

On the Efficiency Costs of De-
tracking Secondary
Schools

Kenn Ariga
Giorgio Brunello
Roki Iwahashi
Lorenzo Rocco

No.35, December 2006

www.carloalberto.org

On the Efficiency Costs of De-tracking Secondary Schools¹

Kenn Ariga², Giorgio Brunello³, Roki Iwahashi⁴ and Lorenzo Rocco⁵

December 2006⁶

¹We are grateful to Giuseppe Bertola and Pietro Garibaldi and to the audience at a seminar at Collegio Carlo Alberto, Moncalieri, for comments and suggestions. Part of this draft was written when the second author was visiting Collegio Carlo Alberto, which provided excellent hospitality. The usual disclaimer applies.

²Kyoto University

³Padova University, Kyoto University, CESifo and IZA

⁴University of the Ryukyus, Okinawa

⁵Padova University

⁶ c Any opinions expressed here are those of the authors and not those of the Collegio Carlo Alberto.

Abstract

During the postwar period, many countries have de-tracked their secondary schools, based on the view that early tracking was unfair. What are the efficiency costs, if any, of de-tracking schools? To answer this question, we develop a two skills - two jobs model with a frictional labour market, where new school graduates need to actively search for their best match. We compute optimal tracking length and the output gain/loss associated to the gap between actual and optimal tracking length. Using a sample of 18 countries, we find that: a) actual tracking length is often longer than optimal, which might call for some efficient de-tracking; b) the output loss of having a tracking length longer or shorter than optimal is sizeable, and close to 2 percent of total net output.

- JEL code: I2, J6
- Key words: mismatch, school tracking

1 Introduction

In most education systems in the developed world, heterogeneous pupils are initially mixed in comprehensive schools - typically primary and lower secondary education. As some stage of the curriculum, however, some form of (self) selection takes place, typically based on ability and past performance, and students are allocated to schools which specialize in different curricula (tracks) or to classes where subjects are taught at a different level of difficulty (streams). The former system is typical of Central European countries, such as Germany, Austria, The Netherlands and Hungary, but exists also in Korea and Japan, and the latter system is typically observed in the US. When no selection whatsoever occurs during upper secondary school, as in some Scandinavian countries, choice and specialization are delayed until college education.

For the sake of brevity, we shall call the allocation of students into different schools or classes as "tracking". School tracking was rather widespread both in Europe and in the US after the Second World War. While Germany is considered as the starkest example of early tracking, with pupils allocated to tracks at age 10, this was not so different in England or Italy after the War, where students were tracked at 11 (Manning and Pischke, 2006, Brunello and Checchi, 2006). In the US, this practice began in response to the influx of immigrant children during the early 20th century. By mid-century, a majority of secondary schools in that country used some form of tracking (Hallinan, 2006).

The postwar education system changed rather abruptly in a number of countries. In Europe, the so called "comprehensive school reform" led to de-tracking or to a substantial delay in the tracking time, starting with Sweden, the UK, Italy, and Norway in the 1960s, and continuing with Finland in the 1970s, France in the 1980s and Spain and Portugal in the early 1990s. Even the countries which have chosen to retain a system with early stratification have taken steps to delay specialization: Germany has introduced both a common orientation phase in each of the three tracks and a system of comprehensive schools which are rather widespread in some Lander, and Belgium and the Netherlands have introduced a common curriculum in the first stages of secondary education¹. In Japan, many vocational high schools have changed their curricula in the 1980s and added more academic courses, with a view of facilitating access to college education. In the US, the so-called "de-tracking movement" began in the 1980s and led to court mandated de-tracking reforms in a number

¹See Green, Wolf and Leney, 1999, and Meghir and Palme, 2004, for a detailed discussion.

of US school districts and states².

Broadly speaking, the de-tracking movement was driven by the perception, among educators and politicians, that early school tracking systems were unfair. It was felt that privileged children were much more likely to end up in the more demanding academic tracks, and then in college, at the expense of working class children, who typically ended up in less demanding and dead-end vocational tracks. Critics of early tracking systems argued that high-performing students gained at the expense of their lower-performance peers. In a well known contribution, Oakes, 1992, concluded that tracking was a way of segregating minority students to lower tracks, and characterized tracking as an elitist practice, that perpetuated the status quo by giving students from privileged families access to elite colleges and high-income careers. This view has been recently refined by Bertocchi and Spagat, 2003, who model school design as the outcome of class struggle.

While well placed, the concern with equality of opportunity prompts the following question: is there a trade-off between equality of opportunity and efficiency, and if yes, what are the efficiency costs of de-tracking secondary schools? It is well known that most educators find that tracking facilitates instruction by making it easier to gear lessons to the ability level of the whole class. Higher effectiveness implies that a more demanding curriculum can be taught, without having to "teach to the middle". When pupils have both academic and practical or vocational talents, the separation of schools into classes which specialize in either talent is likely to generate returns from specialization. Since tracking is often done by (measured) ability, selection generates peer effects. These effects are positive for the high - ability group, because the less able are separated out, and negative for the low ability - group, because the more able are placed in different classes/schools. If peer effects are non-linear, so that the gain for the high ability group more than compensates the loss for the low ability group, then tracking increases efficiency (Minter Hoxby, 2000). However, as remarked by Manning and Pischke, (2006), "...we know very little about the different impact of peer group effects on different types of students empirically, so it is difficult to judge a priori whether this leads to lower or higher average performance in a selective system" (p.6).

It seems natural to argue that there are efficiency costs in the de-tracking of secondary schools if average school performance is higher in tracked systems. Hanushek and Woess-

²We are aware that stratification occurs also between private and public schools, and tracking is only one dimension of the problem. See Card and Rothstein, 2006.

mann, 2006, use the standardized cognitive test scores of 15 years old pupils in a large number of developed countries and show that, on average, performance in these tests is not affected by the extent of tracking. Is this the end of the story? We believe there are at least three reasons why the appropriate answer is no. First, one can argue that looking at test scores at 15 is too early, as tracking in most countries does not start before 15, if ever (see Brunello and Checchi, 2006). Second, it might be reductive to focus only on cognitive skills, as vocational schools in tracked systems develop to a higher extent more practical and vocational skills. Third, and most important for the purpose of this paper, an important test of school quality that goes beyond standardized test scores while at school is earnings and employment prospects after labour market entry (Card and Krueger, 1990). We conclude that an evaluation of the efficiency costs of school tracking cannot ignore the transition from school to work.

In this paper we show that the finding that average school performance - measured by test scores in cognitive skills - is not affected by the extent of school tracking does not imply that optimal tracking time should be zero: quite the contrary, in a competitive labour market with (ex-post) heterogeneous workers and jobs, optimal tracking time should be at its maximum possible value. We also show that the presence of frictions in the labour market can induce a central planner which maximizes total net output by choosing the appropriate education policy to delay tracking with respect to the benchmark frictionless economy, even in the absence of any concern for equality of opportunity. By implication, we argue that the evaluation of the efficiency costs of de-tracking schools should take into proper account the presence of labour market frictions.

We have shown in previous papers (Brunello, Giannini and Ariga, 2004; Ariga, Brunello, Iwahashi and Rocco, 2005) that complete tracking is not efficient when the allocation of students to tracks is noisy, so that the more talented students do not necessarily end up in the academic track. In these papers we have argued that the size of the noise is likely to be higher the earlier selection takes place, because at the early stages of a pupil's life it is easier to confound ability with maturity. This argument emphasizes misallocation at school but ignores mismatch in the labour market.

In this paper we focus instead on the school to work transition. While there is no misallocation at school, the labour market is not perfectly competitive, but is characterized by frictions, and by the fact that individuals need to search for the best job match and firms can post wages to attract applications. In this market, it takes time and other resources for a

worker to land a good job, and for a firm to fill a vacancy. As surveyed by Rogerson, Shimer and Wright, 2005, there are different ways of characterizing such a market, depending on the assumptions one makes on the search process and on wage determination. We focus here on a directed search model, where firms can use their wage policies to attract applicants and workers can place their job application in a non-random way at a positive cost.

We develop a two-period model of schooling and matching, and contrast the optimal tracking time which would emerge in a competitive labour market with optimal tracking in a frictional labour market where search is directed and firms post wages. Using numerical examples, we show that the latter is shorter than the former. Therefore, a frictional labour market could induce an efficient reduction of tracking relative to the first best allocation, even when no misallocation occurs during the schooling period and the government does not care about equality of opportunity. We also compute for 18 countries the output loss generated by the observed school design, which typically deviates from the optimal design, and find that this loss is close to 2 percent of total net output. Given that most developed countries spend about 4-5 percent of their GNP on schooling, this is a sizeable number. While on average actual tracking length is much shorter than optimal length, there is substantial cross country heterogeneity, with about half of the countries in our sample having "too long" tracking, and the other half experiencing the opposite. Less ambiguously, we find that the actual share of students enrolled in the academic track is always below the optimal value.

The paper is organized as follows. Section 2 describes the model setup and characterizes schooling. Section 3 introduces the labour market, and Section 4 describes the properties of the symmetric equilibrium. Section 5 compares decentralized matching with the social optimum and discusses policies - such as taxes and subsidies - which could help the government to bridge the gap between the two. Finally, Section 6 presents the numerical solutions based on the more general model. Conclusions follow.

2 Set-up

Since we are interested mainly in the school to work transition, we consider a two period model without discounting. In the initial period, individuals are young and go to school³. There are no separations and the match lasts until the economy finishes at the end of the second period. In the second period, school graduates enter in the labour market, which consists of an endogenous number of job slots (firms). If a satisfactory match occurs, production takes place and wages are paid. If no match happens, the vacant job slot remains empty, and school graduates end up in a reservation sector, that pays a given level of income, which we normalize to zero⁴.

School consists of a sub-period of comprehensive education, of endogenous length $1 - \tau$, and of a sub-period of selective education, with individuals allocated to two tracks, A (academic) and V (vocational). The academic track specializes in the production of academic skills, and the vocational track specializes in the development of practical skills. Let η be the share of pupils in the A track. While in principle the schooling period includes all schooling from primary school to college, it is best to restrict our attention to schooling until upper secondary education.

In real life, the allocation of students to track is the result of the interaction between parental decisions, individual decisions and allocation mechanisms decided by the central or local government, such as admission criteria⁵. In this model, we drastically simplify the allocation process by assuming that the government determines ex-ante the number of slots available in each school. We also assume that individuals are homogeneous, and that the allocation of pupils to schools is random⁶. While these are strong simplifications, they allow us to focus on the school to work transition rather than on the allocation process, which is much more relevant when equality of opportunity is the question of interest (see Brunello and Checchi, 2006).

The production technology consists of two jobs, G (general) and T (technical), which differ in the importance played by academic and vocational skills. Firms choose which job to

³A dynamic version of the model is under preparation.

⁴Alternatively we could have a wage constraint. Suppose that workers can always find a job with a given wage in the secondary labour market. Then no wage offer below the given wage will be accepted. We can show that most of the analysis below carries through after adequate normalization.

⁵Individuals differ in their innate talent and the allocation process generates a non-random distribution of abilities in each type of school, which is the source of peer effects.

⁶In our model, who goes to which school is irrelevant.

set up at the beginning of the second period, after having observed the supply of graduates, and bear a fixed setup cost c . Search operates as follows: first, firms announce and commit to a wage level, which varies between jobs but not within jobs. Second, applicants observe the posted wage offers and decide their job application policy, which can deviate from random applications at a positive cost. Third, firms receive the applications and make offers. Offers are accepted⁷, wages are paid and production takes place. Unmatched individuals go to the secondary sector. On the other end, unmatched jobs and remain unfilled.

The timing of the model is as follows: the government decides the optimal composition of schools into A and V types and the optimal tracking time τ so as to maximize expected total net output. Individuals are allocated to school and complete their education in the first period⁸. Firms decide whether to enter in this market and the type of job they wish to set up. Search occurs in the manner sketched above. Finally, production takes place, wages are paid and the economy terminates.

2.1 Schooling

Individuals are born with two skills, a general and a vocational skill. While comprehensive schools teach both skills at the same pace, tracked schools specialize in one skill, which they teach more effectively than comprehensive schools. By focusing on a single skill, each track does not develop the other skill, which suffers with respect to the comprehensive benchmark. In a world with heterogeneous individuals, the higher effectiveness of tracked schools could be generated by the lower variance of abilities in the track, or by the presence of peer effects. In the current setup, we entirely by-pass these thorny issues by assuming that individuals are ex-ante homogeneous.

The technology of production of human capital in schools is given, and depends on whether the system is comprehensive or stratified. We assume that, when schools are comprehensive, the human capital accumulated by each pupil in each skill is equal to 1 at the end of school. With tracking, however, we distinguish whether the attended school is academic or vocational. In the former case, accumulated human capital in the academic

⁷As discussed below, offers are never rejected

⁸We assume that on the job training cannot completely undo the relative advantage that each type of graduate has for a job type.

skill is⁹

$$h_A^G = 1 + \sigma_A^G \tau \quad (1)$$

and in the vocational skill is¹⁰

$$h_A^T = 1 - \sigma_A^T \tau \quad (2)$$

where σ_i^j are given parameters. In reality, σ_i^j depend on school resources, that are controlled by the government. In our model, however, we treat these resources as given, and assume that the government can only choose the technology of production - either comprehensive or stratified - and the composition of pupils in each track when schooling is stratified. This is a simplified assumption, which sharpens our focus on the comparison of alternative school designs by setting school resources as given.

