Why is the Timing of School Tracking so Heterogeneous?

by

Kenn Ariga (Kyoto Institute of Economic Research)

Giorgio Brunello (Padova University, CESifo and IZA)

Roki Iwahashi (University of the Ryukyus, Okinawa)

Lorenzo Rocco (Padova University)

Abstract

Secondary schools in the developed world differ in the degree of differentiation and in the first age of selection of pupils into different tracks. In this paper, we account for the heterogeneity of tracking time with a simple stochastic model which conjugates the returns from specialization with the costs of early selection. We calibrate the model for 20 countries – including most of Europe, the US and Japan - and show that the model performs rather well in replicating the observed heterogeneity, with the remarkable exception of Germany.

- JEL: H52, H73
- Keywords: tracking, secondary schools

Introduction

Secondary school design varies significantly across OECD countries. As argued in the 2003 PISA Report (OECD, 2003), some countries have non - selective school systems, that try to provide all students with similar learning opportunities by requiring that each school caters for the full range of student performance. Other countries address diversity explicitly by forming groups of students through selection either between schools (as in Germany or Japan) or between classes within schools (as in the US), with the aim of serving students according to their academic potential and interests.

Table 1 highlights the heterogeneity of secondary school design in a sample of OECD countries, and reproduces three indicators developed by the OECD: the number of school types or distinct education programs available to 15 year olds (A); the percentage of 15 year-olds enrolled in programs that give access to vocational studies at the next program level, or direct access to the labor market (B); and the first age of selection in the education system (C).

The aspect of tracking or streaming we focus on in this paper is the age at which decisions about different school types are made. In some countries - Germany for instance - decisions are taken early on, and they usually imply the allocation of pupils to vocational and general tracks. In other countries - the US or the UK for instance - no formal differentiation takes place until completion of secondary high school at age 16. Early trackers provide students and parents with a range of options, and late trackers offer a single comprehensive school. In the former group of countries, a consistent percentage of students either continue after graduation with additional vocational studies or enter the labor market.

Secondary school design involves the decision to sort pupils into tracks, the number of tracks and the timing of selection into each track. Design has changed over time. After the war, the majority of European countries reformed their secondary education by delaying the moment of specialization into tracks and lines. This widespread comprehensive school reform has taken place at different times and with different degrees across Europe: in the 1960s in Sweden, the UK, Italy, and Norway, in the 1970s in Finland, in the 1980s in France and in the early 1990s in Spain and Portugal. Even the countries which retain a system with early stratification have taken

steps to delay specialization: Germany has introduced a common orientation phase in each of the three tracks, and Belgium and the Netherlands have introduced a common curriculum in the first stages of secondary education[1]. 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[2].

Table 1. Indicators of school tracking

	Number of school types	% enrolled in programs that give access to vocational studies, or direct access to labor market	First age of selection
Austria	4	42.9	10
Australia	1	8.9	16
Belgium	4	22.8	12
Canada	1	-	13
Czech Republic	5	16.9	11
Germany	4	-	10
Denmark	1	0	16
Spain	1	0	16
Finland	1	0	16
France	-	9.5	15
Greece	2	19.9	15
United Kingdom	1	-	16
Hungary	3	19.6	11
Ireland	4	17.8	15
Italy	3	26.7	14
Japan	3	25.4	15
Netherlands	4	61.3	12
Portugal	3	8.8	15
Canada	1	-	13
Korea	3	26.7	14
United States	1	0	16

Note: see OECD (2004)

---

[1] See Green, Wolf and Leney, 1999, and Meghir and Palme, 2004, for a detailed discussion.
[2] We have changed in the table the number of school types for Japan from 2 to 3. Similarly to Korea, Japan has the so called *kosen* schools, vocational colleges lasting 5 years after the end of junior high schools.

Why do countries choose different secondary school designs? In principle, differentiation can improve teacher effectiveness, because teachers can target instruction at the level more closely aligned with student needs than is possible in more heterogeneous environments (see Figlio and Page, 2000). If this is so, the next question is why do some countries track later than others, and by so doing lose the potential gains from specialization. In a different paper - see Brunello, Giannini and Ariga, 2005 - we have shown that a countervailing force is that selection of pupils into tracks is far from noiseless: the earlier the tracking occurs, the easier it is to confound ability with maturity, and the most likely that pupils get misallocated. While several countries allow misallocated students to change tracks, in practice this is often difficult, time consuming, and costly (see Dustmann, 2004, for Germany).

When selection is based on measured criteria - such as the results in standardized tests - individual choice of alternative tracks is constrained by the outcome of these tests. The degree of involvement of teachers and parents in the selection process varies by countries, but perceived ability often remains a key factor[3]. The presence of noise implies that selection contains an element of risk, and that the relative importance of this risk increases when selection is anticipated. We posit that the trade-off between specialization and noise is an important factor in the choice of the first age of selection: by anticipating too much, the government increases the costs of misallocating pupils, whose specific abilities are not yet well developed; by delaying too much, it loses the advantages of specialization[4].

In this paper, we account for the observed heterogeneity of secondary school design in OECD economies by building a simple model, which uses the illustrated trade-off to determine two key variables: the first age of selection after primary school and the relative "size" of general versus vocational jobs. The model has three actors: the government, which in each country chooses the first age of selection; the parents, who care about their offspring's education and can vote on the relative size of each type of schools; and firms, which translate the relative supply of graduates into

---

[3]Checchi and Flabbi, 2005, illustrate the differences in selection in Germany and Italy.
[4]An alternative story is that school differentiation "..has served the purpose of perpetuating the pre-existing social order through a process of exclusion of the emerging middle classes from the more prestigious, academically oriented institutions". (Bertocchi and Spagat, 2003)

general and vocational jobs, with their attached wages[5].

When the government is the repository of all the decisions relevant to school design, it can elect to maximize either productive efficiency – total output net of schooling costs – or welfare. When households are risk averse, the two need not coincide. An alternative, and we believe more realistic characterization of school design, is that parents have some voice in the decision process. Typically, the relative size of each type of school can be affected by the parental demand for general and vocational education. We model parental involvement as follows: the government moves first and sets the tracking time, by taking into account parental preferences on the number of slots in general and vocational schools. These preferences are mediated in a democratic process by majority voting. We label this sequential process as "constrained welfare maximization".

We start by discussing productive efficiency and welfare maximization as useful benchmarks. Then we move to constrained welfare maximization. Since, in spite of its simplicity, the model cannot be solved analytically, we turn to numerical solutions and proceed in two steps: first, we calibrate the model for each country and compare attained with real outcomes. Second, we impose that each government chooses the observed first age of selection and ask which voter in the distribution of parents (and pupils) would select the relative size of tracks – and the selection threshold - consistent with the observed selection age. Since this "pivotal" parent is not necessarily the median voter in the original distribution of talents, we conclude by arguing that parental attention to school design is not homogeneous across countries, and that this heterogeneity is related to the country specific average educational attainment and GNP per capita.

The paper is organized as follows: we setup our theoretical model in Section 1, and discuss its implications in Section 2. Numerical solutions are discussed in Section 3. Conclusions follow.

---

[5] The basic structure of the model is drawn from Brunello, Ariga and Giannini, 2005, where the focus was on the effects of technical progress and up-skilling on the optimal age of first selection.

