The Optimal Timing of School Tracking*

Giorgio Brunello† Massimo Giannini‡ Kenn Ariga§

January 2004

Abstract

We develop a simple model which determines the optimal timing of school tracking as the outcome of the trade off between the advantages of specialization, which call for early tracking, and the costs of early selection, which lead to later tracking. We calibrate the model for Germany and study how relative demand shifts toward more general skills and changes in the (exogenous) rate of technical progress affect the optimal tracking time as well as the efficient allocation of students to general and vocational tracks.

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

*We are grateful to Gianni DeFraja, Charlotte Lauer, Paolo Manasse and to the audience in Milan, Tor Vergata and the CESifo / PEGP Conference in Munich for comments and suggestions. A special thanks to Richard Romano for his very constructive comments and for providing a detailed proof of Proposition 1. The usual disclaimer applies.

†University of Padova, CESifo and IZA. Address: Department of Economics, via del Santo 33, 35100 Padova, Italy

‡University of Molise

§Columbia University and Kyoto Institute of Economic Research

1. Introduction

Most primary and secondary school systems in the developed world consist of an initial period of exposure to the same curriculum followed by diversification of curricula into separate tracks. In Europe, there are vocational and general or academic tracks, with allocation into tracks often based on previous performance and / or on ability tests (see Shavit and Muller, 1998, and Green, Wolf and Leney, 1999[1]). Tracking starts relatively early, after primary school, in Germany and the Netherlands and later on in France. In the US, secondary schools are comprehensive but it is common practice to separate students into different courses or course sequences (tracks) based on their level of achievement or proficiency as measured by some set of tests or course grades (see Gamoran, 1987 and Epple, Newlon and Romano, 2002). In Japan, stratification starts at the post-compulsory stage in upper secondary education, with elite schools at the top and vocational schools at the bottom of the hierarchy (see Ishida, 1998).

International differences in school design have recently been associated in the economic literature to differences in economic performance. Krueger and Kuman 2002, for instance, have argued that the emphasis placed by Europe on specialized, vocational education may reduce the rate of technological adoption and lead to slower economic growth than in the United States, where the schooling system provides more general and comprehensive education. The broad idea is that general education is more suitable to induce (directed) technical change (see Acemoglu, 2000). Since general education is more flexible and versatile, it also encourages organizational change and the adoption of high performance holistic organizations in production (see Lindbeck and Snower, 2000, and Aghion, Caroli and Penalosa, 2000).

This literature looks at the effects of school design on technical and organizational change. It is natural to ask, however, whether and how these changes

[1] Vocational education is directly related to a specific occupation, with a substantial part of the curriculum devoted to learning practical skills to be used immediately upon graduation. General education has no immediate connection with any occupation, but provides basic knowledge that can be used to learn different occupations. See Bertocchi and Spagat, 2003.

affect in turn endogenous school design. The timing of tracking has changed in several European countries after the Second World War. In the UK there has been a shift in the mid 1960s from selection at 11 to selection at 16 (see Heath and Chieng, 1998). In Germany, where tracking by ability starts relatively early, reforms in the 1970s have increased compulsory education from 8 to 9 years, in an effort to make the system more comprehensive (see Muller, Steinmann and Ell, 1998). In France, direct orientation to apprenticeships after two years of lower secondary school was abolished in the 1980s (Goux and Maurin, 1998). All these reforms have gone in the direction of delaying tracking. Moreover, the fraction of the population in vocational secondary education to that in general secondary education has declined monotonously in most of post-war Europe (Bertocchi and Spagat, 2003).

Technical progress leads to skill depreciation, and the degree of obsolescence is likely to be higher the more specialized and tied to a specific set of techniques skills are. While skills learnt in vocational schools can be easily transformed into the corresponding occupations in the labor market, they are less flexible and transferrable than general skills (Shavit and Muller, 1998). As argued by Aghion, Caroli and Penalosa, 1999, organizational change is skill biased. Non hierarchical firms

> "..rely on direct, horizontal communication among workers and on task diversification as opposed to specialization. They hence require multi-skilled agents, who can both perform varied tasks and learn from other agents' activities." (p.1651)

One clear implication of organizational change is the relative demand shift toward more general and versatile skills (*upskilling*), which are better provided by general education.