By spending a fraction of schooling time τ in the academic track, pupils improve their academic skills relative to a comprehensive school at the rate σ_A^G , at the price of having less developed vocational skills. The stock of these skills declines at the rate σ_A^T with the time spent in an academic track. The opposite occurs for pupils enrolled in vocational tracks. In this case we have respectively

$$h_V^T = 1 + \sigma_V^T \tau \quad (3)$$

and

$$h_V^G = 1 - \sigma_V^G \tau \quad (4)$$

A special but interesting case that we shall consider at length in the paper is $\sigma_i^j = \sigma$. This case has the convenient property that average individual human capital is constant (and equal to 1) across comprehensive and tracked schools.

This property implies that, on average, each individual does not gain in terms on human capital from a higher degree of tracking, consistently with the empirical results of Hanushek

⁹With $(1 - \tau)$ of the time spent in a comprehensive school, accumulated human capital is $h_A^G = (1 - \tau) + \zeta_A^G \tau$, where $\zeta_A^G > 1$, which we re-write as $h_A^G = 1 + (\zeta_A^G - 1)\tau = 1 + \sigma_A^G \tau$

¹⁰Accumulated human capital is $h_A^V = (1 - \tau) + \zeta_A^V \tau$, where $\zeta_A^V < 1$, which we re-write as $h_A^V = 1 - (1 - \zeta_A^V)\tau = 1 - \sigma_A^V \tau$

and Woessmann, 2006¹¹. Moreover, when pupils are uniformly distributed between tracks, average academic skills do not depend on tracking length, again consistent with Hanushek and Woessmann, 2006.

2.2 Jobs

The production technology is linear, with one unit of output produced by one unit of human capital. Suppose that a job j matches with a graduate of school i . The resulting output flow is

$$y_i^j = h_i^j$$

for $i = A, V$ and $j = G, T$. The human capital functions specified above imply

$$\begin{aligned} y_A^G &> y_A^T \\ y_V^T &> y_V^G \end{aligned}$$

so that graduates of A schools are more productive at type G jobs, and graduates of V schools have a comparative advantage in T jobs. We also have that

$$\begin{aligned} y_A^G &> y_V^G \\ y_V^T &> y_V^G \end{aligned}$$

so that jobs of type G yield higher output when matched with A graduates, and jobs of type T produce more with V graduates.

3 Matching

Our characterization of job matching with two job types and two worker types follows previous work by Shi, 2002, and Shimer, 2005. Following Shi, 2002, we assume that firms having the same job type post the same wage for each worker type, and workers having the same school record select the same application probabilities. This assumption implies that all the heterogeneity we are concerned with is between jobs and school types. Shi justifies

¹¹Notice however that Hanushek and Woessmann consider only academic skills.

this with the presence of a large number of workers and jobs, and with the fact that it is difficult for agents to coordinate their decisions in a large market. This is equivalent to restrict our attention to "...symmetric, mixed strategy equilibria where ex-ante identical firms and workers use the same strategy..."¹² (p.470).

We start with some notation. At the end of the first period, there exists a unit mass of graduates, with a share η coming from A schools and the share $1 - \eta$ coming from V schools. The number of vacant job slots for each job type is v^G and v^T respectively. Let μ be the share of of type G vacancies on total vacancies v , so that $\mu = \frac{v^G}{v}$. Each worker has a single application at hand, and she can control the probability that such application lands on the preferred job slot¹³. Firms receive applications for their job slot, and select one - if any - for a job offer, which is certainly accepted.

3.1 The ball urn matching mechanism

Denote with p_i^j the probability that type i worker's application reaches a type j job slot. We impose

$$p_A^G + p_A^T = 1, \quad p_V^G + p_V^T = 1, \quad 0 \leq p_i^j \leq 1 \quad (5)$$

For simplicity, we also write $p_A = p_A^G$, $p_V = p_V^T$. Denote by D_i^j the total flow of applications received by type j job from type i workers. Then we have

$$\begin{aligned} D_A^G &= p_A \eta \\ D_A^T &= (1 - p_A) \eta \\ D_V^T &= p_V (1 - \eta) \\ D_V^G &= (1 - p_V) (1 - \eta) \end{aligned} \quad (6)$$

We define the queue length λ_i^j as the ratio of applications i to job slots j . More in detail

¹²Rogerson, Shimer and Wright, 2005, argue that in large markets the mixed strategy equilibrium is the natural outcome. This introduces a coordination friction, as more than one worker can apply for the same job.

¹³The assumption that each worker has only one application in hand is not only crucial for the tractability of the subsequent analysis, but also for its normative implications. If, as Albrecht and co-authors, 2003, do, we allow workers to send multiple applications, then they may reject some of the received offers, thereby imposing a negative externality on the economy.

$$\begin{aligned}\lambda_A^G &= \frac{p_A \eta}{v^G}, \lambda_V^G = \frac{(1 - p_V)(1 - \eta)}{v^G} \\ \lambda_V^T &= \frac{p_V(1 - \eta)}{v^T}, \lambda_A^T = \frac{(1 - p_A)\eta}{v^T}\end{aligned}\tag{7}$$

If workers do not spend any effort to control the probability p_j , applications are distributed randomly over available job slots so that $\bar{p}_A = \mu$, $\bar{p}_V = 1 - \mu$. Since λ_i^j is the average number of applications by type i workers per type j job slot, in a large market the probability that type j job slot does not receive any application from type i workers is given by $\exp(-\lambda_i^j)$.¹⁴ Hence

$$(\lambda_i^j) \equiv 1 - \exp(-\lambda_i^j), \quad '(\lambda_i^j) = \exp(-\lambda_i^j) = 1 - (\lambda_i^j)\tag{8}$$

is the probability that a vacancy of type j receives at least one applications from type i worker. There are four distinct cases to consider: a) a job slot receives applications from both types of workers, which occurs with probability $(\lambda_A^j)(\lambda_V^j)$; b) only from type A workers with probability $(\lambda_A^j)[1 - (\lambda_V^j)]$; c) only from type V workers with probability $(\lambda_V^j)[1 - (\lambda_A^j)]$; d) from no worker, with probability $[1 - (\lambda_A^j)][1 - (\lambda_V^j)]$. Since each worker can receive at best only one offer, offers are never rejected.

Define the net profit of a match as

$$\pi_i^j = y_i^j - w_i^j - c\tag{9}$$

A crucial aspect for the subsequent analysis is how firms rank worker types. Later in the paper we show that, if $y_i^j > y_i^k$, then $\pi_i^j > \pi_i^k$ is also true. This implies that firms with a job slot G receiving applications both from workers A and T strictly prefer to send their offer to type- A worker.

With

$$\phi(x) = \frac{1 - \exp(-x)}{x} = \frac{(x)}{x}$$

¹⁴We have

$$\sum_{N=1}^{\infty} \exp(-\lambda_i^j) \frac{\lambda_i^j}{N} = 1 - \exp(-\lambda_i^j)$$

the probabilities that an application generates an offer vary with the type of worker and job¹⁵, and are given by

$$\begin{aligned} r_A^G &= \phi(\lambda_A^G) ; r_V^G = \phi(\lambda_V^G) (1 - (\lambda_A^G)) \\ r_V^T &= \phi(\lambda_V^T) ; r_A^T = \phi(\lambda_A^T) (1 - (\lambda_V^T)) \end{aligned} \quad (10)$$

To illustrate, consider type A graduates. They are the preferred choice of type G job slots, but the second choice of type T slots. Therefore, the probability that an application sent by one of these graduates generates an offer from a type G job slot is equal to the probability than at least one application is received, $\phi(\lambda_A^G)$, divided by the number of applications per job slot λ_A^G , or to $\phi(\lambda_A^G)$. Next consider type V graduates. If they send their application to job slots G , they will receive an offer only conditional on no application from the other type of graduates. Since the latter event occurs with probability $(1 - (\lambda_A^G))$ and we assume that events are independent from each other, V graduates receive an offer with probability $\phi(\lambda_V^G) (1 - (\lambda_A^G))$.

3.2 Workers

Individuals are risk neutral and maximize their expected income, net of the cost of deviating from random matching. They search for jobs using mixed strategies, which consist of placing their applications to either type of job with a positive probability. We define the cost of deviating from random applications in such a way as to rule out corner solutions, that is, application strategies that concentrate on a single job. The expected utility is

$$U_A = p_A r_A^G w_A^G + (1 - p_V) r_A^V w_A^V - \delta(p_A, \mu) \quad (11)$$

where

$$\delta(p_A, \mu) \equiv - \log \left\{ \left(\frac{p_A}{\eta} \right)^{\mu} \left(\frac{1 - p_A}{1 - \eta} \right)^{\mu - 1} \right\} \geq 0$$

for type A graduates, and

$$U_V = p_V r_V^T w_V^T + (1 - p_V) r_V^G w_V^G - \delta(p_V, \mu) \quad (12)$$

¹⁵We also have

$$\phi'(x) = \frac{1}{x^2} [(1+x) \exp(-x) - 1] < 0 ; \lim_{x \rightarrow 0} \phi(x) = 1, \quad \lim_{x \rightarrow \infty} \phi(x) = 0$$

where

$$\delta(p_V, \mu) = \log \left\{ \left(\frac{p_V}{\eta} \right)^\mu \left(\frac{1 - p_V}{1 - \eta} \right)^{\mu - 1} \right\}$$

for type V graduates.

The optimal choice of application probabilities by workers solves the following first order conditions

$$r_A^G w_A^G - r_A^V w_A^V = \frac{\partial \delta(p_A, \mu)}{\partial p_A} = \frac{p_A - \mu}{p_A(1 - p_A)} = \rho(p_A, \mu) \quad (13)$$

where $\rho(x, \bar{x}) = \frac{x - \bar{x}}{x(1 - \bar{x})}$

$$r_V^T w_V^T - r_V^G w_V^G = \frac{\partial \delta(p_V, \mu)}{\partial p_V} = \frac{p_V - (1 - \mu)}{p_V(1 - p_V)} \quad (14)$$

The rather unusual specification of the costs δ of controlling applications in (11) and (12) rules out corner solutions for the probabilities p_A and p_V , and the possibility that the marginal search costs in (13) and (14) explode. Further discussion on this working hypothesis is relegated to Section 1 in the Appendix at the end of the paper.

Each type of worker chooses these probabilities so as to equate the marginal benefit to the marginal cost of the effort required¹⁶. It is useful to re-write these conditions as follows

$$\begin{aligned} E_A^G - E_A^V &= \frac{p_A - \mu}{p_A(1 - p_A)} \\ E_V^T - E_V^G &= \frac{p_V - (1 - \mu)}{p_V(1 - p_V)} \end{aligned} \quad (15)$$

where

$$\begin{aligned} E_A^G &= \frac{(\lambda_A^G)w_A^G}{\lambda_A^G}; E_A^T = \frac{(\lambda_A^T)(1 - \lambda_V^T)w_A^T}{\lambda_A^T} \\ E_V^T &= \frac{(\lambda_V^T)w_V^T}{\lambda_V^T}; E_V^G = \frac{(\lambda_V^G)(1 - \lambda_A^G)w_V^G}{\lambda_V^G} \end{aligned} \quad (16)$$

and E_i^j are the expected returns from applying to different jobs. When the cost of deviating from random matching can be disregarded ($\eta = 0$), as assumed in competitive search models, school graduates select their application policies so as to arbitrage away differences

¹⁶Notice that each individual worker takes the probability r_i^j of receiving an offer as given, since this probability is the result of the aggregate behavior of all labour market participants. Clearly, r_i^j is endogenous and determined at the equilibrium.

in expected returns. This opportunity is partially precluded by the presence of positive costs ($c > 0$), which grant firms a monopsonistic position in their wage posting policy. We shall see below that this has important implications on efficiency.