1. The Model

Our theoretical model draws from Brunello, Giannini and Ariga, 2005, and focuses on the trade-off between the returns to specialization and the risk associated to early selection. The static economy we consider is populated by parents and pupils. Each household has one parent and one pupil. Pupils differ in their true ability A, which is log-normally distributed. Therefore, log ability α is distributed according to a standard normal. Parents and pupils share the same ability. This is rather extreme but not wholly unrealistic, given the importance of genes in the transmission of ability. Since parents take decisions on behalf of their kids, this assumption guarantees that they decide as if they were the kids themselves.

The frame of the model consists of two periods. In the first period, kids go to school and graduate. In the second period, they work and earn wages. Schools are run by the government, a realistic assumption in the case of secondary education, and the monetary costs required to operate them do not vary with the design of the curriculum and the age of first selection. Given that costs are constant across alternative designs, we simply set them to zero. Since schools cost nothing, there are neither tuition fees nor taxes raised by the government. Therefore, we can completely ignore the public finance aspects of school design[6].

1.1 Schools

The government chooses whether to differentiate schools into different curricula – or tracks – and when to do so during the time spent in primary and secondary schools. Let this time be equal to 1. Then $\tau \in [0,1]$ corresponds to the first age of selection: the lower τ, the lower this age. Differentiation consists of two tracks, H (high) and L (low). In a number of – mainly European - countries, the H school corresponds to general education and the L school to vocational education. In other countries – such as the US - streaming occurs between classes of the same school. When $\tau = 1$ (no differentiation), all schools are comprehensive for the entire period of

---

[6] In a world with taxes, this is equivalent to assuming that taxes are lump – sum.

time. When $\tau < 1,$ all schools are comprehensive for time length τ and are divided in the L and H track for the rest of the time. By assumption, there is no further stratification within each type of school[7]. Independently of design, each school combines individual ability with the effectiveness of teaching to produce human capital. There are no peer effects[8]. When schools are fully comprehensive and there are no tracks, the effectiveness of teaching is hampered by the heterogeneity of talent. Tracking improves effectiveness by grouping individuals by ability. This assumption is often made by those proposing ability grouping, who argue that by narrowing the range of student abilities within a classroom, tracking allows teachers to target instruction at the level more closely aligned with student needs than is possible in more heterogeneous environments[9].

Let the effectiveness of teaching in comprehensive and tracked schools be respectively equal to 1 and to $R > 1$. When students spend an initial proportion τ of their time at school in mixed ability classes and the complementary proportion $(1-\tau)$ in stratified schools composed of two tracks, their individual human capital at the end of the schooling period is

$$H = A^{\tau}\left(RA\right)^{1-\tau} \qquad [1]$$

Taking logs

$$h = \alpha + (1-\tau)\ln R = \alpha + (1-\tau)\beta \qquad [2]$$

The allocation of students to each track is based upon observed ability, which differs from true ability because of a random noise. Let observed log ability θ when selection takes place be related to true log ability by

$$\theta = \alpha + \varepsilon \qquad [3]$$

---

[7]These assumptions follow Brunello, Giannini and Ariga, 2005.
[8] Peer effects play a relevant role in Brunello, Giannini and Ariga, 2005.
[9] The empirical evidence in support of this view is mixed at best. See Gamoran and Mare, 1989, for a positive view, Woessmann and Hanushek, 2005, for a more skeptical view, and Figlio and Page, 2000 for an exhaustive review of the empirical literature.

where ε is an exogenous shock independent of α and normally distributed with mean zero and variance b^2. We capture the idea that the noise of selection increases the earlier selection takes place by letting

$$b = \mu(1-\tau) \tag{4}$$

where μ is a suitable parameter. Conditional on the age of first selection, a higher value of this parameter suggests that selection relies to a lesser extent on true ability. This could happen either because the testing procedure is noisy or because of the presence of other factors which affect outcomes independently of ability. An example is parental influence and networking. It follows that observed ability θ is also normally distributed with zero mean and variance $\sigma^2 = 1 + b^2$. Denote with Φ and ϕ the distribution and density associated to the random variable θ.

Selection is such that individuals with observed ability higher than the threshold value θ^* are allocated to H schools, and the rest is allocated to L schools. At the end of school, graduates enter the labor market and are hired by firms, which observe the school type (the same type if schools are fully comprehensive, H or L type if schools are divided into tracks at some point in time), and infer ability from the observed type[10]. Suppose that the graduate has spent all her education in a comprehensive school ($\tau = 1$). In this case her expected human capital is

$$Eh = E(\alpha) = 0 \tag{5}$$

If instead the graduate has spent part of her time in a comprehensive school and part in an H track, her expected human capital is[11]

---

[10] This assumption is plausible in the static context of this model, and removes the complications associated to labor market screening.

[11] $E(h_H \mid \theta \geq \theta^*) = (1-\tau)\beta + E(\alpha \mid \theta \geq \theta^*) = (1-\tau)\beta + \dfrac{1}{\sigma^2} E(\theta \mid \theta \geq \theta^*)$

$$Eh_H = E(h_H \mid \theta \geq \theta^*) = (1-\tau)\beta + \frac{1}{\sigma}\frac{\phi(\frac{\theta^*}{\sigma})}{1-\Phi(\frac{\theta^*}{\sigma})} \qquad [6]$$

because ability is time invariant and firms know that the graduate must have measured ability higher than θ^* to qualify for the H track. Similarly, for graduates of L tracks we have

$$Eh_L = (1-\tau)\beta - \frac{1}{\sigma}\frac{\phi(\frac{\theta}{\sigma})}{\Phi(\frac{\theta}{\sigma})} \qquad [7]$$

Notice that, when $\beta = 0$, average human capital in the population is zero, as in the case of comprehensive system. In this case, we should have no tracking, because there is no return from specialization in spite of the positive cost associated to noisy selection.

A delay in the tracking time has two contrasting effects on individual human capital: on the one hand, it reduces the expected human capital of both ability groups, because students spend less time together and have fewer opportunities to enjoy the higher effectiveness of teaching. On the other hand, later tracking reduces the noise in the selection process, which positively affects human capital by reducing the variance σ^2.

1.2 Individuals

Pupils and parents are risk averse and care about the expected utility of schooling. Let individual utility be a function of log wages, with no effort to complete schooling, and let $1-\Phi_i = \text{Prob}\left\{\varepsilon_i \geq \frac{\theta^* - \alpha_i}{b}\right\}$ be the probability that an individual with ability α_i be selected for the H school. Finally, let each individual know his own true ability. Then the expected utility from schooling is

$$EU_i = \Pr ob\left\{\varepsilon_i \geq \frac{\theta^* - \alpha_i}{b}\right\}w_G + \Pr ob\left\{\varepsilon_i < \frac{\theta^* - \alpha_i}{b}\right\}w_V \qquad [8]$$

where w_G and w_V are the log wages paid to graduates of H and L schools. Since parents have the same ability of their offspring and care about the welfare of the latter, expression (8) is also their expected utility, minus the non-pecuniary net benefits that parents are assumed to obtain from the direct involvement in the democratic decisions concerning the school of their offspring.

Suppose that $w_G > w_V$. Then, if access to H and L schools were free, all individuals would enroll in track H, unless the effort costs of completing the curriculum in that track are high enough to discourage part of the student population. With no individual effort in the model, we assume that the allocation of students to tracks is not based on free choice but on a noisy ability test: performance in the test higher than or equal to the required standard θ^* qualifies the candidate for the higher - ability track, and lower performance implies assignment to the lower - ability track. In practice, selection by test needs not be an entry exam, but can be based on the quality of the leaving certificate from the previous school, on orientation and evaluation by teachers and on selection during the first year after entry.