School design clearly depends on a host of non-economic factors, including historical heritage. In this paper we take an economic perspective and focus exclusively on efficiency issues. By so doing, we are aware that we are only looking at one side of the problem. Nonetheless, we believe that our approach can be

a useful complement to non-economic approaches as well as to other economic approaches which consider the important distributional effects of school design.

We develop a simple model which determines the optimal timing of school tracking as the outcome of the trade off between the advantages of specialization, which call for early tracking, and the costs of early selection, which call instead for later tracking. The optimal tracking time is the time which maximizes total output net of schooling costs. We use the model to study how relative demand shifts toward more general skills and changes in the (exogenous) rate of technical progress affect the optimal tracking time as well as the allocation of students to schools.

Tracking is associated to selection, and the key factor in the selection process is perceived ability[2]. In a world of imperfect information, selection conveys information about individual ability to the labor market. Tracking also leads to ability grouping, with higher - achieving students being separated from lower - achieving students. It is still an open issue whether separating students into different tracks leads to better educational outcomes than mixing students of different ability. Epple, Newlon and Romano, 2002, briefly review the empirical literature and conclude that, relative to the outcomes of mixed classes, students assigned to low tracks are hurt by tracking while those assigned to high tracks gain. Our model is consistent with these findings.

As shown by Hoxby, 2001, peer effects have distributional effects but no efficiency implications if individual outcomes, such as human capital, are affected linearly by the mean of peers' outcomes in that variable. Efficiency implications can only be drawn from models which are either nonlinear in peers' mean achievement or in which other moments of the peer distribution matter (Hoxby, 2001, p.2)[3].

In our model, the presence of nonlinear peer effects implies that tracking

[2] In Germany, "..the decision about school track is taken by both parents and the local educational authorities...but children's measured ability remains the most important factor determining the selection process. This takes the form of a primary school recommendation for a secondary school track, generally based on a pupil's marks in the core subjects of German and mathematics.." (Schnepf, 2002)

[3] See also Epple and Romano 1998.

has a positive "specialization" effect. In the absence of a countervailing factor, however, positive specialization would lead to immediate tracking. We identify this factor by noticing that the allocation of individuals to tracks is affected by noise in the selection process, and that the relative importance of noise is higher the earlier the selection takes place. Misallocation due to imperfect testing reduces both the quality of the signal offered by schools to the labor market and the peer effects in human capital formation. As remarked by Judson, 1998[4],

> "..innate ability is measured with difficulty and with increasing clarity as education proceeds. Any test given will be a noisy signal, and the less education the person has had, the noisier the signal will be. Before primary school it is very difficult to discern levels of talent, but identification of talent is easier after a few years of primary school, still easier after high school, and so on.." (p.340)

The earlier selection is carried out, the higher the risk of misallocating individuals to the wrong track. We call this the "noise" effect of tracking. The trade-off between the positive "specialization" and negative "noise" effect generates an endogenous optimal tracking time.

The importance of ability tracking for school performance has already been studied in the literature, most recently by Epple, Newlon and Romano, 2002. These authors, however, ignore noise in the selection process and treat both the threshold ability required for the allocation of pupils to tracks and the tracking time as exogenous parameters. Allocation of individuals to tracks can be carried out either by prices (tuition fees) of by quantitative restrictions such as tests. Selection by test implies that individuals with a test score higher than the selected threshold are admitted to the high track and individuals with a lower score are allocated to the low track. Fernandez [1998] shows that allocation by tests should be preferred to allocation by prices when individuals are liquidity constrained. In the absence of constraints, however, the two selection methods are equivalent.

[4] See also Bedard, 1997 and Allen and Barnsley, 1993.

In spite of the very simple structure of the model, its stochastic nature implies that we can offer relatively few analytical results. Therefore, we resort to calibration and focus on the German institutional setup to study how the optimal tracking time and the relative share of graduates from general schools vary with changes in the size of the peer effect, the noise in the selection process, the (exogenous) rate of technical progress and the upskilling of labor from less to more general and versatile tasks. Conclusions follow.