3.3 Firms

Firms with job slots post wages subject to the application strategy of workers (eq.(13)). Each firm knows that by altering the wage offered to each worker type, it can affect the arrival rate of job applications by both worker types. Job slot G yields expected profits equal to

$$\mathcal{L}_G = (\lambda_A^G)(y_A^G \square w_A^G) + (\lambda_V^G)(1 \square (\lambda_A^G))(y_V^G \square w_V^G) \square c \quad (17)$$

where c is the cost of setting up a job. Replacing the wages in (15) with (13) and using (14) allows us to re-write this expression as

$$\begin{aligned} \mathcal{L}_G = & (\lambda_A^G)y_A^G \square \lambda_A^G \left[E_A^T + \rho \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) \right] \\ & + (\lambda_V^G)(1 \square (\lambda_V^G))y_V^G \square \lambda_V^G \left[E_V^T + \rho \left(\frac{v(1 \square \mu)\lambda_V^T}{1 \square \eta}, 1 \square \mu \right) \right] \square c \end{aligned}$$

Importantly, each employer with a job slot G assumes that she cannot affect expected values E_A^T and E_V^T , which accrue to the other job, even though these values are endogenous to the model and determined at equilibrium (see Rogerson, Shimer and Wright, 2005, for a similar hypothesis). Each employer maximizes profits by choosing the queue length of each type of worker. The first order condition for λ_A^G is

$$\begin{aligned} & '(\lambda_A^G) \square y_A^G \square (\lambda_V^G)y_V^G \square E_A^T \\ = & \left[\rho \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) + \frac{v\mu\lambda_A^G}{\eta} \rho_1 \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) \right] \end{aligned} \quad (18)$$

where

$$\rho_1 \equiv \frac{\partial \rho(x, \bar{x})}{\partial x} = \frac{x^2 \square 2x\bar{x} + \bar{x}}{x^2(1 \square x)^2} > 0$$

The first order condition for λ_V^G is similar

$$\begin{aligned} & '(\lambda_V^G)(1 \square (\lambda_A^G))y_V^G \square E_V^T \\ & = \left[\rho \left(\frac{v\mu\lambda_V^G}{1 \square \eta}, \mu \right) + \frac{v\mu\lambda_V^G}{1 \square \eta} \rho_1 \left(\frac{v\mu\lambda_V^G}{1 \square \eta}, \mu \right) \right] \end{aligned} \quad (19)$$

The following two equations are the corresponding first order conditions for type T slots:

$$\begin{aligned} & '(\lambda_V^T) \square y_V^T \square (\lambda_A^T)y_A^T \square E_V^G \\ & = \left[\rho \left(\frac{v(1 \square \mu)\lambda_V^T}{1 \square \eta}, 1 \square \mu \right) + \frac{v(1 \square \mu)\lambda_V^T}{1 \square \eta} \rho_1 \left(\frac{v(1 \square \mu)\lambda_V^T}{1 \square \eta}, 1 \square \mu \right) \right] \end{aligned} \quad (20)$$

$$\begin{aligned} & '(\lambda_A^T)(1 \square (\lambda_V^T))y_A^T \square E_A^G \\ & = \left[\rho \left(\frac{v(1 \square \mu)\lambda_A^T}{\eta}, 1 \square \mu \right) + \frac{v(1 \square \mu)\lambda_A^T}{\eta} \rho_1 \left(\frac{v\mu\lambda_A^T}{\eta}, 1 \square \mu \right) \right] \end{aligned} \quad (21)$$

So far, we have taken it for granted that firms with a type G job slot prefer to hire a type A graduate. Actually, this is not obvious. Although output in the match of a G job with a A graduate is higher than in a match with a V graduate, the firm hiring a G graduate may have to pay a relatively high wage to attract such worker, and end up with a lower net profit. Fortunately, the lemma below shows that this event does not occur.

Lemma 1 *When firms post wages as in eqs. (16)-(19), $y_i^j \square w_i^j$ is strictly increasing in y_i^j . Hence $\text{sign}(\pi_i^j \square \pi_{i'}^j) = \text{sign}(y_i^j \square y_{i'}^j)$ for $i \neq i'$.*

Proof. See Section 2 in the Appendix. ■

Section 3 in the Appendix shows that conditions (16)-(19) can be combined to yield the following two equations

$$'(\lambda_A^G) \square y_A^G \square y_V^G + \sum = \square^G \quad (22)$$

where

$$\begin{aligned} \sum & \equiv 1 \square (\lambda_A^G) \square 1 \square (\lambda_V^G)) y_V^G \square (1 \square (\lambda_V^T))(1 \square (\lambda_A^T))y_A^T \\ & \square^G = \left[\frac{v\mu\lambda_A^G}{\eta} \rho_1 \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) \right] \\ & \square \left[\rho \left(\frac{\eta \square v\mu\lambda_A^G}{\eta}, 1 \square \mu \right) + \frac{\eta \square v\mu\lambda_A^G}{\eta} \rho_1 \left(\frac{\eta \square v\mu\lambda_A^G}{\eta}, 1 \square \mu \right) \right] \end{aligned}$$

and

$$v(\lambda_V^T) \left[y_V^T - y_A^T \right] \sum = \square^T \quad (23)$$

where

$$\square^T = \left[\frac{v(1 - \mu)\lambda_V^T}{1 - \eta} \rho_1 \left(\frac{v(1 - \mu)\lambda_V^T}{1 - \eta}, 1 - \mu \right) \right] \\ \square \left[\rho \left(1 - \frac{v(1 - \mu)\lambda_V^T}{1 - \eta}, \mu \right) + \left(1 - \frac{v(1 - \mu)\lambda_V^T}{1 - \eta} \right) \rho_1 \left(1 - \frac{v(1 - \mu)\lambda_V^T}{1 - \eta}, \mu \right) \right]$$

3.4 Zero Profit Conditions

While firms have some monopsonistic power in the labour market, there are no barriers to entry in the product market, and new firms and job slots are driven to enter or exit the market by the presence of expected positive or negative profits. To tie down the number of vacancies, we need the following zero profit condition for job G

$$\pi_G = (\lambda_A^G)(y_A^G - w_A^G) + (\lambda_V^G)(1 - (\lambda_A^G))(y_V^G - w_V^G) - c = 0$$

which can be re-written after some manipulations as

$$\pi_G = y_A^G - g \left[\lambda_A^G \right] (y_A^G - y_V^G) - g \left[\lambda_A^G + \lambda_V^G \right] y_V^G \\ + \frac{v\mu(\lambda_A^G)^2}{\eta} \rho_1 \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) + \frac{v\mu(\lambda_V^G)^2}{1 - \eta} \rho_1 \left(\frac{v\mu\lambda_V^G}{1 - \eta}, \mu \right) = c \quad (24)$$

where $g(x) = (1 + x) \exp(-x)$. Similarly for the type T job slot

$$\pi_T = y_V^T - g \left[\lambda_V^T \right] (y_V^T - y_A^T) - g \left[\lambda_V^T + \lambda_A^T \right] y_A^T \\ + \frac{v(1 - \mu)(\lambda_V^T)^2}{1 - \eta} \rho_1 \left(\frac{v(1 - \mu)\lambda_V^T}{1 - \eta}, 1 - \mu \right) \\ + \frac{v(1 - \mu)(\lambda_A^T)^2}{\eta} \rho_1 \left(\frac{v(1 - \mu)\lambda_A^T}{\eta}, 1 - \mu \right) = c \quad (25)$$

Finally, the total number of applications must be equal to the total number of applicants, under our maintained assumption that each graduate has a single application in hand. Therefore

$$\lambda_A^G v \mu + \lambda_A^T v (1 - \mu) = \eta \quad (26)$$

$$\lambda_V^T v (1 - \mu) + \lambda_V^G v \mu = 1 - \eta \quad (27)$$

The system of equations (22)-(27) determine the six endogenous variables of the model, λ_A^G , λ_A^T , λ_V^G , λ_V^T , v and μ . This completes the description of the model.

4 The Symmetric Equilibrium

In order to be able to provide an analytical treatment of the equilibrium, we simplify the model by imposing that $\sigma_i^j = \sigma$. This is equivalent to assuming

$$y_A^G = y_V^T = y_1 \quad \text{and} \quad y_V^G = y_A^T = y_2$$

These restrictions imply that the output gain from a more productive match relative to a less productive match is the same, independently of the type of job and skill. It turns out that the equilibrium is symmetric and

$$\begin{aligned} \eta &= \frac{1}{2} & \mu &= \frac{1}{2} \\ \lambda_A^G &= \lambda_V^T = \lambda_1 \\ \lambda_V^G &= \lambda_A^T = \lambda_2 \end{aligned} \quad (28)$$

This equilibrium is symmetric in the sense that the two jobs, G and T , specialize to the same extent in the use of one of the two available skills. Therefore, application rates differ not across jobs but only between the most preferred and the less preferred match. Vacancy creation is such that half of the new vacancies is in job G and the other half is in job T . Since η is a policy parameter, this equilibrium requires that the government, which designs the school system, sets the number of slots in each track so as to equalize the composition of graduates from each track with the composition of vacancies. Therefore, $\eta = \frac{1}{2}$.

4.1 Existence and Uniqueness of the Symmetric Equilibrium

Under symmetry, conditions (22) and (23) are equivalent, and eq. (22) can be written as

$$v(\lambda_1)(y_1 - y_2) + \sum = \pi^G$$

where $\sum = 0$ and

$$\begin{aligned} \pi^G &\equiv \pi\left(\lambda_1, v, \frac{1}{2}\right) = \left[v\lambda_1\rho_1\left(v\lambda_1, \frac{1}{2}\right)\right] - \left[\rho\left(v\lambda_2, \frac{1}{2}\right) + v\lambda_2\rho_1\left(v\lambda_2, \frac{1}{2}\right)\right] \\ &= \frac{2(v\lambda_1)^3 - 3(v\lambda_1)^2 + 3v\lambda_1 - 1}{2(v\lambda_1)^2(1 - v\lambda_1)^2} \end{aligned}$$

or

$$\exp[-\lambda_1](y_1 - y_2) = \frac{2(v\lambda_1)^3 - 3(v\lambda_1)^2 + 3v\lambda_1 - 1}{2(v\lambda_1)^2(1 - v\lambda_1)^2}$$

The two zero profit conditions (24) and (25) are also equivalent. Under symmetry, we have

$$\begin{aligned} \pi_G &= \pi_T = y_1 - g(\lambda_1)(y_1 - y_2) - g(\lambda_1 + \lambda_2)y_L^A \\ &+ \frac{2(v\lambda_1)^2 - v\lambda_1 + \frac{1}{2}}{2v(1 - v\lambda_1)^2} + \frac{2(v\lambda_2)^2 - v\lambda_2 + \frac{1}{2}}{2v(1 - v\lambda_2)^2} = c \end{aligned}$$

Using the adding up constraint $v(\lambda_1 + \lambda_2) = 1$, the symmetric equilibrium is characterized by the following system of two equations and two unknowns:

$$\begin{aligned} y_1 - (1 + \lambda_1)\exp[-\lambda_1](y_1 - y_2) - (1 + \frac{1}{v})\exp\left[-\frac{1}{v}\right]y_2 \\ + \frac{[2(v\lambda_1)^2 - 2v\lambda_1 + 1]^2}{2v[v\lambda_1(1 - v\lambda_1)]^2} = c \end{aligned} \quad (29)$$

$$\exp[-\lambda_1](y_1 - y_2) = \frac{2(v\lambda_1)^3 - 3(v\lambda_1)^2 + 3v\lambda_1 - 1}{2(v\lambda_1)^2(1 - v\lambda_1)^2} \quad (30)$$

We show the following

Proposition 1 *An equilibrium of the symmetric case always exists and is unique if (sufficient condition)*

$$0.003 < \frac{\sigma}{-\tau} < 10.65$$

Proof. See Section 4 in the Appendix. ■

4.2 The Benchmark: Competitive Symmetric Equilibrium

Before discussing the properties of the symmetric equilibrium with frictions, it is useful to characterize the competitive equilibrium with no frictions as the benchmark case. When workers can immediately locate their preferred job, the monopsonistic power of firms in a world with frictions - which allows them to post wages in order to influence the size of job queues - dissolves. In such equilibrium, graduates from school V work with certainty in jobs T , and graduates of schools A work with certainty in jobs G . With only these matches taking place, wages do not vary across jobs and are determined by the zero profit condition

$$y_1 \square w = c$$

With firms making zero profits, net output in this economy is simply

$$NY_C = 1 + \sigma\tau \square c$$

which allows us to establish the following remark

Remark 1 *When the labour market is perfectly competitive and firms make zero profits in equilibrium, it is optimal to start school tracking as early as possible ($\tau = 1$).*

From the vintage point of competitive equilibrium with no frictions, policies which delay tracking because of equality of opportunity incur efficiency losses.

5 Decentralized Matching versus Social Efficiency

In the matching model described in this paper, firms have monopsonistic power and post wages, and workers queue for their preferred jobs. In this section, we compare the outcome

of the decentralized equilibrium with the outcome which would obtain if the government could allocate workers to jobs and set the number of vacancies, using the same matching technology. This implies that the government can determine both queue lengths λ_1 and λ_2 , total vacancies v and the distribution of vacancies between the two jobs, μ . Re-write first the two key conditions for the symmetric decentralized equilibrium:

$$y_1 \square g(\lambda_1)(y_1 \square y_2) \square g\left(\frac{\lambda_1}{z}\right)y_2 + \frac{\lambda_1 [2z^2 \square 2z + 1]^2}{2z^3(1 \square z)^2} = c \quad ((29'))$$

$$\exp[\square \lambda_1](y_1 \square y_2) = \frac{(2z \square 1)(z^2 \square z + 1)}{2z^2(1 \square z)^2} \quad ((30'))$$

where $z \equiv v\lambda_1$, and $\frac{(2z \square 1)}{2z(1 \square z)} \frac{(z^2 \square z + 1)}{z(1 \square z)} > \frac{(2z \square 1)}{2z(1 \square z)}$.

The corresponding conditions for optimal policy are¹⁷

$$y_1 \square g(\lambda_1)(y_1 \square y_2) \square g\left(\frac{\lambda_1}{z}\right)y_2 = c \quad (31)$$

$$\exp[\square \lambda_1](y_1 \square y_2) = \frac{(2z \square 1)}{2z(1 \square z)} \quad (32)$$

By comparing these conditions we establish the following

Proposition 2 *Denote by $*$ the optimal policy, and by \dagger the matching equilibrium. First, optimal policy is unique. Second*

$$\begin{aligned} \lambda_1^* &> \lambda_1^\dagger \\ v^* &< v^\dagger \end{aligned}$$

i.e., in the decentralized matching equilibrium job queues are shorter and vacancies more abundant than in the socially optimal equilibrium.