1.3 Firms

Firms can only observe the school the individual has graduated from. Since true individual ability cannot be observed, each graduate is paid her expected productivity, conditional on the type of school. Each firm has two types of job or task, G and V. When schools are comprehensive, graduates are equally productive in either job. With tracking, however, firms always prefer to allocate H graduates to G jobs and L graduates to V jobs. The production technology is Cobb Douglas and log output is given by

$$y = a + \lambda(n_G + Eh_H) + (1 - \lambda)(n_V + Eh_L) \qquad [9]$$

There is no mismatch, in the following sense: there are no frictions, wages in the labor market are flexible enough to allocate workers to jobs, and the government has enough information to match the relative supply of graduates to relative demand. Removing one of these assumptions would introduce mismatch, and the risks associated to it. Given the long run nature of our model, we are implicitly assuming that temporary mismatches dissipate over time. Therefore, the only element of risk we focus on in this paper takes place before labor market entry, and is associated to the selection of the school track.

Profit maximization yields

$$w_G = \ln \lambda + y - n_G \; ; \; w_V = \ln(1-\lambda) + y - n_V \tag{10}$$

Absence of mismatch implies that the number of slots for G - type jobs is exactly equal to the number of graduates from schools of type H. Therefore, $n_G = \ln(1-\Phi)$. Relative wages in this economy satisfy the following condition

$$w_V - w_G = \ln \frac{1-\lambda}{\lambda} + \ln \frac{1-\Phi}{\Phi} = B \tag{11}$$

where B is a convenient notation. Log output can be written as

$$y = a + \lambda \ln(1-\Phi) + (1-\lambda)\ln \Phi + \beta(1-\tau) + \frac{\phi}{\sigma}(\frac{\lambda}{1-\Phi} - \frac{(1-\lambda)}{\Phi}) \tag{12}$$

The case of no tracking corresponds to the equality $\lambda = 1 - \Phi$ and to $\tau = 1$. Since λ is a constant and Φ is a function of θ, we can write

$$1 - \Phi\left(\hat{\theta^*}\right) = \lambda$$

because $\sigma = 1$ when $\tau = 1$. Using this into (12) when $\theta^* = \hat{\theta^*}$, we obtain

$$y_c = a + \lambda \ln \lambda + (1 - \lambda) \ln \lambda \qquad\qquad [13]$$

where the subscript c is for comprehensive schools. As $\tau \to 1$, $\theta^* \to \hat{\theta}^*$ and $y \to y_c$. The intuition is as follows: when tracking occurs at the last instant, it does convey information to firms via its signaling role. On average, however, there is no social value in this information, because $(1 - \Phi)Eh_H + \Phi Eh_L = 0.$

2. Alternative modes of determination of the tracking time and the selection threshold

In this section, we illustrate the implications of the model in two benchmark cases, productive efficiency and welfare maximization, which occur when the government sets the two control variables, τ and θ^*, to maximize either total output or a well defined social welfare function. In this command economy, parents and students have no active role to play. A more plausible assumption is that parents have the opportunity to express their views and preferences about the relative size of H and L schools. In most developed countries, parents are increasingly concerned with the risks associated to vocational education, which is perceived as too specialized, and try to have their offspring admitted to schools with a more general and flexible curriculum. The ensuing academic drift, a pervasive feature of schooling in the last part of the previous century (see Green, Wolf and Leney, 1999), can be interpreted at least in part as the outcome of parental pressure for wider access to this type of education.

Our model illustrates well why parents care about the value of θ^*, the key parameter associated to the relative size of each school type. Consider for instance the parents of students with observed ability below the critical threshold θ^*, and let $w_G > w_V$. A lower value of the threshold has two effects: first, it increases the probability that the offspring of these parents has access to better paid jobs via education in H schools, because it increases the size of these schools and reduces the passing hurdle. Second, it leads to lower wages in either type of job, because it

reduces expected human capital. While the parents of pupils with an observed ability very much below the threshold are likely to prefer a higher threshold, and the associated higher wages, because their offspring is too far from the passing line, the parents of pupils with an observed ability near the threshold may go for a lower passing line, because of the higher probability of access in the H school.

We model parental influence on the relative size of schools by assuming that θ^* is the outcome of a democratic vote over alternatives. If individual preferences are single peaked with respect to the variable of interest, we can apply the median voter theorem, which implies that the preferred policy is the one selected by the parent (and offspring) who occupies the median position in the ability distribution. Parental voice is important to determine the relative size of H and L schools, but is probably less relevant for establishing the first age of selection. Therefore, we assume that the relative size of schools is determined by voting, but that the government chooses the first age of selection τ to maximize the social welfare function subject to the best response of voters. We shall call this characterization of the decision mechanism "constrained welfare maximization".

2.1 Productive efficiency

Productive efficiency requires that the government chooses θ^* and τ to maximize log output. Letting $\tau \in [0,1]$, we have

$$\varphi_1(\theta^*, \tau) = \frac{\partial y}{\partial \theta^*} = -\lambda\Phi + (1-\lambda)(1-\Phi) + \frac{1}{\sigma}\left\{\lambda\Phi\left(\frac{\phi}{1-\Phi} - \frac{\theta^*}{\sigma}\right) + (1-\lambda)(1-\Phi)\left(\frac{\phi}{\Phi} + \frac{\theta^*}{\sigma}\right)\right\} = 0 \qquad [14]$$

where the term in braces is positive because of the properties of the truncated normal (see Maddala, 1983). Therefore an internal solution requires that

$$-\lambda\Phi + (1-\lambda)(1-\Phi) < 0 \qquad\qquad [15]$$

Or

$$\frac{1-\Phi}{\Phi} < \frac{\lambda}{1-\lambda}$$

which implies that $B < 0$. As shown by Brunello, Giannini and Ariga, 2005, this condition must be satisfied to have a finite number of G and V jobs. Therefore, productive efficiency in the presence of tracking requires that the wages paid to workers in G jobs be strictly higher than the wages paid to V jobs[12].

Using (14), the first order condition with respect to τ can be written as

$$\varphi_2(\theta^*, \tau) = \frac{\partial y}{\partial \tau} = \frac{\partial y}{\partial \theta^*} \times \left(-\theta^* \phi \frac{\partial \sigma}{\partial \tau} \frac{1}{\sigma^2}\right) - \beta - \frac{\partial \sigma}{\partial \tau} \frac{1}{\sigma^2} \left(\frac{\lambda \phi}{1-\Phi} - \frac{(1-\lambda)\phi}{\Phi}\right) = 0 \qquad [16]$$

By making use of the fact that $\frac{\partial y}{\partial \theta^*} = 0$ at the optimum, the above equation can be rearranged to yield

$$\frac{\left[1 + \mu^2 (1-\tau^*)^2\right]^{\frac{3}{2}}}{\mu^2 (1-\tau^*)} = \frac{\phi}{\beta} \left(\frac{\lambda}{1-\Phi} - \frac{(1-\lambda)}{\Phi}\right) \qquad [17]$$

where τ^* is the optimal value of τ. Notice that, when $\beta \to 0$, $\tau \to 1$ and schooling is fully comprehensive.

The combination of threshold and tracking time which maximizes productive efficiency $\left(\tau_p, \theta_p^*\right)$ is obtained by jointly solving equations (14) and (16). It is useful to define the following relationship

$$1 - \Phi\left(\frac{\overline{\theta^*}}{\sigma}\right) = \frac{1}{2} \qquad [18]$$

---

[12] This result is from Brunello, Giannini and Ariga, 2005.

which is satisfied when $\overline{\theta^*} = 0$. Evaluating the first order condition (14) at $\overline{\theta^*}$ yields

$$\frac{\partial y}{\partial \theta^*} = \frac{1}{2} - \lambda + \frac{\overline{\phi}}{\sigma} > 0 \quad \text{if} \quad \lambda < \frac{1}{2} + \frac{\overline{\phi}}{\sigma}$$

Hence, $\theta_p^* > 0$ if λ is small enough.