2. The Model

2.1 Setup

Consider a simplified economy with an exogenous number of individuals and job slots. Each individual lives for two periods. In the first (preliminary) period she goes to school and in the second period she is matched to a job slot supplied by a firm. The exogenous number of individuals is normalized to 1. There is a given number of public schools M, each with one teacher and $\frac{1}{M}$ students. The monetary cost Z of running each school does not vary with school design. In the rest of the paper we normalize this cost to zero for the sake of simplicity.

The assumption of public schools is quite accurate for most countries if we focus on primary to upper secondary education, but less accurate if we consider also tertiary education. While our model can be extended to include college, we prefer to focus our attention on primary and secondary education. In many developed countries this coincides with compulsory education, which justifies our assumption of an exogenous length of time spent at school.

Let the period spent at school be equal to one unit, and define $\tau \in [0, 1]$ as the time when students are separated into tracks, or tracking time. Then τ is also the period spent in a comprehensive school and $(1 - \tau)$ is the time spent in a stratified school. While a comprehensive school provides the same curriculum to everybody, in a stratified school students are allocated to two different tracks, H (high - ability) and L (low - ability), each with its own specialized curriculum.

In the US, the H and L tracks are segregated classes which coexist within the same comprehensive school. In most European countries, they correspond to general (academic) and vocational education.

When $\tau = 1$ all M schools are comprehensive for the entire period of time. When $\tau < 1$ the M schools are comprehensive for time length τ and are divided into MX classes or schools in the L track and $M(1-X)$ classes in the H track for the rest of the time, where X is the percentage of pupils going to L tracks. By assumption, there is no further stratification within each type of school.

Risk neutral individuals care only about (expected) wages and differ in their endowed ability, which cannot be observed by firms when recruitment takes place. While we can think of several types of ability, in this paper we focus only on cognitive ability, and assume that individuals differ in their endowment of this single type[5]. Firms only know the school the individual has graduated from. Since ability cannot be observed, each individual is paid her expected productivity. In this environment, firms make zero expected profits and the efficient social outcome is produced by the school design which maximizes total net output[6].

When individual utility depends only on expected wages after school and admission to H and L schools is free and left to individual choice, all individuals should enrol in track H if the wage of graduates from these tracks is expected to be higher than the wage gained by L graduates. We assume that 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 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.

2.2 Schools

[5] See Brunello and Giannini, 2004a, 2004b for models with two ability types.
[6] This design also maximizes a utilitarian welfare function.

Using small letters for logarithms, let true ability $A \in (0, \infty)$ be log-normally distributed across individuals, and define $\alpha = \ln(A) \sim N(0,1)$. Let observed log ability θ when the test takes place be related to true log ability by:

$$\theta = \alpha + \epsilon \tag{1}$$

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 the test increases the earlier the test is taken by letting

$$b = \mu\,(1 - \tau) \tag{2}$$

where μ is a suitable parameter[7]. It follows that observed ability is normally distributed with zero mean and variance $1 + b^2$.

Since α and ε are both normally distributed, the conditional density ψ of α given θ is

$$\psi(\alpha|\theta) = \left(\frac{2\pi b^2}{1+b^2}\right)^{-\frac{1}{2}} \exp\left[\frac{-\frac{1}{2}(\alpha - \frac{\theta}{1+b^2})^2}{\frac{b^2}{1+b^2}}\right] \tag{3}$$

and the conditional mean is a linear function of observed ability θ (see Anderson and Moore 1979)

$$E\left[\alpha|\theta\right] = \frac{\theta}{1+b^2} \tag{4}$$

If observed log ability is positive, expected log true ability is higher the lower the variance of the noise. If on the other hand observed log ability is negative, expected log true ability falls as the variance of the noise declines.