Proof. See Section 6 of the Appendix. ■

Optimal policy and the matching equilibrium are illustrated in Figure 1. Here, we find it convenient to use the plane $(\lambda_1, z \equiv v\lambda_1)$. As shown in the Appendix, the first order condition for λ_1 is negatively sloped both in the matching equilibrium (we call it "Foc match" in the figure) and in the command economy ("Foc opt"). Moreover, the latter

¹⁷The derivation is in Section 5 of the Appendix.

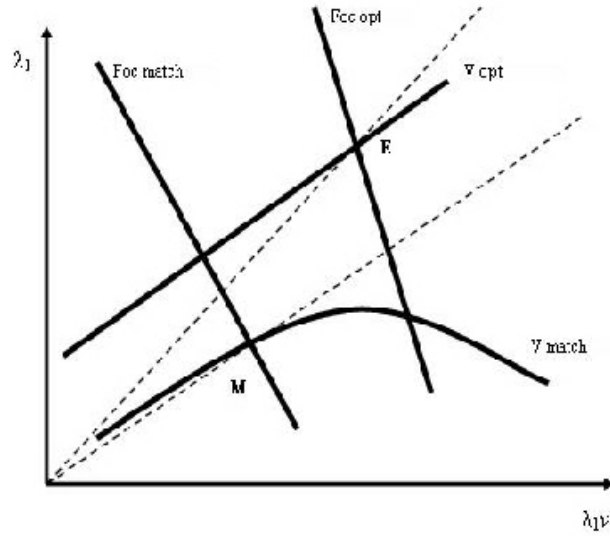


Figure 1: Matching Equilibrium and Socially Optimal Policy

is always strictly above the former. On the other hand, the zero profit condition in the command economy ("V opt") is positively sloped, and lies strictly above the zero profit condition for the matching equilibrium ("V match"), which is hump-shaped and with a slope which in absolute value is smaller than the slope of "Foc match". Thus, the combination $(\lambda_1, v\lambda_1)$ in the command economy, described by point E , lies above the combination in the matching equilibrium, which is at point M . Moreover, point M lies always below "Foc opt". This proves the ranking of queue lengths. Moreover, the slope of the lines connecting the origin with M and E is equal to $(\frac{1}{v})$. Clearly, the slope of the line passing through E is higher than the slope of the line passing through M . This proves the relationship between vacancies.

Such result is easy to interpret: under the decentralized matching equilibrium, job slots earn monopsonistic profits, which can only be diluted away by excessive entry of firms creating new vacancies. Applicant queues are too short as a result. The other side of the coin is the sub-optimal level of adjustment in the application probability, reflecting the fact that the wage differential between the more and the less productive match is smaller than

the socially optimal level, due to monopsonistic wage setting. Inspection of the first order conditions also reveals the following

Lemma 2 *The decentralized matching equilibrium is constrained inefficient when $\beta > 0$*

Proof. Straightforward from comparison of (29'), (30'), (31) and (32). ■

Job slots in the decentralized matching model face an upward sloping quasi - supply schedule. In the standard competitive search model, where $\beta = 0$, firms can choose the arbitrary size of the queue as far as they offer the equilibrium expected return. Just as monopsonistic firms have to increase their wage offers to attract more workers, the firms in our model have to pay a higher expected return to increase the number of applicants.

Ceteris paribus, an increase in β , the cost of adjusting application probabilities, increases profits, because such an increase reduces the response of the queue to a change in the expected return from each application: i.e., the supply (job application) schedule becomes less elastic. Since the wedge between labor productivity and wage is increasing in the inverse of supply elasticity (it vanishes as the supply becomes infinitely elastic), profits are increasing in β .

The distortions typical of a search market result in excessive profits and hence in excessive entry of vacancies in the market. It is a feature of our matching model that the equilibrium fails to achieve constrained social efficiency, unlike the standard competitive search model (see Rogerson, Shimer and Wright, 2005). Moreover, the inefficiency is such that there are too many vacancies and too few applicants per job slot. Given monopsonistic wage setting, workers under-invest in adjusting their application probabilities. Hence we have excessive mismatch.

5.1 Features of the Symmetric Equilibrium

One feature of the symmetric equilibrium is that total employment is increasing in the number of vacancies. Using the fact that $\lambda_1 + \lambda_2 = \frac{1}{v}$, total employment e is given by

$$e = v [(\lambda_1) + (\lambda_2)(1 - (\lambda_1))] = v \left[1 - \exp\left(-\frac{1}{v}\right) \right] \quad (33)$$

and the employment/vacancy rate $\frac{e}{v}$ decreases as vacancy creation speeds up.

Changes in queue lengths only affect the composition of employment, not its total. This is because

$$\frac{\partial(e_A^G + e_T^V)}{\partial\lambda_1} = v [1 - (\lambda_1)] > 0$$

and total employment does not depend on the application rate. A higher application rate by workers A and V to jobs G and T increases the number of successful matches and reduces the number of less successful matches (of workers A with T jobs and workers V with G jobs).

5.2 Optimal Tax Policy

The decentralized equilibrium in our model not only generates mismatch but also fails to achieve constrained efficiency. This is due to the fact that firms have some monopsonistic power in their wage posting, which affects the application strategy selected by workers. As a result, in the symmetric equilibrium job queues for the most productive matches are too short and vacancies are too many, compared to the constrained social optimum. Facing this outcome, the government could try to attain the constrained social optimum by using an appropriate combination of taxes and transfers. There is a small literature which explicitly considers taxes and education as alternative tools to attain equity and efficiency, with a special emphasis on redistribution (see for instance Benabou, 2002, and Hanushek, Ka Yui Leung and Tilmaz, 2003). Here, we are not concerned with redistribution, but with efficiency, and in particular with the existence of a proper combination of taxes and subsidies which can allow the government to attain the constrained social optimum for any education policy in place.

For this to happen, we need two instruments. Suppose that the government can always meet his budget constraint, either by raising revenue with non-distortionary taxation or by giving away lump sum subsidies so as to dispose of excessive revenue. First, we know that in the decentralized equilibrium workers do not adjust adequately their application probabilities in the direction of the most productive matches, because of the compression of the wage distribution induced by monopsonistic wage setting. A negative income tax - or a wage subsidy - can be used to increase the wage premium and provide the right incentives to job searchers. To illustrate, consider equation (13), which describes the optimal choice of application probabilities in the absence of taxes and subsidies, and let s^j be the subsidy paid to each type of graduate. Wages gross of subsidy for type A graduates are

$$\begin{aligned}\tilde{w}_A^G &= (1 + s^A)w_A^G \\ \tilde{w}_A^T &= (1 + s^A)w_A^T\end{aligned}$$

and the expected return from a match between an A graduate and a G job is

$$\tilde{E}_A^G = \frac{(\lambda_A^G)(1 + s^A)(w_A^G \square w_A^T)}{\lambda_A^G}$$

If we apply the same scheme to the other type of graduates, equation (13) can be re-written as

$$\begin{aligned}E_A^G \square E_A^T &= \frac{1}{(1 + s^A)} \left[\frac{p_A \square \mu}{p_A(1 \square p_A)} \right] \\ E_V^T \square E_V^G &= \frac{1}{(1 + s^V)} \left[\frac{p_V \square (1 \square \mu)}{p_V(1 \square p_V)} \right]\end{aligned}$$

Second, we notice that the other source of inefficiency is excessive entry by firms, because job slots earn excessive profits due to their monopsonistic power in the labor market. To correct this inefficiency, the government can impose a tax on vacancy creation, t . Consider the symmetric matching equilibrium. The following Proposition shows that there exists a combination of subsidies and taxes which can move the matching equilibrium to the socially efficient allocation.

Proposition 3 *Denote by $(\Lambda_1, \Lambda_2, V)$ the socially optimal values of $(\lambda_1, \lambda_2, v)$. Then constrained efficiency can be restored starting from the decentralized symmetric matching equilibrium by using the following wage subsidy s and tax on vacancies t*

$$1 + s = \frac{\square(\Lambda_1, V, \mu)}{(\Lambda_1, V, \mu)} \quad (34)$$

$$t = \Theta(\Lambda_1, \Lambda_2, V, \mu) \square \Psi(\Lambda_1, \Lambda_2, V, \mu) \quad (35)$$

where

$$\equiv \left[\frac{V\Lambda_1 \square \frac{1}{2}}{\Lambda_1(\frac{1}{2} \square \mu V\Lambda_1)} \right]$$

$$\square \equiv [2\mu V\Lambda_1\rho_1(2\mu V\Lambda_1, 1 \square \mu)]$$

$$\square [\rho(1 \square 2\mu V\Lambda_1, 1 \square \mu) + (1 \square 2\mu V\Lambda_1)\rho_1(1 \square 2\mu V\Lambda_1, 1 \square \mu)]$$

$$\Psi \equiv \frac{\mu}{2V} [\log(2\Lambda_1) \square \log(2\Lambda_1)] + \frac{\mu}{2V} [\log(2\Lambda_2) \square \log(2\Lambda_2)] = 0$$

$$\Theta \equiv 2 \mu V \Lambda_1^2 \rho_1(2\mu V \Lambda_1, \mu) + 2 \mu V \Lambda_2^2 \rho_1(2\mu V \Lambda_2, \mu)$$

Proof. See Section 7 of the Appendix ■

In the symmetric equilibrium, queue lengths to the most productive match are too short and the number of vacancies is too large compared to the socially optimal policy. Hence, the government needs to set up a wage subsidy to encourage additional search by school graduates, and a tax to discourage vacancy creation.

How big are these taxes and subsidies? In the symmetric equilibrium, the optimal subsidy is given by

$$1 + s = \frac{\square(\Lambda, V)}{(\Lambda, V)} = \frac{(2z \square 1)(z^2 \square z + 1)}{2z^2(1 \square z)^2} \div \frac{(2z \square 1)}{2z(1 \square z)}$$

or

$$s = \frac{2z^2 \square 2z + 1}{z(1 \square z)} \geq 2 \quad \text{since} \quad z \equiv v\lambda_1 \geq \frac{1}{2}$$

The optimal tax on vacancy creation is instead

$$t = \frac{\lambda_1 [2z^2 \square 2z + 1]^2}{2z^3(1 \square z)^2}$$

To illustrate with a numerical example, let $\beta = .01, v = 1, \lambda_1 = .7$; then $t = .038$. It follows that constrained efficiency can be restored only with a very large subsidy, equal at least to twice the wage. This, and the fact that taxes and subsidies are used for redistribution, suggests that in practice constrained efficiency is not attained, and that the government operates its education choices subject to the decentralized equilibrium.

6 The General Model: Numerical Solutions

The symmetric model is tractable enough for us to derive analytical results concerning existence and uniqueness, as well as the key properties of the decentralized equilibrium vis a vis the socially optimal outcome. When we drop symmetry and go back to the general model, however, tractability is lost, and we are forced to illustrate the properties of the general model using numerical solutions. The general model contains two policy variables, τ and η , six parameters: $\sigma_i^j, i = A, V, j = G, T,$ and c , and four endogenous variables:

λ_A^G , λ_V^T , v and μ . We reduce the space of parameters by imposing the following restrictions in the technology of production of human capital

$$\sigma_A^G = \sigma_A^T = \sigma_A$$

$$\sigma_V^G = \sigma_V^T = \sigma_V$$

Therefore, we allow the "productivity of tracking", measured by parameter σ , to vary between tracks, but impose that the rate of accumulation or de-cumulation of skills within each track is the same.

6.1 Total Net Output and Taxes with Decentralized Matching and in the Command Economy

Before discussing how we proceed to estimate the vector of parameters, we compare total output net of hiring and search costs in the decentralized and in the command economy equilibrium when each parameter is allowed to vary with respect to a baseline configuration. Let this baseline be

$$\sigma_A = \sigma_V = 0.08; \quad \eta = 0.05; \quad c = 0.35; \quad \tau = 0.5; \quad \eta = 0.5$$

where we also fix the policy variables, and allow one parameter at a time to vary while the rest remains constant at the baseline. The results are illustrated in Figure 2. As expected, total net output is always higher in the command economy equilibrium than in the decentralized matching equilibrium: by directly selecting vacancies, queue lengths and wages, the government can eliminate the inefficiencies associated to excessive vacancy creation and limited search activity, and reduce the share of bad matches. However, total employment is higher in the decentralized equilibrium than in the command optimum, due to the excessive production of vacancies in the former case. Notice that net output decreases both in the decentralized and in the command equilibrium when the costs of setting up vacancies and of searching for the right match increase, and increases when the productivity of human capital formation in the academic track is allowed to raise.