2.2 Welfare Maximization

The government ignores individual ability, but knows the distribution of true ability α. Suppose that the welfare function is utilitarian. Then

$$W = \int EU_i = \ln \lambda + y - \ln\left(1 - \Phi(\frac{\theta^*}{\sigma})\right) + B \int \Phi(\frac{\theta^* - \alpha_i}{b})\phi(\alpha_i)d\alpha_i \qquad [19]$$

It must be that

$$\int \Phi(\frac{\theta^* - \alpha_i}{b})\phi(\alpha_i)d\alpha_i = E\left\{\Pr ob\left[\alpha_i + \varepsilon_i \leq \theta^*\right]\right\} = \Phi \qquad [20]$$

because in equilibrium the average probability of being enrolled in n L school is equal to the proportion of the population enrolled in that school, Φ. The welfare function can be written more conveniently as

$$W = (1 - \Phi)\ln \lambda + \Phi \ln(1 - \lambda) + (1 - \lambda - \Phi)\ln\frac{\Phi}{1 - \Phi} + \beta(1 - \tau) + \frac{1}{\sigma}\left(\frac{\lambda\phi}{1 - \Phi} - \frac{(1 - \lambda)\phi}{\Phi}\right)$$

The first order condition for θ^* is

$$\frac{\partial W}{\partial \theta^*} = \ln\frac{1-\lambda}{\lambda} + \ln\frac{1-\Phi}{\Phi} + \frac{(1-\lambda-\Phi)}{\Phi(1-\Phi)} + \frac{1}{\sigma\Phi(1-\Phi)}\left[\lambda\Phi\left(\frac{\phi}{1-\Phi}-\frac{\theta^*}{\sigma}\right)+(1-\lambda)(1-\Phi)\left(\frac{\phi}{\Phi}+\frac{\theta^*}{\sigma}\right)\right]=0 \qquad [21]$$

Since the term in brackets is positive, the rest must add up to a positive value. This is the case if condition (15) holds, because $\frac{(1-\lambda-\Phi)}{\Phi(1-\Phi)}$ is also negative. The first order condition for τ is

$$\frac{\partial W}{\partial \tau} = \frac{\partial W}{\partial \theta}\times\left(-\phi\frac{\partial\sigma}{\partial\tau}\frac{\theta}{\sigma}\right)-\beta-\frac{\partial\sigma}{\partial\tau}\frac{1}{\sigma^2}\left(\frac{\lambda\phi}{1-\Phi}-\frac{(1-\lambda)\phi}{\Phi}\right)=0 \qquad [22]$$

Notice that this condition is equivalent to (16) because $\frac{\partial W}{\partial \theta}=0$ at the optimum. However, since (21) is different from (14), the solution to the welfare problem $\left(\theta^*_w,\tau_w\right)$ is different from $\left(\theta^*_p,\tau_p\right)$.

Proposition: *The government which maximizes a utilitarian social welfare function chooses a threshold of the test lower than the one which maximizes productive efficiency ($\theta^*_w < \theta^*_p$). The associated tracking time, however, can be earlier ($\tau_w < \tau_p$) or later ($\tau_w > \tau_p$) than the time which maximizes productive efficiency.*

Proof: using equation (14), we obtain that the first order condition (21) reduces to

$$\frac{\partial W}{\partial \theta^*}\bigg|_{\theta^*=\theta^*_p} = \ln\frac{1-\lambda}{\lambda} + \ln\frac{1-\Phi}{\Phi} = B < 0$$

which proves the first part of the Proposition. This result and equation (16) yield

$$\frac{\partial W}{\partial \tau}\bigg|_{\tau=\tau_p} = \left[\ln\frac{1-\lambda}{\lambda} + \ln\frac{1-\Phi}{\Phi}\right]\left(-\frac{\partial\sigma}{\partial\tau}\frac{\theta^*_p}{\sigma}\right)$$

the sign of which depends on the sign of θ^*_p. Hence, τ_w is lower than τ_p if $\theta^*_p > 0$ and higher if $\theta^*_p < 0$. QED

Since individual utility is concave in earnings, households with ability not much below the selection threshold are willing to pay to reduce the risk of being allocated to the L school and V jobs, where pay is lower. The social planner can reduce this risk by reducing the threshold and by increasing access to H schools. By so doing, however, it also affects the welfare of both low and high ability households, who are far from the initial threshold, because wages decline. The former group of households is far away from the standard and unlikely to end up in H schools, and the latter faces little risk of ending up in L schools. The government can try to compensate the negative effects of a lower threshold by varying the tracking time, but such policy is not beneficial for all. For instance, a delay of tracking increases the wage in G jobs but reduces the wage in V jobs. When the threshold θ^*_p is low and negative, ablest students suffer a relatively higher damage in terms of low wages, and the tracking time increases. Conversely, when the threshold is positive, it is the least able who suffer most, and tracking time declines with respect to productive efficiency.

2.3. Constrained Welfare Maximization

We start the characterization of constrained welfare maximization by discussing the political mechanism for the determination of the selection standard θ^*. One might ask why we do not extend such mechanism to include tracking time. We refrain from doing so for two reasons. The first is realism: while parents may exert pressure on local communities to add a general or a vocational school, and thus alter the number of slots available to either type of education, it seems unlikely that they can pressure the government to change the tracking time. Second, and more technically, individual utility functions are not single peaked with respect to τ. In this case, the voting outcome does not satisfy the transitivity condition, and Arrow's Impossibility Theorem might apply.

In the calibration section, we show that, for the relevant values of the key parameters, expected utility (8) is single peaked in the threshold standard θ^*. Therefore, we propose a decision mechanism which combines both utilitarian welfare maximization by the social planner and individual direct participation. The procedure is

two-stage. First, the government sets the optimal tracking time by maximizing welfare. Second, parents vote over the standard θ^*, given the tracking time. Since the government is a rational decision maker, it anticipates both the voting outcome and the fact that τ influences such outcome.

2.3.1 Voting on the threshold

The voting literature has identified the so-called "voting paradox": given that the probability of being the pivotal voter is small, a rational agent should not have an incentive to vote when voting is costly, unless some non-pecuniary returns are at stake (see Chang, 2004). One such non-pecuniary return is the sense of civic duty (see Sinnott, 2003), which we expect to be higher among better educated people[13]. The view that more educated individuals are less prone to crime and more able to fully participate to democratic life is a standard tenet of human capital theory, which argues that education provides not only private but also social returns. In his review of the recent literature, Hanushek, 2002, states that "..the more educated are more prone to be civically involved, to vote in local and national elections, and to be a better informed and more responsible electorate...(Hanushek, 2002, p.51).

Since ability and education are positively correlated, the higher participation by better educated individuals to social activities such as voting suggests that it is the individuals in the lower tail of the distribution of ability who are more likely to abstain from choice. If that is the case, the median voter in the truncated distribution has a positive value of log ability, $\alpha_P > 0$, and lies to the right of the median voter in the unconditional distribution, who has zero log ability. We shall call the individual with ability α_P the "the pivotal voter".