If the government sets the test standard θ^* to allocate individuals to tracks, the expected log true ability of individuals is $E\left[\alpha|\theta \geq \theta^*\right]$ and $E\left[\alpha|\theta < \theta^*\right]$ in H and L tracks respectively. Using the Law of Iterated Projections (Sargent 1979) we get

$$E\left[\alpha|\theta \geq \theta^*\right] = E\left[E\left[\alpha|\theta\right]|\theta \geq \theta^*\right] = \frac{1}{1+b^2}\frac{\int_{\theta^*}\theta\phi(\theta)d\theta}{1-\Phi(\theta^*)} = \frac{1}{1+b^2}E\left[\theta|\theta \geq \theta^*\right] = m_h \tag{5}$$

[7] The specification (2) should be considered as a convenient linearization of the relationship between the size of the noise and the time when selection occurs. The true relationship need not be linear.

$$E\left[\alpha|\theta<\theta^*\right]=E\left[E\left[\alpha|\theta\right]|\theta<\theta^*\right]=\frac{1}{1+b^2}E\left[\theta|\theta<\theta^*\right]=m_l \tag{6}$$

Since the unconditional mean of α is equal to zero by assumption, m_h and m_l are positive and negative respectively. We can establish the following Remark.

Remark 1: *The expected log true ability of pupils in H and L tracks is increasing in the threshold* θ^*.

Proof. See Appendix. ■

An increase in the selection standard θ^* eliminates from H tracks individuals in the lowest observed ability group, who are allocated to L tracks, where they belong to the highest observed ability group (see Betts [1998]). Therefore, expected conditional observed ability of either group increases. Since expected conditional true ability increases with expected conditional observed ability, the former increases as well for both groups.

Each school combines individual ability with the effectiveness of teaching to produce human capital. Since by assumption the number and quality of schools and teachers are given, we posit that effectiveness varies with the average ability of the class (*peer effect*)[8]: the abler the class the more effective is instruction provided by a teacher of given quality. If an individual spends all her first period in a comprehensive school ($\tau=1$), her human capital at the end of the period is

$$H_c=A\exp[\beta E(\alpha)] \tag{7}$$

where $\exp\beta E(\alpha)$ is the peer effect. In the selected specification peer effects are convex and their impact on individual human capital is higher for abler individuals. Therefore, winners in the H track win more than losers in the L track lose, and there are average gains from tracking. The (log) human capital accumulated in this type of schools is

$$h_c=\beta E(\alpha)+\alpha=\alpha \tag{8}$$

[8] Zimmer and Toma, 2000; Hoxby, 2001; Zimmermann, 2003; Hanushek, Klain, Markman and Rivkin, 2001 is a non exhaustive list of recent contributions on peer effects.

Next consider schools stratified into tracks. Pupils in H tracks have an observed ability θ higher than θ^*. If they spend all their time in such tracks their log individual human capital is

$$h_h = \beta E\left[\alpha|\theta \geq \theta^*\right] + \alpha = \beta m_h + \alpha > \alpha \tag{9}$$

Similarly for L tracks we have

$$h_l = \beta E\left[\alpha|\theta < \theta^*\right] + \alpha = \beta m_l + \alpha < \alpha \tag{10}$$

Notice that $h_h > h_c > h_l$. Therefore an implication of tracking is that the human capital of high ability students increases while the expected human capital of low ability students falls with respect to no tracking. This feature of the model is consistent with the existing empirical literature reviewed in the introduction.

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. The individual log human capital at the end of the schooling period is

$$h_H = \tau h_c + (1-\tau)h_h = \alpha + (1-\tau)\beta m_h \tag{11}$$

if the student is assigned to the H track and likewise for students allocated to L tracks, except from the fact that we allow skills accumulated in the lower - ability track to depreciate at the rate g, where g is the rate of exogenous technical progress[9].

The asymmetric obsolescence effects of technical progress can be justified as follows. First, ability lessens the adverse effect of technological change (see Galor and Moav, 2000). Second, if we interpret skills developed in the L tracks as vocational, these skills are less flexible and adjustable than the general skills developed in the H track, and they depreciate faster. In the second period, the

[9] Since we are only concerned with the relative effect of technical change on vocational and general skills, we find it convenient to normalize the obsolescence of general skills to zero.