The general version of our model accommodates two types of mismatch: the first type is the mismatch between worker types and job types which occurs when the total supply of each type of worker is equal to the total demand. In the symmetric model, $\eta = \mu$, but workers still can end up in a bad match. The second type occurs only in the general

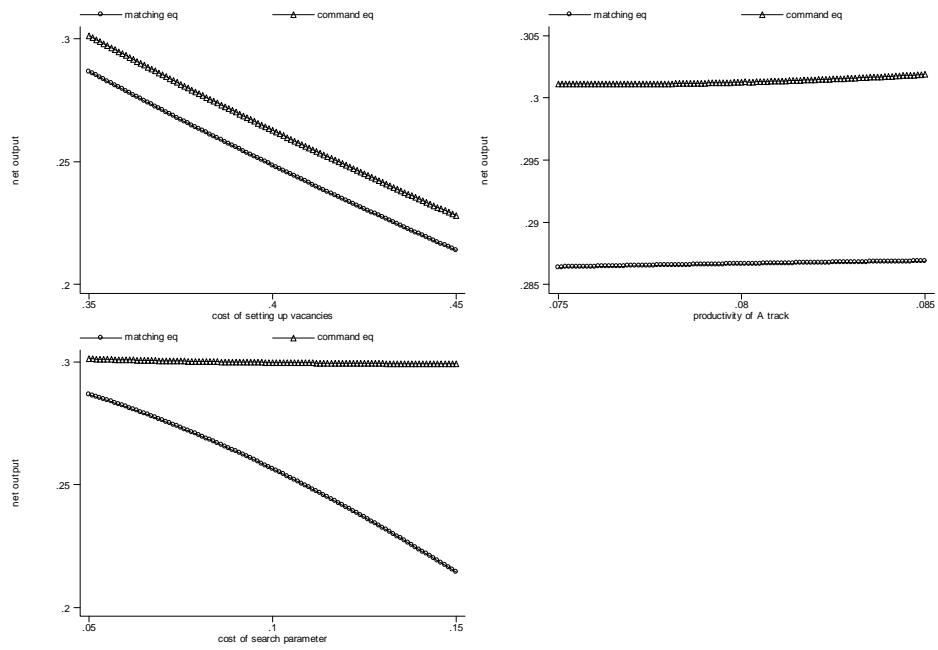


Figure 2: Net output in the matching equilibrium and the command economy

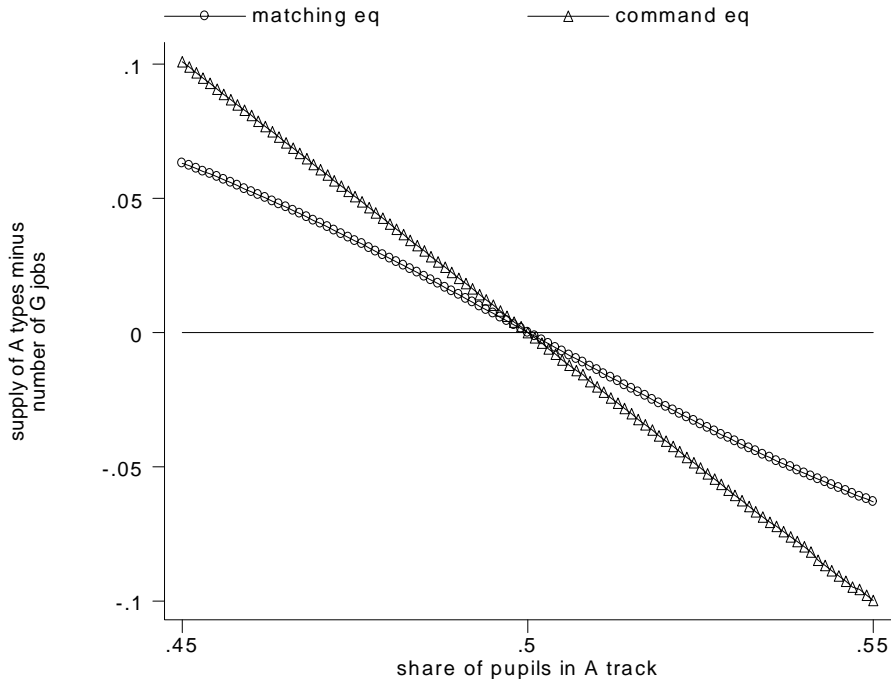


Figure 3: The supply of A graduates and the imbalance between demand and supply

model when $\eta \neq \mu$, and implies that the supply of each worker type does not match the (endogenous) demand. We call this second type of mismatch imbalance, and we illustrate in Figure 3 how it varies with changes in the share η , both with decentralized matching and in the command economy.

First of all, in both matching equilibrium and command economy, the absolute size of the imbalance increases as we deviate from the symmetric case $\eta = \frac{1}{2}$. In both equilibria, an increase in η brings about a more than proportionate increase in demand, measured by the share μ of vacancies for type G jobs. By the same token, if the supply of the A skill is reduced below $\frac{1}{2}$, the demand is reduced even more. Interestingly enough, this ‘overshooting’ phenomenon is even more pronounced in the command economy. The other side of the coin is the virtual invariance of the total number of mismatches $e_A^T + e_V^G$ as η varies, implying that an increase in the mismatch of one type of worker is largely offset by a counter-balancing decline in the mismatch of the other type of worker. To put it differently, in both equilibria the matching mechanism minimizes total mismatch, and firms respond strongly to changes in supply by adjusting the type of job slots on offer. Notice that shifting

the share of vacancies from one type of job to the other is costless¹⁸.

If our numerical examples are a good indication of the more general properties of the matching equilibrium described in the paper, one intriguing implication is that policy discussions about mismatch in the labor market, which typically focus on imbalances between supply and demand for each type of skill, are somewhat misplaced, as mismatch can occur even in the absence of such imbalances.

Finally, we ask what would be the combination of wage subsidies and taxes on job slots that would allow the decentralized equilibrium to attain the command economy. Again, we keep the policy variables and the productivity of human capital formation at their baseline values and allow β and c to vary about their baseline values. As illustrated by Figure 4, it turns out that the wage subsidy required to attain the social optimum is remarkably high, at about 60 times the wage. On the other hand, the tax on job slots varies with the underlying parameters in the range between 10 and 40 percent of output.

What should we conclude from this? Clearly, the estimated wage subsidies are unreasonably high. This result, and the common observation that taxation is used mainly for redistribution, suggests that attaining the social optimum by combining taxes and subsidies is either not feasible or not consistent with actual government policy. We take stock by limiting our numerical solutions of the model to the decentralized matching equilibrium.

6.2 Numerical solution: the strategy

One straightforward strategy would be to produce independent estimates of the parameters, and use these estimates to solve the model for the endogenous variables. Unfortunately, this is a close to impossible task for some of the parameters, such as β and σ_V , for which we have no clear empirical counterpart. The former enters in the measure of the cost of search faced by job searchers, and the latter measures the productivity of vocational schools in the production of vocational skills. While we do have statistics on cognitive skills, our information on non - cognitive skills is still at its infancy¹⁹. An alternative strategy is to treat these parameters as unknowns, and estimate them using a minimum distance estimator, which minimizes the distance between the numerical and actual value of some

¹⁸Total vacancies do not change very much anyway in both equilibria.

¹⁹See for instance Heckman, Stixrud and Urzua, 2006.

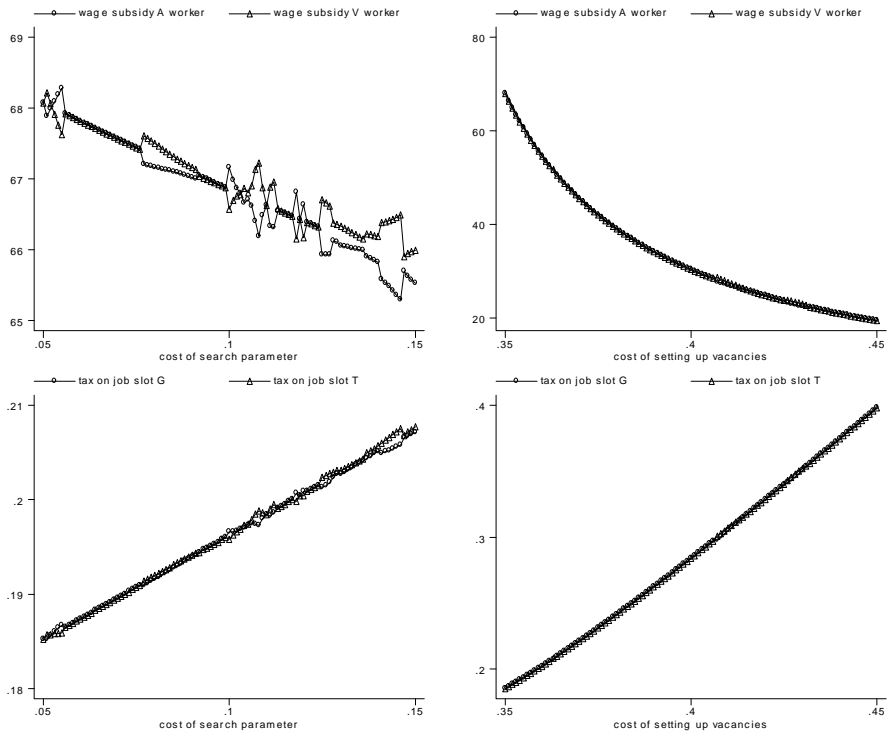


Figure 4: Taxes and subsidies as functions of the key parameters

measurable indicator. A natural candidate is employment in jobs G and T . Based on our model, this is given by

$$e^G = e_A^G + e_V^G = v^G [(\lambda_A^G) + (\lambda_V^G)(1 - (\lambda_A^G))] \quad (36)$$

$$e^T = e_A^T + e_V^T = v^T [(\lambda_A^T) + (\lambda_V^T)(1 - (\lambda_A^T))] \quad (37)$$

Suppose that our empirical sample of observations on sectorial employment covers N countries, and denote as $D_j = \{e_j^G, e_j^T\}$, $j = 1, 2, \dots, N$, the vector of observed employment rates. Define the following objective function

$$Z = \frac{1}{\omega_G} \sqrt{\frac{\sum_{j=1}^N (e_j^G - \hat{e}_j^G)^2}{N^2}} + \frac{1}{\omega_T} \sqrt{\frac{\sum_{j=1}^N (e_j^T - \hat{e}_j^T)^2}{N^2}} \quad (38)$$

where $\omega_G = \frac{std(e_j^G)}{mean(e_j^G)}$, $\omega_T = \frac{std(e_j^T)}{mean(e_j^T)}$, and e_j^i and \hat{e}_j^i , $i = G, T$, are the actual and computed values of employment in each sector. Further assume that the policy variables are set at their actual values. Then the parameters of interest are obtained by minimizing the weighted sum of the distance between computed and actual employment in each sector

$$\{ \lambda_A, \lambda_V, c, \sigma_A, \sigma_V \} = \arg \min[Z]$$

To implement this strategy, we need empirical information on employment in G and T jobs, and estimates of actual tracking time τ and the share of students in academic tracks η .

6.3 The Data

We collect data on employment by job type, tracking time and the share of students in the academic track for a sample of 20 countries, which spans 4 continents and covers mainly developed countries. Since we need employment in G and T jobs, we need first of all to map existing occupations into the two jobs. We proceed by selecting two benchmark countries, Germany and the US, and by restricting our attention to high school and college graduates. Our data for Germany are from the European Community Household Panel (ECHP). For the US, we use the March supplements of the Current Population Survey. For

both countries, we select the years 1996 and 2000, and average our results to smooth away potential business cycle effects.

For the US, we include in the high school category the individuals with some college but no degree as well as those with an associate degree. We eliminate public sector jobs and retain only individuals aged 25 to 34, who are most likely to be in the transition from school to work. We also assign some occupations to the secondary labour market, which in our model is observationally equivalent to unemployment. These occupations include agricultural work, sales and services, plant assembly and elementary blue collar jobs. Among the remaining occupations, we assign to G jobs the occupations where the percentage of employees with a college degree is higher than 50%. The underlying idea is that individuals enrolled in academic tracks are much more likely to go on to college than individuals enrolled in vocational tracks. The remaining occupations are assigned to T jobs. For the US, the occupations mapped to G jobs are managers, executives and professionals, where the percentage of college graduates in the year 2000 was 56% and 78% respectively. On the other hand, the occupations mapped into T jobs are technicians, clerks and specialized blue collar labor. Since Germany turns out to have a distribution of college graduates by occupation similar to the US, we confirm the same mapping, and extend such mapping to the other countries in the dataset.

To obtain e^G and e^T , we compute the total labor force by adding to employment the unemployed and those assigned to the secondary sector, and divide actual employment in the two jobs by this total. It is impossible to cover 20 countries with the same dataset. We use the ECHP for most European countries, and the tabulations from ILO (www.laborsta.org) for the rest of the sample. Needless to say, available tabulations do not fit perfectly our selection of data, especially with reference to educational attainment and public versus private employment. Compared to US and ECHP data, our estimates for the rest of the sample covers both public and private sector and all educational groups, but restrict age to less than 35.

The data on τ and η are drawn from Table 1 in Brunello and Checchi, 2006. Tracking length is measured as the percentage of total schooling time from primary to upper secondary education spent in a tracked school, and refers to the year 2002. Notice that we have assigned the value 0 to both τ and η in the United States and Canada. This is not to say that US schools do not have tracking. However, since tracking occurs within a comprehensive schooling system, there is no obvious measure that we can use. Therefore,

we exclude this country - as well as Canada - from the numerical exercise, thereby reducing the number of available countries to 18. The table reports for the same year the share of students in upper secondary general education, and data on average tax rates on earnings - both drawn from OECD sources. Since actual employment subsidies expressed as shares of GDP are very small in most countries, we set them to zero as a first approximation.

