrite expression (4.3) as follows:

$$(p_{HH} + p_M)[r(\theta_H, e) - r(\theta_H, e_g)] > p_M[r(\theta_H, e_g) - r(\theta_L, e)] \quad (\text{B.1})$$

if $p_{HH} + p_M + p_M = 1$ (which is assumed here for ease of exposition). Assume that (B.1) is strict. The terms in the square brackets have the same signs so that $r(\theta_H, e) > r(\theta_H, e_g)$ and $r(\theta_H, e_g) > r(\theta_L, e)$. Now, as $r(\theta, e)$ is increasing in e , the LHS is decreasing in e_g while the RHS is increasing in e_g . Hence, for some e_g , equality in (B.1) is attained. An analogous argument holds if the inequality in (B.1) is reversed. Moreover, without uncertainty, welfare in the two systems is equal only when $e_g = e$. With uncertainty, the welfare under specialized decreases while the welfare under general remains the same. Hence, equality will hold for some $e_g < e$. ■

Proof of Proposition 9. A specialized system dominates a general one whenever the distortion

$$(p_{HH} + p_M)r(\theta_H, e) + p_M r(\theta_L, e) > r(\theta_H, e_g) + (p_{HH} + p_M)\kappa \quad (\text{B.2})$$

Then comparing (4.3) and (B.2) it is clear that for a fixed e equality in (B.2) holds for $e_g < e_g^*$ as $\kappa > 0$. Now, the RHS of (B.2) is increasing in e_g while the LHS is decreasing in e_g . Therefore, at $e_g = e_g^*$,

$$(p_{HH} + p_M)r(\theta_H, e) + p_M r(\theta_L, e) < r(\theta_H, e_g^*) + (p_{HH} + p_M)\kappa$$

so that adverse selection favours a general system. ■

Proof of Proposition 10. The result requires that

$$(p_{HH} + p_{LH} + p_{HL})r_g \Delta_g < \left[p_{HH} + \frac{1}{2}(p_{LH} + p_{HL}) \right] r_h \Delta_h + \frac{1}{2}(p_{LH} + p_{HL})r_l \Delta_l. \quad (\text{B.3})$$

A sufficient condition for this to hold is (by lemma 6)

$$(p_{HH} + p_{LH} + p_{HL})r_g \Delta_g < \left[p_{HH} + \frac{1}{2}(p_{LH} + p_{HL}) \right] r_h \Delta_h + \frac{1}{2}(p_{LH} + p_{HL})r_l \Delta_l. \quad (\text{B.4})$$

Factoring Δ_h and rewriting gives

$$(p_{HH} + p_{LH} + p_{HL})r_g \Delta_g < \left(\left[p_{HH} + \frac{1}{2}(p_{LH} + p_{HL}) \right] r_h + \frac{1}{2}(p_{LH} + p_{HL})r_l \right) \Delta_h. \quad (\text{B.5})$$

This is true when Lemma 6 and (4.14) hold. ■

Appendix C

Appendix for Chapter 5

C.1 Proofs

Proof Lemma 7.

Consider first $\frac{\partial b_1^*}{\partial d} > 0$, $\frac{\partial \tau_1^*}{\partial d} > 0$. The first-order conditions for the unconstrained problem are

$$h'(\delta(\tau_1)y_1 + b_1 - d)\delta'(\tau_1) = 1 \quad (\text{C.1})$$

$$h'(\delta(\tau_1)y_1 + b_1 - d) + \beta(1+r)h'(\delta(\tau_2)y_2 - (1+r)b_1) = 0 \quad (\text{C.2})$$

The implicit function theorem assures us of the existence of the functions $\tau_1^*(d)$ and $b_1^*(d)$.

Differentiating the F.O.C's fully and separating the relevant partials provides the results.

These two derivatives are given by

$$\begin{aligned} \frac{\partial \tau_1}{\partial d} \underbrace{(h''(\delta(\tau_1)y_1 + b_1 - d)[\delta'(\tau_1)]^2 y_1 + h'(\delta(\tau_1)y_1 + b_1 - d)\delta''(\tau_1))}_{a} + \frac{\partial b_1}{\partial d} \underbrace{h''(\delta(\tau_1)y_1 + b_1 - d)\delta'(\tau_1)}_{b} \\ = \underbrace{h''(\delta(\tau_1)y_1 + b_1 - d)\delta'(\tau_1)}_{c} \end{aligned}$$

$$\frac{\partial \tau_1}{\partial d} \underbrace{(\delta'(\tau_1)y_1)}_{d} + \frac{\partial b_1}{\partial d} \underbrace{(2 + r)}_{e} = 1$$

We have two equations and two unknowns. Using the labels in the underbraces and rewriting in matrix form we have the following.

$$\begin{bmatrix} a & b \\ d & e \end{bmatrix} \begin{bmatrix} \frac{\partial \tau_1}{\partial d} \\ \frac{\partial b_1}{\partial d} \end{bmatrix} = \begin{bmatrix} c \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{\partial \tau_1}{\partial d} \\ \frac{\partial b_1}{\partial d} \end{bmatrix} = \frac{1}{ae - bd} \begin{bmatrix} ec - b \\ -dc + a \end{bmatrix}$$