The equilibrium pair (θ, τ) is determined by backward induction. Consider first the majority voting outcome. Assume that individual utility is single peaked in θ for any τ, and for the relevant configuration of parameters – as we show in the calibration section of the paper. Then, the median voter theorem can be applied, and the majority voting outcome coincides with the alternative most preferred by the

---

[13]Milligan, Moretti and Oreopoulos, 2003, present empirical evidence supporting this view for the US.

voter with ability α_P, denoted $\theta(\alpha_P, \tau)$.

2.3.2 Government Choice of τ

Next, consider the government problem. By anticipating $\theta(\alpha_P, \tau)$, the government sets τ to maximize W. Formally, the problem can be written as

$$\max_{\tau} W(\theta^*, \tau) \quad \text{s.t.} \ \theta^* = \theta^*(\alpha_P, \tau) \tag{23}$$

Unfortunately, this problem does not have a simple analytical solution. Therefore, we characterize the equilibrium outcome by estimating the key parameters and solving the model numerically.

3. Numerical solutions

How well does the model of constrained welfare maximization fare in explaining the observed variation in tracking time among OECD countries? To answer this, we solve expression (23) numerically for the pair $\left(\theta_c^*, \tau_c\right)$, where the subscript c is for constrained maximization, after estimating for each country in our sample the three key parameters λ, β and μ. These estimates are explained below and are reported in Table 2, which highlights the important cross – section variation in the parameters associated to the relative demand for jobs, the return to track specialization and the risk of misallocation to tracks.

3.1 Estimates of the key parameters

With a Cobb Douglas production function, λ is the share of the total wage bill paid out to workers in G jobs. Hence, $\lambda = \dfrac{W_G N_G}{W_G N_G + W_V N_V}$ can be estimated by allocating occupations to G and V jobs. To the former jobs we allocate managers, professionals, technicians and clerks, and to the latter jobs we allocate sales workers,

craft and related trades workers, and plant and machine operators. Elementary occupations and agricultural labor are excluded from the data.

Table 2. Estimates of the key parameters

	λ	μ	β
Austria	.535	.588	.037
Australia	.551	2.685	.056
Belgium	.744	.794	.046
Canada	.603	.994	.056
Czech Republic	.614	.770	.031
Germany	.574	.575	.050
Denmark	.703	2.288	.038
Spain	.543	2.100	.057
Finland	.665	2.383	.041
France	.625	1.346	.046
Greece	.527	1.506	.058
United Kingdom	.671	2.525	.047
Hungary	.626	.621	.049
Ireland	.533	1.456	.054
Italy	.501	1.218	.039
Japan	.561	2.059	.054
Korea	.505	1.342	.058
Netherlands	.777	0.771	.033
Portugal	.520	1.259	.019
United States	.640	2.436	.043

This classification can be justified as follows: first, elementary occupations do not require secondary education; second, white collar jobs often require college education, and this is more likely for individuals who have completed general

secondary education. While the European data are drawn from the 2001 wave of the European Community Household Panel, a comparable dataset which covers most of Europe, the non European data are from national sources. It turns out that the estimated λ is barely over $\frac{1}{2}$ in Italy and Korea and higher than 0.7 in Belgium and The Netherlands. Japan and Germany have similar values, just above $.55$, and so do the US and the UK, with values close to $.65$. The lower value of the share of white collar jobs partly reflect the fact that both Germany and Japan have a higher share of manufacturing labor than the US and the UK.

The parameter β measures the relative effectiveness of tracked schools. Using (6) and (7), we can express expected human capital from either type of track as

$E\bar{h} = \Phi Eh_L + (1-\Phi)Eh_H = \beta(1-\tau)$. In principle, we would like to have a measure of average human capital at the end of secondary school, and to relate such measure to the years spent in stratified schools. Unfortunately, there are no data which cover our broad range of countries. The closest we can get is with PISA 2003, a comparative project which produces comparative data on the performance of 15 year olds on a battery of standardized tests (see OECD, 2004).

For each student included in the data, we compute her average test score ($SCORE$) and regress it on a set of controls, including gender, grade, nationality, age at the time of the test, and family background indicators such as parental education, parental occupation and information on the availability of books at home. We define time in the stratified school $TIMET$ as

$$TIMET = (18 - \text{first age of selection })/12$$

where 18 is the average year of completion of secondary school and 12 is the average number of years spent from primary to upper secondary education. The variable TIMET is equal to 0 in the event of no tracking whatsoever and to 1 in the event of tracking from the earliest possible age. Since β is a measure of school effectiveness, we assume

$$\beta = \gamma + \delta CT$$

where *CT* is the country specific proportion of certified teachers. The underlying idea is that the relative effectiveness of schools depends on the average quality of teachers, as documented by the literature (see Rivkin, Hanushek and Kain, 2005), and that this quality is higher when teachers have been certified by the relevant national authority. Therefore, we estimate

$$\ln SCORE = \eta_1 + \eta_2 X + (\gamma + \delta CT)*TIMET + \varepsilon$$

where X is the vector of individual controls, and adjust standard errors for clustering. We find that $\gamma = -.141$ (standard error: 0.057) and $\delta = .200$ (standard error: 0.067). Evaluated at the sample country average, the country – specific value of β is shown in the last column of Table 2. According to our estimates, school effectiveness is higher in stratified than in comprehensive schools, and the relative benefit of stratification ranges from 1.9 percent per year of school in Portugal to 5.8 percent in Greece and Korea.

In order to estimate the parameter μ, recall that $Var(\theta) = Var(\alpha) + \mu^2 (1-\tau)^2$. We use the dispersion of the test scores in PISA 2003 as a proxy of the variance of observed ability. Notice that the observed scores in the PISA dataset are bounded between 0 and 800, while the variable θ in the model is unbounded. Therefore, we normalize test scores by defining a new variable, *z*, as the logistic transformation $\ln \frac{SCORE}{(800-SCORE)}$. Since true ability α is unobserved, we regress *z* on the variables in vector X as well as on additional controls, such as a dummy indicating whether the student has a computer at home, whether she speaks a foreign language at home and whether she has repeated one or more grades. We take the variance of the residuals *RES* from this regression as a proxy of the variance of observed ability net of true ability $Var(RES) = Var(\theta) - Var(\alpha)$, and compute the implied value of μ as

$$\mu = \sqrt{\frac{Var(RES)}{(1-\tau)^2}} \ .$$

The results are in the second column of Table 2. We notice the wide cross country variation in this parameter, which measure the relative risk of misallocation for any age of first selection into tracks. This risk is particularly low in Germany, Austria and Hungary, and relatively large in Australia, the UK and the US, which suggests that selection by ability takes place more efficiently in the former group of countries at any age of selection. Given the substantially lower risk, it is not surprising that selection occurs relatively earlier in this group than in the latter group.
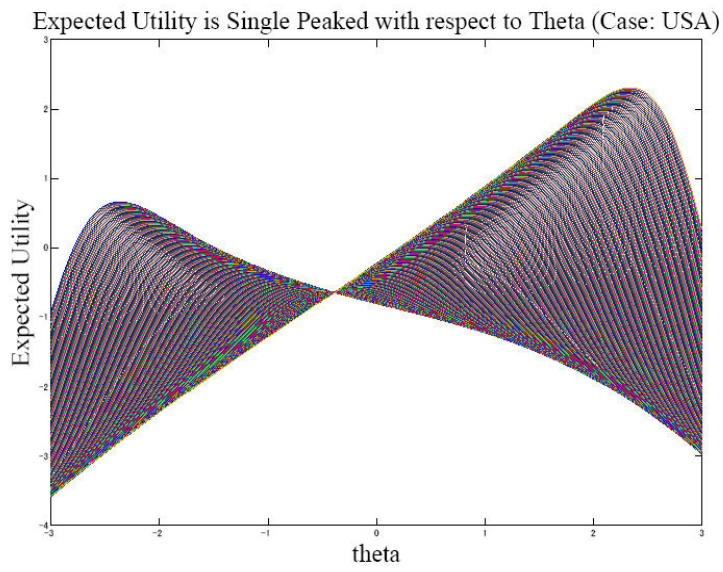
3.2. Solving the model of constrained maximization

Our proposed characterization of the school design problem makes sense only if the individual utility function is single peaked with respect to the selection threshold. Since single peaked-ness cannot be computed analytically, we check whether it holds for the estimated parameters λ, μ and β. Figure 1 plots expected utility EU as a function of θ^* using the estimated parameters fro the United States, and allowing α to vary in the domain $[-3, +3]$. As it turns out, the calibrated functions are always single peaked. Therefore, we can apply for the existing configuration of parameters the median voter theorem.