Table 1. Values of employment shares, taxes, subsidies, tracking length and share of students in general tracks.

Country	e^G	e^T	τ	η	t
Austria	.109	.588	.154	.370	.240
Belgium	.239	.439	.667	.273	.141
Canada	.212	.437	0	1	.166
France	.178	.535	.250	.437	.154
Germany	.156	.548	.692	.370	.209
Italy	.085	.492	.385	.732	.181
Japan	.187	.532	.250	.751	.066
Spain	.190	.376	.167	.720	.140
UK	.359	.426	.154	.279	.174
US	.320	.297	0	1	.157
Finland	.288	.373	.250	.428	.249
Denmark	.196	.522	.250	.470	.304
Holland	.397	.399	.500	.308	.105
Ireland	.265	.391	.182	.763	.126
Greece	.171	.393	.250	.600	.056
Portugal	.245	.514	.250	.712	.101
Russian Fed	.133	.564	.217	.671	.13
Poland	.119	.472	.385	.391	.064
Mexico	.044	.487	.455	.886	.063
New Zealand	.154	.465	.154	.628	.205

6.4 The Results

With 18 available data points, we cannot possibly produce country specific estimates of all the parameters. A more modest approach is to fix σ_i and c to be constant across countries and allow β to take three different values, starting from some initial value. Let β_L , β_M and β_H be the low, medium and high value of β . Conditional on the policy and tax variables

taking their actual observed values, the minimum distance estimator produces the following estimates

$$\sigma_A = 0; \sigma_V = 0.065; c = 0.35; L = 0.018; M = 0.046; H = 0.095$$

Interestingly, the distance between actual and simulated employment in jobs G and T is minimized when the productivity of human capital formation is significantly higher in the vocational track, which specializes in the production of vocational skills, than in the academic track, which is not significantly more effective than the comprehensive track in the production of cognitive skills. This empirical result squares well with the findings by Hanushek and Woessmann, 2006, that tracking does not affect average performance in cognitive test scores, and implies that the gains from specialization accruing to tracking, if any, lie in the less prestigious and practically oriented track.

With these estimates in hand, we solve the model for the endogenous variables, including employment in the two jobs and total net output. Table 2 compares actual employment shares with the computed values for each country in the sample. It turns out that the correlation between actual and computed values is rather high for T jobs (0.87) but somewhat lower for G jobs (0.53). Overall, we take these results as indicating that our model fits reasonably well actual data, and that this fit is significantly better for technical jobs typically filled by individuals with vocational skills.

Table 2. Values of actual and predicted employment shares.

Country	actual e^G	estimated e^G	actual e^T	estimated e^T
Austria	.109	.099	.588	.594
Belgium	.239	.257	.439	.462
France	.178	.091	.535	.550
Germany	.156	.221	.548	.494
Italy	.085	.109	.492	.525
Japan	.187	.125	.532	.497
Spain	.190	.253	.376	.394
UK	.359	.311	.426	.392
Finland	.288	.310	.373	.401
Denmark	.196	.112	.522	.529
Holland	.397	.223	.399	.480
Ireland	.265	.271	.391	.374
Greece	.171	.301	.393	.390
Portugal	.245	.121	.514	.505
Russian Fed	.133	.125	.564	.500
Poland	.119	.251	.472	.445
Mexico	.044	.134	.487	.486
New Zealand	.154	.150	.465	.472

Finally, we ask what are the values of the education policy variables, τ and η , which maximize total net output under the estimated configuration of parameters. These values are reported in Table 3 below, together with the percentage deviation between total net output when the policy variables are at their optimum level, and total net output when the policy variables take their current value. Our results suggests the following:

- optimal track length when the labour market is frictional is generally shorter than in a hypothetical country characterized by a perfectly competitive labour market, possibly with the exception of Belgium. The gap is rather small in Germany, Finland and the UK, but quite large in France, Denmark, Austria and New Zealand;
- on average, actual tracking length (0.331) is much shorter than the level which maximizes total output in a decentralized matching equilibrium (0.471). However, this average masks substantive cross country heterogeneity: in about half of the countries in our sample, actual tracking length is too long compared with the efficient length. The opposite occurs in the remaining half of the sample. These findings warn us against drawing easy conclusions on whether the current tracking length is "too long"

or "too short". We are afraid that the existing relationship between total net output and tracking time in imperfectly competitive labour markets is too complex to warrant easy generalizations;

- on average, the actual share of pupils in the academic track (0.530) is much lower than the level which maximizes total output in a decentralized matching equilibrium (0.870). Compared to tracking length, this result is less ambiguous, as optimal η is always larger than actual η . How do explain this? Again, the relationship between net output and η is complex, but we speculate that, by increasing the share of students in academic schools, who have more versatile skills ($\sigma_A = 0$) than students in vocational schools ($\sigma_V > 0$), the government can reduce the output losses generated by bad matches²⁰.
- on average, the output loss incurred by having both a tracking time and a share of students in the academic track different from their optimal values is close to 2 percentage points of total net output, a sizeable loss, especially if we consider that most developed OECD countries spend between 4 and 5 percent of their GNP for public education.

²⁰When both η and τ are set at their optimal values, the supply of vacancies declines, and the share of G jobs increases.

Table 3. Values of actual and optimal education variables, and percentage output loss.

Country	actual τ	optimal τ	actual η	optimal η	% output loss
Austria	.667	.031	.277	.522	7.9
Belgium	.500	.998	.299	.891	2.6
France	.250	.074	.437	.799	2.4
Germany	.692	.969	.370	.904	4.2
Italy	.385	.206	.732	.979	1.3
Japan	.250	.151	.751	.976	0.6
Spain	.167	.710	.620	.969	0.3
UK	.154	.921	.279	.906	0.4
Finland	.250	.958	.428	.893	0.4
Denmark	.250	.029	.470	.830	2.0
Holland	.500	.895	.308	.916	3.6
Ireland	.182	.165	.763	.872	0
Greece	.250	.839	.600	.919	0.3
Portugal	.182	.141	.712	.973	1.0
Russian Fed	.250	.116	.671	.961	1.0
Poland	.250	.812	.391	.915	1.6
Mexico	.455	.405	.886	.991	0.6
New Zealand	.154	.074	.628	.949	0.6

7 Conclusions

An important pattern in school design after the second world war has been the detracking of secondary schools, motivated mainly by equity considerations: early tracking systems generate early stratification, with working class students confined to less prestigious and rewarding vocational schools, and students from privileged families having access to more elitist academic tracks and college.

As usual, equity comes at a price in terms of efficiency. We have argued in this paper that one cannot evaluate the relative efficiency of school tracking by looking only at what happens in schools. The transition from school to work is also important. In an ideal frictionless labour market, such transition is so smooth that it might be efficient to push tracking to cover the full length of schooling. We show that, in a more realistic imperfectly competitive labour market with frictions, the hazards in the transition from school to work may lead governments to delay tracking, even in the absence of equity considerations. This has two implications: first, an alternative interpretation of the de-tracking movement can be framed in terms of efficiency, rather than equality of opportunity; second, deviations

from complete tracking driven by equity considerations do not necessarily entail efficiency losses, as it might be efficient to delay tracking with respect to the benchmark model with perfectly competitive labour markets.

Our characterization of market frictions builds on the directed competitive search model, as discussed by Shi, 2002, and Rogerson, Shimer and Wright, 2005. We innovate with respect to this setup by introducing the assumption that workers bear a positive cost of adjusting applications with respect to random matching. This innovation implies that the supply of workers to jobs becomes upward sloping, and that firms can use the associated monopsonistic power to extract rents from workers. These rents can only be washed away by excessive firm entry.

We have restricted our analysis to the symmetric equilibrium, which we prove to be existent and unique under some relatively mild restrictions. Our numerical solutions for the more general model show that optimal tracking time when the labour market is frictional is generally postponed with respect to the optimal tracking time in a perfectly competitive labour market. On average, the actual length of educational tracks is also shorter than that which maximizes total output in a decentralized matching equilibrium. However, this average masks substantive cross country heterogeneity, because in about half of the countries in our sample, actual tracking length is too long compared with the efficient length, while the opposite occurs in the remaining half of the sample.

We also find that, on average, the actual share of pupils in the academic track is much lower than the level which maximizes total output in a decentralized matching equilibrium. Compared to tracking length, this result is less ambiguous, as the optimal share is always larger than the actual share. Perhaps more important for the purposes of this paper, we find that the output loss incurred by having both a tracking time and a share of students in the academic track different from their optimal values is close to 2 percentage points of total net output. At the forefront of this paper we asked whether de-tracking secondary schools has a significant efficiency cost. Our answer has three parts: first, we should not take the competitive labour market as a benchmark, because labour market frictions are a relevant ingredient of the school to work transition; second, the actual configuration of tracking time and allocation of students to tracks deviates from the optimal configuration prevailing in an imperfect labour market, which entails a sizeable average efficiency loss. Third and last, while there are countries in our sample where school tracking is too short compared to the optimum, there are also countries where tracking is too long.

Tractability has required that we strip down the schooling part of the model considerably, and that we consider a two-period model. In future work, we intend to move to a multi-period environment, possibly with a richer school structure. The current model can also be used to explore the consequences on optimal school design induced by the introduction of unexpected productivity shocks, which hit some of the jobs available in the economy. Typically, a consequence we expect is a higher demand for versatility, which could lead governments to further de-track secondary schools, independently of equality considerations²¹.

²¹This is not a new question, see for instance Lamo, Messina and Wasmer, 2006.

References

- [1] Albrecht, J., Gautier, P., and Vroman, S., 2003, Matching with Multiple Applications, *Economics Letters*, 78, 67-70
- [2] Ariga, K., G.Brunello, R.Iwahashi, and Rocco, L., 2005. Why Is the Timing of School Tracking So Heterogeneous? *IZA discussion paper* n.1854
- [3] Benabou, R., 2002, Tax and Education Policy in an Heterogeneous-agent Economy: What Level of Redistribution Maximize Growth and Efficiency?, in *Econometrica*, 70,2, 481-517
- [4] Bertocchi, G. and Spagat, M., 2003, The Evolution of Modern Education Systems, in *Journal of Development Economics*.
- [5] Brunello, G., K,Ariga and M.Giannini. 2006. The Optimal Timing of School Tracking, forthcoming in P. Peterson and L.Wößmann, (eds), *Schools and the Equal Opportunity Problem*, MIT Press, Cambridge MA. (also IZA Discussion Paper 955)
- [6] Brunello. G. and Checchi, D., 2006, Does School Tracking Affect Equality of Opportunity? New International Evidence, forthcoming in *Economic Policy*.
- [7] Card, D. and Rothstein, J., 2006, Racial Segregation and the White-Black Test Score Gap, *NBER wp* 12078
- [8] Card, D. and Krueger, A., 1990, Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy*: 1-39.
- [9] Delacroix, C. and Shi, S., 2006, Directed search on the job and the wage ladder, *International economic Review*, 651-699.
- [10] Green, A., Wolf A. and Leney, T. (1999), *Convergence and Divergence in European education and training systems*, Bedford Way Papers, Institute of Education, University of London, 1999.
- [11] Hallilan, M., The Detracking Movement, *Education Next*, 2006, 1-5
- [12] Hanushek, E., Ka Yui Leung, C. and Yilmaz, K., 2003, Redistribution through Education and Other Transfer Mechanisms, in *Journal of Monetary Economics*, 50, 1719-50.

- [13] Hanushek, E. and L.Wößmann. 2006. Does Educational Tracking Affect Performance and Inequality? Differences-In-Differences Evidence Across Countries. *Economic Journal* 116: C63-C76.
- [14] Heckman, J., Stixrud, J., and Urzua, S., The Effects of Cognitive and Non-cognitive Ability on Labour Market Outcomes and Social Behaviour, *NBER Working Paper* 12006.
- [15] Lamo, A., Messina, J., Wasmer, E., 2006, Are specific skills an obstacle to labour market adjustment? ECB Working Paper.
- [16] Manning, A. and J.S.Pisckhe. 2006. Comprehensive versus Selective Schooling in England in Wales: What Do We Know? *IZA Discussion Paper* n.2072.
- [17] Minter Hoxby, C. 2000. Peer effect in the classroom: learning from gender and race variation. *NBER wp.7867*
- [18] Nickell, S., Dynamic Labor Demand, in Layard, R. and Ashenfelter, O., eds. 1986, *Handbook of Labor Economics*, North Holland.
- [19] Oakes, J. 1992. Can Tracking Research Inform Practice? Technical, Normative, and Political Considerations. *Educational Researcher* 21(4): 12-21
- [20] Palme, M. and C.Meghir. 2005. Assessing the effects of Schooling on wages using a social experiment. *American Economic Review* 95(1): 414-424
- [21] Pissarides, C., 1985, Unemployment, vacancies and real wages, *American Economic Review*, 660-689
- [22] Rogerson, R., Shimer, R. and R.Wright, 2005, Search theoretic models of the labour market, *Journal of Economic Literature*, 959-988
- [23] Shi, S., 2002, A directed search model of Inequality with heterogeneous skills and skill based technologies, *Review of Economic Studies*, 467-91
- [24] Shimer, R., 2005, The cyclical behaviour of unemployment and vacancies, *American Economic Review*, 25-49
- [25] Yashiv, E., 2000, The determinants of equilibrium unemployment, *American Economic Review*, 1297-1322.