To show the result, first calculate the following

$$ae - bd = (1 + r)(h''(\delta(\tau_1)y_1 + b_1 - d)[\delta'(\tau_1)]^2 y_1 + (2 + r)(h'(\delta(\tau_1)y_1 + b_1 - d)\delta''(\tau_1)) < 0 \quad (\text{C.3})$$

$$ec - b = (1 + r)(h''(\delta(\tau_1)y_1 + b_1 - d)\delta'(\tau_1)) < 0 \quad (\text{C.4})$$

$$-dc + a = h'(\delta(\tau_1)y_1 + b_1 - d)\delta''(\tau_1) < 0 \quad (\text{C.5})$$

Given these we have

$$\frac{\partial \tau_1}{\partial d} = \left(\frac{1}{ae - bd} \right) (ec - b) > 0 \quad (\text{C.6})$$

$$\frac{\partial b_1}{\partial d} = \left(\frac{1}{ae - bd} \right) (-dc + a) > 0 \quad (\text{C.7})$$

The other two results are obtained in the same manner. ■

Proof Proposition 11.

It is tedious but straightforward to show that $\frac{\partial^2 V(w_0)}{\partial \bar{b}^2} < 0$ so that a stationary point is a unique maximum. Replacing $\bar{w} = y_2 - (\bar{b} - L)(2 + r)$ and $\underline{w} = y_2 - \bar{b}(2 + r)$, we can express the impact on welfare of a change in \bar{b} as follows (using both the envelope theorem and Leibniz integral rule).

$$\frac{\partial V(w_0)}{\partial \bar{b}} = \beta \underbrace{\int_{\underline{y}}^{y_2 + (1+r)b_0^* - \bar{b}(2+r)} [h'(y_1 - (1+r)b_0^* + \bar{b}) - h'(y_2 - (1+r)\bar{b})] f(y_1) dy_1}_{\geq 0, \text{ effect of easing debt constraint when borrowing is constrained}} \quad (\text{C.8})$$

$$\beta \underbrace{\int_{y_2 + (1+r)b_0^* - (\bar{b}-L)(2+r)}^{\bar{y}} [h'(y_1 - (1+r)b_0^* + \bar{b} - L) - h'(y_2 - (1+r)(\bar{b} - L))] f(y_1) dy_1}_{\leq 0, \text{ increases in the borrowing constraint give the politician more leverage}}$$

As the debt limit becomes larger (smaller), the positive (negative) terms shrink and eventually approach zero as there is no positive probability of hitting the debt (political) constraint. Specifically,

$$\begin{aligned} \lim_{\bar{b} \rightarrow y_2} y_2 + (1+r)b_0^* - \bar{b}(2+r) &\leq \lim_{\bar{b} \rightarrow y_2} (y_2 - \bar{b}) = 0 < \underline{y} \\ \Rightarrow \lim_{\bar{b} \rightarrow y_2} \int_{\underline{y}}^{y_2 + (1+r)b_0^* - \bar{b}(2+r)} f(y_1) dy_1 &= 0 \Rightarrow \frac{\partial V(w_0)}{\partial \bar{b}} < 0 \end{aligned}$$

Similarly,

$$\begin{aligned} \lim_{\bar{b} \rightarrow y_2 + L - \bar{y}} y_2 + (1+r)b_0^* - (\bar{b} - L)(2+r) &\leq \lim_{\bar{b} \rightarrow y_2 + L - \bar{y}} y_2 + L - \bar{b} = \bar{y} \\ \Rightarrow \lim_{\bar{b} \rightarrow y_2 + L - \bar{y}} \int_{y_2 + (1+r)b_0^* - (\bar{b} - L)(2+r)}^{\bar{y}} f(y_1) dy_1 &= 0 \Rightarrow \frac{\partial V(w_0)}{\partial \bar{b}} > 0 \end{aligned}$$

where the last inequalities in each come from the fact that only the relevant term in (C.8) is zero at these limits (which is easy to show). Since $V(\bar{b})$ is continuous, there is some point where the derivative is zero.

To show that the optimal debt limit is decreasing in initial wealth, it is sufficient to show that $\frac{\partial^2 V}{\partial \bar{b} \partial w_0} < 0$. Again using Leibniz rule we have the following.

$$\frac{\partial V^2}{\partial \bar{b} \partial w_0} = \beta \int_{\underline{y}}^{y_2 + (1+r)b_0^* - \bar{b}(2+r)} h''(y_1 - (1+r)b_0^* + \bar{b}) \left[-(1+r) \frac{\partial b_0}{\partial w_0} \right] f(y_1) dy_1 \quad \text{†(C.9)}$$

$$\beta \int_{y_2 + (1+r)b_0^* - (\bar{b} - L)(2+r)}^{\bar{y}} h''(y_1 - (1+r)b_0^* + \bar{b} - L) \left[-(1+r) \frac{\partial b_0}{\partial w_0} \right] f(y_1) dy_1$$

which is always less than zero when period 0 borrowing is not increasing in wealth or

$$\frac{\partial b_0}{\partial w_0} \leq 0. \quad \blacksquare$$