On the other hand, single peaked-ness does not hold with respect to τ. As shown by Figure 2, if we evaluate expected utility at the US values of the relevant parameters and plot it over the full range of τ after fixing θ^* at the value consistent with constrained welfare maximization, we obtain a non-concave function, with a global maximum in the corner $\tau = 0$.

Next, we use the estimated parameters to compute, for each country, the tracking time and the share of G jobs - which depends on the selected tracking time as well as the threshold θ^* - in the cases of productive efficiency, welfare maximization and constrained welfare maximization.

Figure 1

Expected Utility is Single Peaked with respect to Theta (Case: USA)

Figure 2

Expected utility as a function of tau
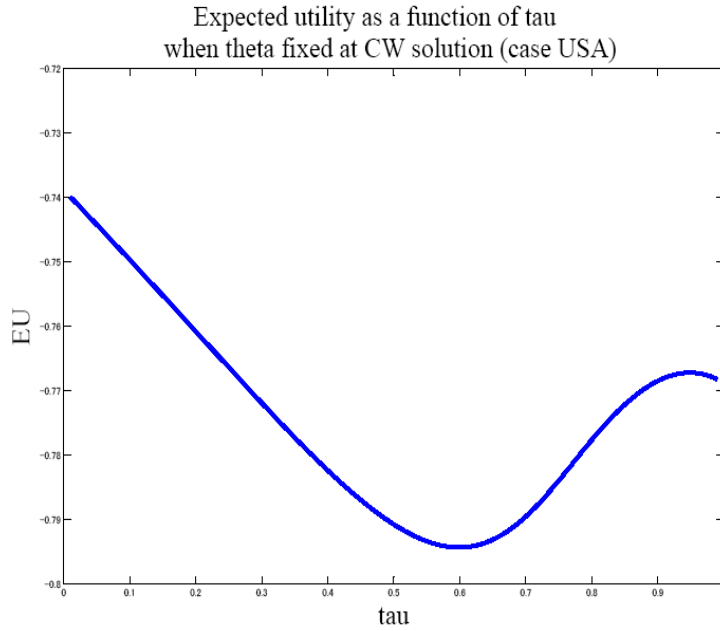when theta fixed at CW solution (case USA)

Table 3 illustrates the results. First, we find that – consistent with the

predictions of the model - welfare maximization yields a higher value of employment in G jobs and a lower value of τ than productive efficiency[14].

Table 3. Calibration outcomes

	τ_p	τ_w	τ_c	τ	$N_{G\,p}$	$N_{G\,w}$	$N_{G\,c}$	N_G
Austria	0.817	0.638	0	0.333	0.192	0.340	0.441	0.436
Belgium	0.889	0.726	0.828	0.500	0.335	0.564	0.573	0.657
Canada	0.909	0.812	0.359	0.583	0.230	0.403	0.489	0.504
Czech Republic	0.918	0.833	0.308	0.416	0.237	0.413	0.502	0.503
Denmark	0.989	0.976	0.994	0.833	0.299	0.510	0.512	0.612
Germany	0.748	0.431	0	0.333	0.216	0.385	0.471	0.465
Spain	0.978	0	0.717	0.833	0.194	0.443	0.438	0.404
Finland	0.989	0.977	0.999	0.833	0.271	0.467	0.500	0.582
France	0.960	0.919	0.977	0.750	0.244	0.424	0.513	0.498
United Kingdom	0.988	0.976	0.999	0.833	0.275	0.473	0.500	0.581
Greece	0.956	0.918	0.560	0.750	0.185	0.327	0.422	0.436
Hungary	0.795	0.530	0	0.416	0.247	0.434	0.507	0.460
Ireland	0.957	0.920	0.560	0.750	0.188	0.332	0.429	0.439
Italy	0.954	0.918	0.456	0.666	0.171	0.303	0.402	0.418
Japan	0.979	0	0.726	0.750	0.204	0.465	0.456	0.481
Korea	0.944	0.897	0.461	0.666	0.174	0.307	0.401	0.423
Netherlands	0.917	0.793	0.735	0.500	0.363	0.604	0.697	0.662
Portugal	0.979	0.963	0.539	0.750	0.181	0.319	0.423	0.331
United States	0.988	0.977	0.841	0.833	0.254	0.439	0.530	0.534
Australia	0.987	0	0.791	0.833	0.198	0.472	0.447	0.587

Second, we notice that constrained maximization does on average a better job in tracking actual values than both productive efficiency and welfare maximization.

---

[14]Employment in G jobs declines with the standard θ^*.

Two important exceptions are Germany and Hungary, two early trackers, where welfare maximization does the best job in predicting the actual age of selection. Define the calibration error as the difference between the actual and calibrated value of the variable of interest. Then an indicator of goodness of fit is the sum of squared errors. In the case of τ, this is equal to 1.77 for productive efficiency, 2.76 for welfare maximization and to 0.93 for constrained welfare maximization. In the case of N_G, these numbers are 1.47, 0.16 and 0.07 respectively.

Third, as shown in Figure 3, the calibrated employment share of G jobs is relatively close to the actual value, possibly with the exception of Portugal and Australia. The simple correlation between calibrated and actual values is 0.812, which suggests that our model does a very good job in mimicking the cross – country ordering of employment in G jobs.
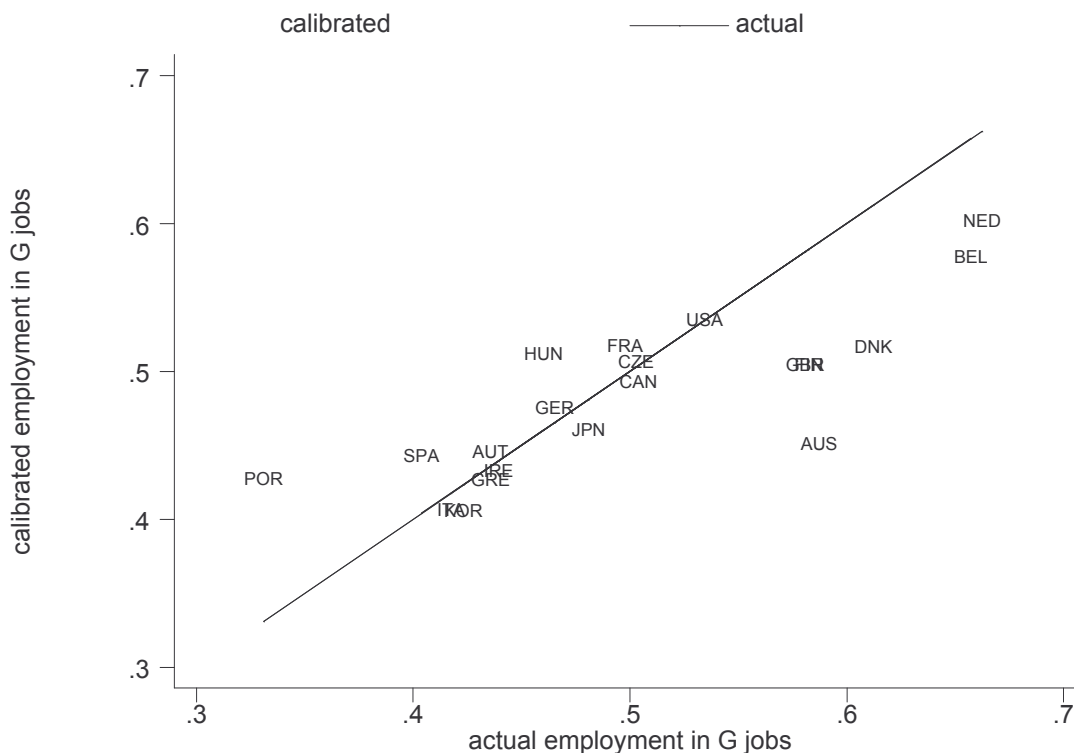
Figure 3