8 Appendices

A.1

Let n be the index of individual job slots. This index is chosen in such a way that job slots belong to type G if $0 \leq n \leq v\mu$, and to type T if $v\mu < n \leq v$. Denote by $\beta(n)$ the application density, which must sum up to unity:

$$\int_0^v \beta(n) dn = 1 \quad (\text{A1.1})$$

Random applications imply that

$$\beta(n) = \frac{1}{v}$$

so that

$$\bar{p}_A = \int_0^{v\mu} \frac{1}{v} dn = \mu$$

is the probability that the application is sent to a type G job slot, and

$$1 - \bar{p}_A = \int_{v\mu}^v \frac{1}{v} dn = 1 - \mu$$

is the probability that the application reaches a type T job slot.

We now posit the following cost function

$$\begin{aligned} C &= \int_0^v \left[\beta(n) - \frac{1}{v} \log(\beta(n)) \right] dn - \int_0^v \left[\frac{1}{v} - \frac{1}{v} \log\left(\frac{1}{v}\right) \right] dn \\ &= \int_0^v \left[\beta(n) - \frac{1}{v} \log(\beta(n)) \right] dn - (1 + \log(v)) \end{aligned} \quad (\text{A1.2})$$

The second term in the RHS is for normalization, so that if workers set $\beta(n) = \frac{1}{v}$, they do not incur any cost. The cost is strictly positive and it increases without bound if we let $\beta(n) \rightarrow 0$. Denote by $E(n)$ the expected returns from applying to the offer posted by a job slot n . The worker's optimal policy is to maximize the following objective function:

$$U = \int_0^v E(n) \beta(n) dn - \int_0^v \left[\beta(n) - \frac{1}{v} \log(\beta(n)) \right] dn - (1 + \log(v)) \quad (\text{A1.3})$$

subject to (A1.1).

Let us assume that each type of job slot has a common wage policy, given by E_A^G and E_A^T for type A workers. The argument is exactly the same for type V workers so we skip it.

Type A worker's problem is reduced to:

$$\mathcal{L}^A \equiv U + \xi \left[1 - \int_0^v \beta(n) dn \right]$$

where ξ is Lagrangian multipliers for (A1.1). The first order condition is

$$\frac{\partial \mathcal{L}^A}{\partial \beta(n)} = E(n) - \left[1 - \frac{1}{v\beta(n)} \right] \xi = 0$$

so that we have

$$\begin{aligned} E_A^G - E_A^T &= \left(1 - \frac{1}{v\beta(n_G)} \right) - \left(1 - \frac{1}{v\beta(n_T)} \right) \\ &= \left(\frac{1}{v\beta(n_T)} - \frac{1}{v\beta(n_G)} \right) = \left(\frac{\beta(n_G) - \beta(n_T)}{v\beta(n_G)\beta(n_T)} \right) \end{aligned}$$

Notice that:

$$\begin{aligned} \int_0^{v\mu} \beta(n) dn &= v\mu\beta(n_G) = p_A \\ \int_{v\mu}^v \beta(n) dn &= v(1 - \mu)\beta(n_T) = 1 - p_A \end{aligned}$$

Hence we have:

$$\begin{aligned} \left[\frac{\beta(n_A) - \beta(n_B)}{\beta(n_A)\beta(n_B)} \right] &= \left[\frac{\frac{v\mu\beta(n_A) - \mu}{v\mu(1 - \mu)}}{\frac{v\mu\beta(n_A)(1 - v\mu\beta(n_A))}{v\mu(1 - \mu)}} \right] \\ &= \left[\frac{v\mu\beta(n_A) - \mu}{v\mu\beta(n_A)(1 - v\mu\beta(n_A))} \right] = \left[\frac{p_H - \mu}{p_H(1 - p_H)} \right] \end{aligned}$$

Consequently

$$\begin{aligned} E_A^G - E_A^T &= \left[\frac{p_A - \mu}{p_A(1 - p_A)} \right] = \rho(p_A, \mu) \\ \rho(x, \bar{x}) &\equiv \frac{x - \bar{x}}{x(1 - x)} \end{aligned}$$

Similarly, for type V workers, we have

$$\begin{aligned}
E_V^T \square E_V^G &= \left[\frac{p_V \square (1 \square \mu)}{p_V(1 \square p_V)} \right] \\
&= \rho(p_V, 1 \square \mu)
\end{aligned}$$

as in the text.

A.2

Proof of Lemma 1

This result is proved in Shimer (2005) when $\eta = 0$, but applies more generally. Suppose that the firm holding the job slot wants to increase the size of the queue of applicants. To do so, it has to raise the posted wage both to compensate applicants for the decline in their matching probabilities, and to attract more applicants. Differently put, an increase in the queue length is more expensive in terms of wages if η is larger. This in turn reduces the desired change in the queue length. Thus, the net profit after optimization is a monotone and concave transformation of gross output, more concave than under competitive search. More formally, we only prove the Lemma for y_A^G and w_A^G , since the other cases are similar. The first order condition for λ_A^G is

$$\begin{aligned}
'(\lambda_A^G) \square y_A^G \square (\lambda_V^G)y_V^G &= \tag{A2.1} \\
E_A^T + \left[\rho \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) + \frac{v\mu\lambda_A^G}{\eta} \rho_1 \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) \right]
\end{aligned}$$

The posted wage must satisfy the worker arbitrage condition

$$(\lambda_A^G)w_A^G = \lambda_A^G \left[E_A^T + \rho \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) \right] \tag{A2.2}$$

First, differentiate (A2.1) with respect to y_A^G to get

$$\begin{aligned}
0 < \frac{\partial \lambda_A^G}{\partial y_A^G} &= \frac{'(\lambda_A^G)}{'(\lambda_A^G) \square y_A^G \square (\lambda_V^G)y_V^G + 2 \frac{v\mu\lambda_A^G}{\eta} \rho_1 \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) + \frac{v\mu\lambda_A^G}{\eta} \rho_{11} \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right)} \\
&< \frac{'(\lambda_A^G)}{E_A^T + \left[\rho \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) + \frac{v\mu\lambda_A^G}{\eta} \rho_1 \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) \right]} \tag{A2.3}
\end{aligned}$$

The second inequality follows from the fact that

$$2 \frac{v\mu\lambda_A^G}{\eta}\rho_1\left(\frac{v\mu\lambda_A^G}{\eta}, \mu\right) + \frac{v\mu\lambda_A^G}{\eta}\rho_{11}\left(\frac{v\mu\lambda_A^G}{\eta}, \mu\right) > 0$$

that both ρ_1 and ρ_{11} are strictly positive, and from rewriting the first term in the denominator using (A2.1). Next, differentiate (A2.2) with respect to λ_A^G . We get

$$\begin{aligned} 0 &< \frac{\partial w_A^G}{\partial \lambda_A^G} = \frac{E_A^T + \rho\left(\frac{v\mu\lambda_A^G}{\eta}, \mu\right) + \frac{v\mu\lambda_A^G}{\eta}\rho_1\left(\frac{v\mu\lambda_A^G}{\eta}, \mu\right) \square '(\lambda_A^G)w_A^G}{(\lambda_A^G)} \\ &= \frac{\left[E_A^T + \rho\left(\frac{v\mu\lambda_A^G}{\eta}, \mu\right) + \frac{v\mu\lambda_A^G}{\eta}\rho_1\left(\frac{v\mu\lambda_A^G}{\eta}, \mu\right)\right] \left(1 \square \frac{\lambda_A^G '(\lambda_A^G)}{(\lambda_A^G)}\right)}{(\lambda_A^G)} \\ &\quad + \frac{\frac{v\mu\lambda_A^G}{\eta}\rho_1\left(\frac{v\mu\lambda_A^G}{\eta}, \mu\right) \frac{\lambda_A^G '(\lambda_A^G)}{(\lambda_A^G)}}{(\lambda_A^G)} \\ &< \frac{\left[E_A^T + \rho\left(\frac{v\mu\lambda_A^G}{\eta}, \mu\right) + \frac{v\mu\lambda_A^G}{\eta}\rho_1\left(\frac{v\mu\lambda_A^G}{\eta}, \mu\right)\right] \left(1 \square \frac{\lambda_A^G '(\lambda_A^G)}{(\lambda_A^G)}\right)}{(\lambda_A^G)} \end{aligned}$$

Hence

$$\begin{aligned} \frac{\partial w_A^G}{\partial y_A^G} &< \frac{'(\lambda_A^G)}{E_A^T + \left[\rho\left(\frac{v\mu\lambda_A^G}{\eta}, \mu\right) + \frac{v\mu\lambda_A^G}{\eta}\rho_1\left(\frac{v\mu\lambda_A^G}{\eta}, \mu\right)\right]} \\ &\quad \times \frac{\left[E_A^T + \rho\left(\frac{v\mu\lambda_A^G}{\eta}, \mu\right) + \frac{v\mu\lambda_A^G}{\eta}\rho_1\left(\frac{v\mu\lambda_A^G}{\eta}, \mu\right)\right] \left(1 \square \frac{\lambda_A^G '(\lambda_A^G)}{(\lambda_A^G)}\right)}{(\lambda_A^G)} \\ &= \frac{'(\lambda_A^G) \left(1 \square \frac{\lambda_A^G '(\lambda_A^G)}{(\lambda_A^G)}\right)}{(\lambda_A^G)} < 1 \end{aligned}$$

Thus we conclude

$$\frac{\partial(y_A^G \square w_A^G)}{\partial y_A^G} > 0$$

QED

A.3

Subtract (19) from (16) to get:

$$'(\lambda_A^G) \square y_A^G \square (\lambda_V^G)y_V^G \square \left['(\lambda_A^T)(1 \square (\lambda_V^T))y_A^T \right] + E_A^G \square E_A^T$$

$$\begin{aligned}
&= \left[\rho \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) + \frac{v\mu\lambda_A^G}{\eta} \rho_1 \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) \right] \\
\Box & \left[\rho \left(\frac{v(1 \Box \mu)\lambda_A^T}{\eta}, 1 \Box \mu \right) + \frac{v(1 \Box \mu)\lambda_A^T}{\eta} \rho_1 \left(\frac{v\mu\lambda_A^T}{\eta}, 1 \Box \mu \right) \right] \tag{A3.1}
\end{aligned}$$

Then use (13) to substitute for $E_A^G \Box E_A^T$ and get:

$$\begin{aligned}
& '(\lambda_A^G) \Box y_A^G \Box (\lambda_V^G) y_V^G \Box ['(\lambda_A^T)(1 \Box (\lambda_V^T)) y_A^T] + \rho \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) \\
&= \left[\rho \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) + \frac{v\mu\lambda_A^G}{\eta} \rho_1 \left(\frac{v\mu\lambda_A^G}{\eta}, \mu \right) \right] \\
\Box & \left[\rho \left(\frac{v(1 \Box \mu)\lambda_A^T}{\eta}, 1 \Box \mu \right) + \frac{v(1 \Box \mu)\lambda_A^T}{\eta} \rho_1 \left(\frac{v(1 \Box \mu)\lambda_A^T}{\eta}, 1 \Box \mu \right) \right] \tag{A3.2}
\end{aligned}$$

which can be re-written as (22) in the text.

A.4

Consider first the solution of the system of two equations in the symmetric equilibrium. Start with defining $z \equiv v\lambda_1 < 1$, so that $v = \frac{z}{\lambda_1}$. Also define $2\Delta = y_1 \Box y_2$. Equation (29) and (30) can then be re-written as follows

$$v = \frac{z}{\log \left[\frac{4\Delta z^2(1 \Box z)^2}{(2z \Box 1)(z^2 \Box z + 1)} \right]} = v(z, \Delta,) \tag{A4.1}$$

$$\lambda_1 = \log \left[\frac{2z^2(1 \Box z)^2 \Delta}{(2z \Box 1)(z^2 \Box z + 1)} \right] \equiv \lambda(z, \Delta,) \tag{A4.2}$$

Notice the following:

$$\lim_{z \rightarrow \frac{1}{2}} v(z, \Delta,) = 0$$

There also exists a value \bar{z}_1 ($\frac{1}{2} < \bar{z}_1 < 1$) such that

$$\frac{2\bar{z}_1^2(1 \Box \bar{z}_1)^2 \Delta}{(2\bar{z}_1 \Box 1)(\bar{z}_1^2 \Box \bar{z}_1 + 1)} = 1 \tag{A4.3}$$

and $\lambda(\bar{z}_1, \Delta,) \rightarrow \infty$.

There exists another cut-off point $\bar{z}_2 > 1$, such that $v(\bar{z}_2, \Delta,) \rightarrow \infty$, but it lies outside the feasible interval $(\frac{1}{2}, 1)$.