human capital of an individual who has enrolled in a L track is

$$H_L = H_c^{\tau} \left[H_l(1-\delta g)\right]^{1-\tau}$$

where δ is a suitable parameter. Using logs and the approximation $\ln(1-x) \simeq -x$[10], we obtain

$$h_L = [\alpha + (1-\tau)\beta m_l] - (1-\tau)\delta g \tag{12}$$

In the second period 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. Suppose that the graduate has spent all her education in a comprehensive school ($\tau = 1$). In this case her expected human capital is

$$Eh_c = E(\alpha) = 0 \tag{13}$$

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

$$Eh_H = E(h_H|\theta \geq \theta^*) = (1-\tau)\beta m_h + E(\alpha|\theta \geq \theta^*) = [1 + (1-\tau)\beta]\, m_h \tag{14}$$

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

$$Eh_L = [1 + (1-\tau)\beta]\, m_l - (1-\tau)\delta g \tag{15}$$

Casual observation of schooling around the world suggests that primary education and often lower secondary education are comprehensive, with tracking starting later on. In principle, however, we could have tracking from the start followed by a period of comprehensive schooling. Suppose for instance that tracking lasts for the period $(1-\tau)$, followed by comprehensive schooling for

[10] This approximation is reasonable since δg is small.

the remaining period τ. Assuming that firms have information on the entire school curriculum, expected human capital would be as in (14) and (15), and so would be depreciation. The only key difference between tracking first and tracking later is that noise and misallocation in selection are higher when tracking starts earlier on.

Expected log human capital in either track varies with tracking time. Differentiation of (14) with respect to τ yields

$$\frac{\partial m_h}{\partial \tau} = -\beta m_h + [1 + (1-\tau)\beta] \frac{2b\mu}{1+b^2} m_h - [1 + (1-\tau)\beta] \frac{\mu}{1+b^2} \frac{\partial E(\theta|\theta \geq \theta^*)}{\partial b} \tag{16}$$

Later tracking reduces the expected human capital of the high ability group, because students in this group spend less time together and have fewer opportunities to enjoy the positive peer effect. On the other hand, later tracking reduces the noise in the selection process, which positively affects human capital (second term on the right hand side). Finally, later tracking also alters the conditional distribution of observed ability, with uncertain effects on expected human capital. Similarly, differentiation of (15) yields

$$\frac{\partial m_l}{\partial \tau} = -\beta m_l + [1 + (1-\tau)\beta] \frac{2b\mu}{1+b^2} m_l - [1 + (1-\tau)\beta] \frac{\mu}{1+b^2} \frac{\partial E(\theta|\theta < \theta^*)}{\partial b} \tag{17}$$

In the case of lower - ability students, later tracking reduces the negative peer effects ($m_l < 0$), with a positive effect on expected human capital.

2.3 Firms

The economy is populated by a given number of identical firms, which produce output by using two types of jobs or tasks, a "G" and a "V" task. G tasks are general and require versatility and high ability. V tasks, on the other hand, are narrowly defined, vocational, and can be filled by less talented individuals. In the absence of tracking both tasks can be filled indifferently by all graduates. With tracking, however, specialization makes graduates of H tracks

more suitable for G tasks and graduates of L tracks a better match for V tasks. Therefore, tracking entains both the creation of different peer groups and specialized education[11]. For convenience we normalize to 1 the number of firms. The Cobb Douglas production technology is given by

$$y = a + \lambda (n_G + Eh_H) + (1 - \lambda)(n_V + Eh_L) \tag{18}$$

where a is the log of the technical level, y is log real output, $\lambda \in (0, 1)$ and n_G and n_V are the log of the number of employees in G and V tasks. Profit maximization yields

$$w_G = \ln \lambda + y - n_G \;\; ; \;\; w_V = \ln(1 - \lambda) + y - n_V \tag{19}$$

where w is the log wage rate. Relative wages in this economy satisfy the following condition

$$w_G - w_V = \ln \frac{\lambda}{(1 - \lambda)} + n_V - n_G \tag{20}$$

Following Katz and Murphy, 1992, $\ln \frac{\lambda}{(1-\lambda)}$ measures relative demand shifts in log quantity units, or upskilling. A demand shift toward more general tasks (a higher value of λ) can be met either by an increase in relative wages or by an increase in the relative supply of general skills or finally by a combination of both. Relative supply depends on the selection threshold, θ^*, and on the optimal timing τ, which are set by the government to maximize net output.