Figure 4 plots the actual versus the calibrated age of first selection. Since the

simple correlation between these two values is again high (0.803), we are confident that our model – in spite of its simplicity – goes a good way in reproducing the actual ordering of countries with respect to the age of selection, with the exception of Belgium and The Netherlands, for which we predict an age of selection significantly higher than the actual age. The figure shows that the gap between calibrated and actual tracking time is very small in 6 countries out of 20 (the US, Japan, Australia, France, Spain and the Czech Republic) and somewhat larger for the majority of countries with an early tracking system (Germany, Austria, Belgium, Hungary and The Netherlands).

Figure 4
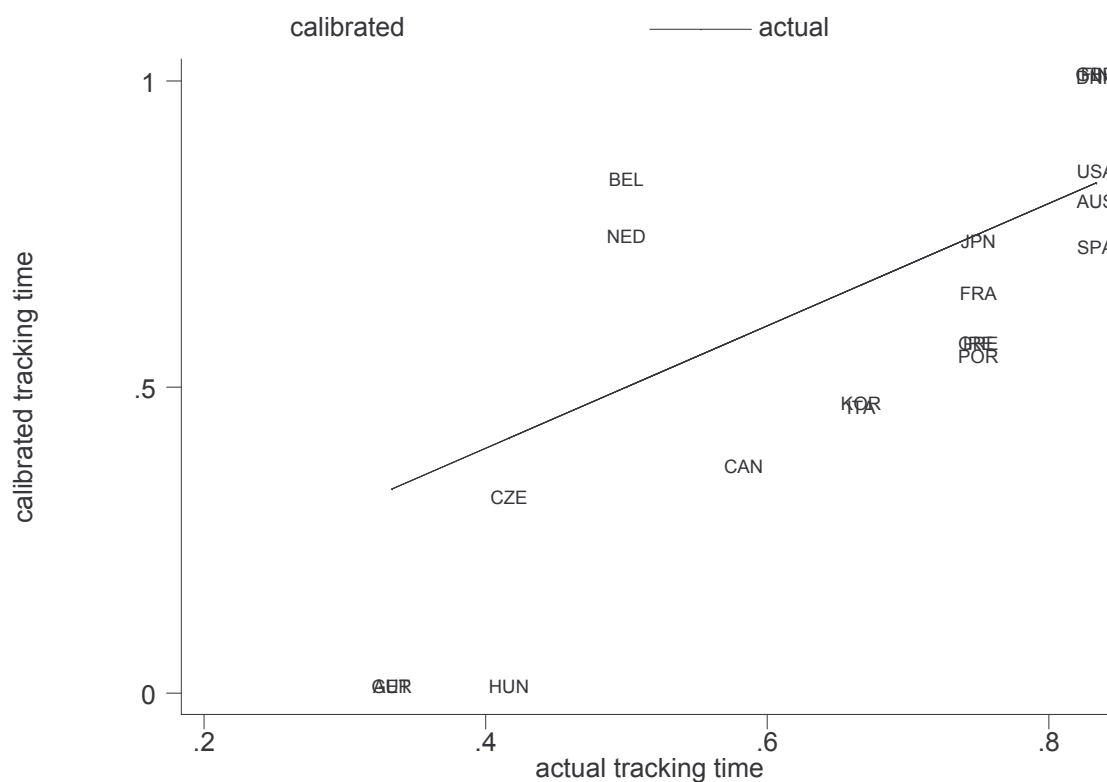
How do we explain the observed discrepancies between predicted and actual values? One natural candidate is that governments in different countries place

different weights on parents and students, for instance by giving priority to less privileged households. While we cannot exclude such possibility, we also recognize that it does not take us too far: after all, we can always accommodate any outcome by properly adjusting preferences, which are typically unobserved. Another possibility is that constrained welfare maximization needs not be the best characterization of the decision process in all the 20 countries in our sample. A closer inspection of our results suggests that this might be true for Hungary and, to a lesser extent, for Portugal, where welfare maximization does a better job in mimicking the actual outcomes.

An alternative explanation is that the implicit assumption that all parents in the ability distribution participate equally to the voting process is overly restrictive. It is well known that informed and educated parents take a closer look at the education of their offspring than less educated parents. If this is the case, then it might be that the median voter is not the parent with true ability equal to 0, as we have assumed so far, but a parent with a higher (or lower) ability. To investigate this point more in detail, we ask which voter would choose the selection threshold which minimizes the weighted sum of the squared difference between actual and calibrated tracking time and the actual and calibrated employment share.

In practice, we choose the value of α^i for each country which minimizes

$$\Omega^i = \xi_\tau \left[\hat{\tau}\left(\alpha^i\right) - \tau \right]^2 + \xi_G \left[\hat{N_G}\left(\alpha^i\right) - N_G \right]^2 \qquad [24]$$

where the hat refers to calibrated values, and ξ is the weight[15]. Let this value be α^i_p, which we report for each country in Table 4. It turns out that the pivotal voter's ability is very close to zero only in the US, Japan and the Czech Republic, is positive for the majority of the countries in the sample, and negative in some countries.

Since we believe that the cross country variation in the ability of the pivotal voter is related to differences in parental participation to the decision process, and we expect participation to be less likely among less educated parents, we investigate the relationship between the educational attainment of parents - who are most likely in

---

[15]The weights are the inverse of the coefficient of variation of each variable ($\xi_\tau = 0.826$ and $\xi_g = 0.174$)

the age group 35-54 - and the ability of the pivotal voter in our cross section of countries by regressing the latter variable on the former and on other candidate variables, which might affect participation.