Thus we need:

$$\frac{1}{2} < z \leq \bar{z}_1 \quad (\text{A4.4})$$

We have:

1. $\frac{\partial \lambda_1}{\partial z} < 0$ for $\frac{1}{2} < z < \bar{z}_1$
2. $\frac{\partial \lambda_1}{\partial z} > 0$ for $z > \bar{z}_2$

Inserting $\lambda(z, \Delta, \cdot)$ in (29) we get:

$$\begin{aligned} & 1 + \Delta \square 2\Delta \left(1 + \frac{z}{v(z, \Delta, \cdot)}\right) \exp \left[\square \frac{z}{v(z, \Delta, \cdot)} \right] \\ & \square \left(1 + \frac{1}{v(z, \Delta, \cdot)}\right) \exp \left[\square \frac{1}{v(z, \Delta, \cdot)} \right] (1 \square \Delta) \\ & + \frac{[2z^2 \square 2z + 1]^2}{2\lambda(z, \Delta, \cdot)z^2(1 \square z)^2} = c \end{aligned} \quad (\text{A4.5})$$

which can be solved for the remaining unknown, z .

For existence, it suffices to notice that the left hand side of (A4.5) goes to ∞ as $z \rightarrow \frac{1}{2}$ and is zero as $z_1 \rightarrow \bar{z}_1$. For unicity, compute the first derivative of (A4.5) in a piecewise fashion. Each piece is a line of (A4.5).

- First piece (1st):

$$\frac{\partial}{\partial z}(1^{st}) = \square \frac{\lambda_1(2z^4 \square 4z^3 + 9z^2 \square 7z + 2)}{2z^3(1 \square z)^3}$$

where $2z^4 \square 4z^3 + 9z^2 \square 7z + 2 > 0$ is always positive in $[1/2, 1]$ and therefore $\frac{\partial}{\partial z}(1^{st}) < 0$.

- Second piece:

$$\frac{\partial}{\partial z}(2^{nd}) = \frac{y_2 \exp(\square \frac{\lambda_1}{z}) \lambda_1^2 (2z^4 \square 5z^3 + 6z^2 \square 4z + 1)}{z^3(1 \square z)(z^2 \square z + 1)(2z \square 1)} + \frac{y_2 \exp(\square \frac{\lambda_1}{z}) \lambda_1 (\square 2z^4 + 4z^3 \square 9z^2 + 7z \square 2)}{z^3(1 \square z)(z^2 \square z + 1)(2z \square 1)}$$

which can be rewritten as:

$$\frac{\partial}{\partial z}(2^{nd}) = \frac{y_2 \exp(\square \frac{\lambda_1}{z}) \lambda_1 [(2z^4 \square 5z^3 + 6z^2 \square 4z + 1) \lambda_1 \square (2z^4 \square 4z^3 + 9z^2 \square 7z + 2)]}{z^3(1 \square z)(z^2 \square z + 1)(2z \square 1)}$$

Define

$$A = (2z^4 \square 5z^3 + 6z^2 \square 4z + 1) \quad \text{and} \quad B = (2z^4 \square 4z^3 + 9z^2 \square 7z + 2)$$

Note that in the interval $[\frac{1}{2}, 1]$ we have that $A < 0$ and $B > 0$. Next consider the component

$$A\lambda_1 \square B$$

and substitute out

$$\lambda_1 = \log \left[\frac{4\Delta z^2(1 \square z)^2}{(2z \square 1)(z^2 \square z + 1)} \right] = \log \left[\frac{4z^2(1 \square z)^2}{(2z \square 1)(z^2 \square z + 1)} \right] + \log \left(\frac{\Delta}{\square} \right)$$

Define

$$C = (2z^4 \square 5z^3 + 6z^2 \square 4z + 1) \log \left[\frac{4z^2(1 \square z)^2}{(2z \square 1)(z^2 \square z + 1)} \right] \square (2z^4 \square 4z^3 + 9z^2 \square 7z + 2)$$

which is an expression only in z . We have that $C < 0$ in $[\frac{1}{2}, 1]$. Then

$$A\lambda_1 \square B = C + A \log \left(\frac{\Delta}{\square} \right)$$

which is negative provided that

$$\frac{\Delta}{\square} > \sim 0.003$$

If such condition is verified, then the second piece is globally negative.

- Third piece:

$$\begin{aligned} \frac{\partial}{\partial z}(3^{rd}) &= \frac{\lambda_1(2z^2 \square 2z + 1)(4z^6 \square 14z^5 + 32z^4 \square 41z^3 + 34z^2 \square 16z + 3)}{2z^4(1 \square z)^3(z^2 \square z \square 1)(2z \square 1)} + \\ &+ \frac{(2z^2 \square 2z + 1)(\square 4z^6 + 12z^5 \square 28z^4 + 36z^3 \square 27z^2 + 11z \square 2)}{2z^4(1 \square z)^3(z^2 \square z \square 1)(2z \square 1)} \end{aligned}$$

Define

$$D = (4z^6 \square 14z^5 + 32z^4 \square 41z^3 + 34z^2 \square 16z + 3) \quad \text{and} \quad E = (\square 4z^6 + 12z^5 \square 28z^4 + 36z^3 \square 27z^2 + 11z \square 2)$$

Now only E is always negative. Hence

$$\frac{\partial}{\partial z}(3^{rd}) = \frac{(2z^2 \square 2z + 1)}{2z^4(1 \square z)^3(z^2 \square z \square 1)(2z \square 1)} [D\lambda_1 + E]$$

Now substitute out for λ_1 as above and define

$$F = D \log \left[\frac{4z^2(1 - z)^2}{(2z - 1)(z^2 - z + 1)} \right] + E$$

which depend only on z and it is always negative. Finally

$$[D\lambda_1 + E] = F + D \log \left(\frac{\Delta}{-} \right) < 0$$

over the entire interval $[\frac{1}{2}, 1]$ provided that $\Delta < \sim 10.65$. Summing up, there exist a unique solution provided that

$$0.003 < \frac{\Delta}{-} < 10.65$$

QED

A.5

Optimal policy under symmetry is characterized as follows

$$\begin{aligned} \mathcal{L} \equiv & v [(\lambda_1) + (\lambda_2)(1 - (\lambda_1))y_2] \\ & + \left[\frac{1}{2} \log(2\lambda_1) + \frac{1}{2} \log(2\lambda_2) + \log(v) \right] \\ & - cv + \theta(1 - v\lambda_1 - v\lambda_2) \end{aligned}$$

The first order conditions are:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \lambda_1} &= v '(\lambda_1)(1 - (\lambda_2)y_2) + \frac{1}{2\lambda_1} - \theta v = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda_2} &= v '(\lambda_2)(1 - (\lambda_1))y_2 + \frac{1}{2\lambda_2} - \theta v = 0 \\ \frac{\partial \mathcal{L}}{\partial v} &= -(\lambda_1) - (\lambda_2)(1 - (\lambda_1))y_2 + \frac{1}{v} - c - \theta(\lambda_1 + \lambda_2) = 0 \end{aligned}$$

Using the first and the second equation and eliminating θv , we get

$$\exp[-\lambda_1] (y_1 - y_2) = \frac{1}{2v} \left[\frac{1}{\lambda_2} - \frac{1}{\lambda_1} \right] \tag{A5.1}$$

The third equation is rewritten as

$$\begin{aligned}
& (\lambda_1) + (\lambda_2)(1 - (\lambda_1))y_2 + \frac{c}{v} \\
&= \theta(\lambda_1 + \lambda_2) + c \\
&= (\lambda_1 + \lambda_2) \left[\theta'(\lambda_2)(1 - (\lambda_1))y_2 + \frac{c}{2v\lambda_2} \right] + c
\end{aligned}$$

where the equality follows from the first equation.

Rearranging terms, we get

$$\begin{aligned}
& y_1 - g(\lambda_1)(y_1 - y_2) - g(\lambda_1 + \lambda_2)y_2 \\
&= \frac{c}{v} \left[1 + \frac{\lambda_1 + \lambda_2}{2\lambda_2} \right] - \lambda_1 \exp[-\lambda_1] (y_1 - y_2) + c \\
&= \frac{c}{2v\lambda_2} [\lambda_1 - \lambda_2] - \lambda_1 \frac{c}{2v} \left[\frac{1}{\lambda_2} - \frac{1}{\lambda_1} \right] + c \\
&= c
\end{aligned}$$

Hence we have

$$\exp[-\lambda_1] (y_1 - y_2) = \frac{c}{2v} \left[\frac{1}{\lambda_2} - \frac{1}{\lambda_1} \right] \quad (\text{A5.2})$$

$$y_1 - g(\lambda_1)(y_1 - y_2) - g(\lambda_1 + \lambda_2)y_2 = c \quad (\text{A5.3})$$

Using the adding up constraint, we obtain:

$$\exp[-\lambda_1] (y_1 - y_2) = \frac{(2v\lambda_1 - 1)c}{2v\lambda_1(1 - v\lambda_1)} \quad (\text{A5.4})$$

$$y_1 - g(\lambda_1)(y_1 - y_2) - g\left(\frac{1}{v}\right)y_2 = c \quad (\text{A5.5})$$

One thing worth noticing is that the first order condition for vacancy creation does not depend on v . The impact of v on optimal v is only through the first order condition on the length of queues. Equation A5.4 can be interpreted naturally as the equality between the marginal benefit and the marginal cost of adjusting the queue length. The marginal benefit from an increase in λ_1 in this symmetric setting is simply the marginal increase in output; a small increase in λ_1 raises the probability that a better matching occurs and reduces by the same amount the probability that the less effective match occurs. This marginal change in probability is simply given by

$$\frac{\partial \exp[-\lambda_1]}{\partial \lambda_1} = -\exp[-\lambda_1]$$

and the net increase in output is simply $(y_2 \square y_1)$. The cost is an increase in adjustment cost, as given by the RHS, which can be written as

$$\left[\frac{p \square \frac{1}{2}}{p(1 \square p)} \right]$$

Equation A5.5 is also an equality between the marginal cost and benefit of vacancy creation. The marginal cost is c . The LHS is the benefit and can be interpreted as follows. An increase in vacancy creation increases the number of job slots but at the same time reduces proportionately the length of queues λ_1 and λ_2 . It is a simple consequence of the envelope theorem that optimal vacancy creation does not depend directly on .

A.6

Proof of Proposition 4

First, λ_1 is strictly decreasing in $z \equiv v\lambda_1$ both in (30') and (32). Moreover, since

$$\frac{(2z \square 1)(z^2 \square z + 1)}{2z^2(1 \square z)^2} > \frac{(2z \square 1)}{2z(1 \square z)}$$

the schedule described by (32) lies strictly above the schedule described by (30'): i.e., given z , λ_1 in (30') is always strictly smaller than λ_1 in (32). While the schedule (31) is also upward sloping, the schedule (29') is hump - shaped. Moreover, since

$$\frac{\lambda_1 [2z^2 \square 2z + 1]^2}{2z^3(1 \square z)^2} > 0$$

eq. (31) lies always above (29'). Hence it follows that

$$\lambda_1^* > \lambda_1^\dagger$$

and

$$v^* < v^\dagger$$

QED

A.7

Proof of Proposition 6

Insert in the symmetric matching equilibrium equations the socially optimal solution (Λ_1, V) , to obtain, for the $A \square G$ match

$$'(\lambda_1)(y_1 \square y_2) = \frac{[2\mu V \Lambda_1 \rho_1 (2\mu V \Lambda_1, \mu)]}{(1+s)} \quad (\text{A7.1})$$

$$\square \frac{[2\mu V \Lambda_1 \rho_1 (2\mu V \Lambda_1, \mu)]}{(1+s)} [\rho (1 \square 2\mu V \Lambda_1, 1 \square \mu) + (1 \square 2\mu V \Lambda_1) \rho_1 (1 \square 2\mu V \Lambda_1, 1 \square \mu)] \quad (39)$$

Notice that

$$1 + s = \frac{\square(\Lambda_1, V)}{(\Lambda_1, V)}$$

so that we have

$$\frac{(\Lambda_1, V)}{(1+s)} = \frac{(\Lambda_1, V)}{\square(\Lambda_1, V)}$$

Inserting this expression on the right hand side of (A7.1), we obtain

$$\begin{aligned} & \frac{[2\mu V \Lambda_1 \rho_1 (2\mu V \Lambda_1, \mu)]}{(1+s)} \\ & \square \frac{[2\mu V \Lambda_1 \rho_1 (2\mu V \Lambda_1, \mu)]}{(1+s^A)} [\rho (1 \square 2\mu V \Lambda_1, 1 \square \mu) + (1 \square 2\mu V \Lambda_1) \rho_1 (1 \square 2\mu V \Lambda_1, 1 \square \mu)] \\ & = \square(\Lambda_1, V) \frac{(\Lambda_1, V)}{\square(\Lambda_1, V)} = (\Lambda_1, V) \end{aligned}$$

Turning to the tax on vacancy creation, this is such that, for job type G

$$\begin{aligned} \pi_1 &= y_1 \square g(\Lambda_1)(y_1 \square y_2) \square g(\Lambda_1 + \Lambda_2) y_2 + \Theta \square t \\ &= y_1 \square g(\Lambda_1)(y_1 \square y_2) \square g(\Lambda_1 + \Lambda_2) y_2 + \Theta \square \Theta + \Psi \\ &= y_1 \square g(\Lambda_1)(y_1 \square y_2) \square g(\Lambda_1 + \Lambda_2) y_2 = 0 \end{aligned}$$

QED