Table 4. The pivotal voter

	α_p^i	$\hat{\tau}$	$\hat{N_G}$
Austria	0.41	0.314	0.386
Belgium	-0.15	0.487	0.629
Canada	0.19	0.561	0.464
Czech Republic	0.08	0.413	0.491
Germany	0.43	0.310	0.403
Denmark	-0.18	0.840	0.614
Spain	0.31	0.831	0.405
Finland	-0.14	0.811	0.576
France	0.13	0.749	0.495
United Kingdom	-0.13	0.828	0.579
Greece	0.30	0.715	0.396
Hungary	0.30	0.412	0.457
Ireland	0.30	0.716	0.402
Italy	0.31	0.633	0.376
Japan	-0.02	0.718	0.464
Korea	0.28	0.632	0.379
Netherlands	-0.34	0.478	0.690
Portugal	0.39	0.771	0.380
United States	-0.02	0.834	0.533
Australia	-0.32	0.682	0.513

After some experimenting, our preferred specification is reported below:

$$\alpha_p = 2.204 - 0.337\, ED - 0.212\, LNGNP + 0.568\, GER$$
$$\quad\ \ {}_{0.118}\quad\ {}_{0.181}\qquad\ {}_{0.054}\qquad\qquad {}_{.113}$$

$$R^2 = 0.560$$

where *ED* is the country specific educational attainment of individuals aged 35 to 54 (source: OECD, *Education at a Glance*) , *LNGNP* is the log of real GNP per capita (source: *World Bank Development Indicators*) and *GER* is a dummy equal to 1 for Germany and Austria and to 0 otherwise.

It turns out that the ability of the pivotal voter decreases both with GNP per capita and with the average educational attainment of adults. Therefore, as expected, parents in wealthier and better educated societies participate more to schooling decisions, and their pivotal voter has a relatively low ability. While differences in educational attainment and GNP per capita do a relatively good job in explaining cross country differences, they still fail to explain Germany and Austria: these two countries have both a pivotal voter with a relatively high positive ability, in spite of their relatively high average income and educational attainment (see Figure 5).

Figure 5
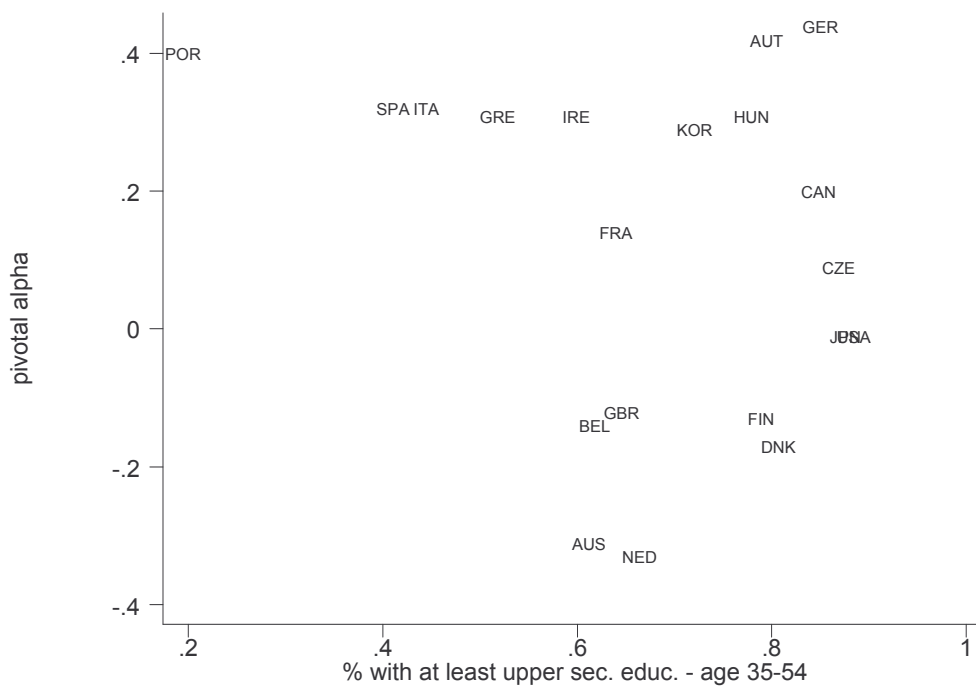
We admit that our simple model cannot fully cover the institutional variation encompassed by 20 countries by adding to the above regression a dummy for Austria and Germany[16].

Conclusions

Secondary schools in the developed world differ in the degree of differentiation and in the first age of selection of pupils into different tracks. In this paper, we account for the heterogeneity of tracking time with a simple stochastic model, which conjugates the returns from specialization with the costs of early selection. We have calibrated the model for 20 countries - most of Europe, Canada, the US, Korea, Japan and Australia - to show that it performs rather well in replicating the observed heterogeneity, with the remarkable exception of some countries with an early tracking system (most notably Germany).

If we are prepared to believe that our model captures the key features of school tracking, the observed discrepancies between actual and calibrated values can be accounted for by the fact that governments in different countries place different weights on parents and students, for instance by giving priority to less privileged households. While we do not exclude such possibility, we observe that it is always possible to accommodate outcomes by properly adjusting preferences, which are typically unobserved. Therefore, we take a different route, and argue that the implicit assumption that all parents in the ability distribution participate equally to the voting process is overly restrictive.

We show that the pivotal voter, who would choose a selection threshold consistent with the observed tracking time, has a true ability which is very close to the median voter of the unrestricted distribution of α in the countries for which our calibrated model fits well (the US, Japan, The Czech Republic). In the remaining countries, the pivotal voter's ability is either positive – suggesting that less able parents do not participate to the voting decision on the relative size of each type of

---

[16] We have also experimented with a measure of the degree of decentralization of schooling decisions, on the ground that more decentralization should help parental participation. While the cross – country correlation between decentralization and the pivotal voter's ability is negative, as expected, it never attracts a statistically significant coefficient.

school - or negative - which points to reduced participation also in the upper tail of the ability distribution, perhaps because the ablest have relatively high opportunity costs of participation. We conclude by showing that the cross country variation in the pivotal voter is related to differences in the educational attainment of adults: with the notable exception of Germany, parents in countries with a higher educational attainment of the population and with a high average income per capita tend to participate more to democratic decisions – including those concerning the schools where their offspring go – and in these countries the pivotal voter has lower ability than in less educated, and less wealthy, countries.

References

- Bertocchi, G. and Spagat, M., 2003, The Evolution of Modern Education Systems, in *Journal of Development Economics.*

- Brunello, G, Giannini, M. and Ariga, K., 2005, The Optimal Timning of School Tracking, forthcoming in P.Peterson and L. Woessmann, *Schools and the Equal Opportunity Problem*, MIT Press, Cambridge USA, 2005.

- Chang, J., 2004, Costly Voting: Voting Composition and Turnout, mimeo

- Checchi, D. and Flabbi, L., 2005, Intergenerational Mobility and Schooling Decisions, mimeo

- Dustmann, C., 2004, Parental background, primary to secondary school transitions and wages, *Oxford Economic Papers*, Bonn.

- Gamoran, A., and Mare, R., 1989, Secondary school tracking and education inequality: compensation, reinforcement or neutrality? *The American Journal of Sociology*, 94, 5, 1146-1183

- 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.

- Figlio, D. and Page, M., 2000, School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Equality? *NBER Working Paper* 8055

- Hanushek, E., 2002, Publicly Provided Education, *NBER Working Paper* 8799.

- Maddala, 1983, *Limited Dependent Variables*, Cambridge University Press.

- Meghir, C. and Palme, 2004, Educational Reform, Ability and Parental Background, *The American Economic Review*

- Milligan, K., Moretti, E. and Oreopoulos, P., 2003, Does education improve citizenship? Evidence from the US and the UK, *NBER Working Paper* 9584.

- OECD, 2004, *The PISA 2003 Study*, Paris

- Rivkin, S., Hanushek, E. e Kain, J., 2005, Teachers, schools and academic achievement, *Econometrica*, 73, 417-458

- Sinnott, R., 2003, Electoral Participation / Abstention: a Framework for Research and Politcy Development, mimeo

- Woessmann, L., and Hanushek, E., 2005, Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence across Countries", in press, *Economic Journal.*