2.4 The Optimal Policy

When schools are comprehensive ($\tau = 1$), graduates have the same expected human capital and can fill indifferently either task. Since perfect competition in the labor market guarantees that $w_G - w_V = 0$, relative employment is simply

$$n_G - n_V = \ln \frac{\lambda}{(1 - \lambda)} \tag{21}$$

[11] While most theoretical contributions on tracking emphasize peer effects, our contribution is the first to introduce the combination of peer effects and specialized education.

Labor supply is defined by

$$\ln(N_G + N_V) = 0 \tag{22}$$

Therefore $n_G = \ln \lambda$ and log net output ny_c is equal to y

$$ny_c = y = a + \lambda \ln \lambda + (1-\lambda)\ln(1-\lambda) \tag{23}$$

With selection, there are $1 - \Phi(\theta^*)$ graduates from the H track and $\Phi(\theta^*)$ graduates from the L track, and total net output ny_s can be re-written as

$$\begin{aligned} ny_s &\equiv \chi(\tau, \theta^*, \lambda, g, \mu, \delta) \\ &= a + \lambda \ln\left[1 - \Phi(\theta^*)\right] + (1-\lambda)\ln \Phi(\theta^*) \\ &+ \left[1 + (1-\tau)\beta\right]\left[\lambda m_h + (1-\lambda)\, m_l\right] - (1-\lambda)\,(1-\tau)\delta g \end{aligned} \tag{24}$$

The government maximizes net output by selecting the optimal values of τ and θ^*. The first order conditions are

$$\begin{aligned} \chi_\tau(\tau, \theta^*, \lambda, g, \mu, \delta) : &-\beta\left[\lambda m_h + (1-\lambda) m_l\right] + (1-\lambda)\delta g \\ &+ \left[1 + (1-\tau)\beta\right] \frac{2\mu b}{1+b^2}\left[\lambda m_h + (1-\lambda) m_l\right] \\ &- \left[1 + (1-\tau)\beta\right] \frac{\mu}{1+b^2}\left[\lambda \frac{\partial E\left[\theta | \theta \geq \theta^*\right]}{\partial b} + (1-\lambda)\frac{\partial E\left[\theta | \theta < \theta^*\right]}{\partial b}\right] \\ &+ \left[\frac{1-\lambda}{\Phi} - \frac{\lambda}{1-\Phi}\right] \frac{\mu b \Phi}{1+b^2}\left(1 - \frac{E\left[\theta^2 | \theta < \theta^*\right]}{1+b^2}\right) = 0 \end{aligned} \tag{25}$$

$$\begin{aligned} \chi_{\theta^*}(\tau, \theta^*, \lambda, g, \mu, \delta) : &-\frac{\lambda\phi}{1-\Phi} + \frac{(1-\lambda)\phi}{\Phi} \\ &+ \left[1 + (1-\tau)\beta\right]\left[\lambda \frac{\partial m_h}{\partial \theta^*} + (1-\lambda)\frac{\partial m_l}{\partial \theta^*}\right] = 0 \end{aligned} \tag{26}$$

Remark 1 implies that an internal solution for the threshold θ^* exists if

$$-\frac{\lambda\phi}{1-\Phi} + \frac{(1-\lambda)\phi}{\Phi} < 0$$

This condition can be rewritten as

$$\frac{\lambda}{(1-\lambda)} > \frac{1-\Phi}{\Phi} = \frac{N_G}{N_V} \tag{27}$$

which implies from (20) that $w_G - w_V > 0$. Therefore, with tracking and selection the graduates of H tracks - which have higher average observed and true ability - are paid in equilibrium a higher wage than the graduates of L tracks. We use this result to establish the following Lemma

Lemma 1: $[\lambda m_h + (1-\lambda)m_l]$ *is positive.*

Proof. See Appendix. ■

This Lemma implies that, at the optimal value of the selection threshold θ^*, a linear combination of the expected abilities of H and L graduates, with weights equal to the relative wage bill of each group, is higher than the expected ability of graduates of a comprehensive school, which is equal to zero by definition. We call this the "specialization effect" of tracking.The first order condition with respect to τ is composed of five terms: the first term is negative and captures the fact that later tracking reduces the gains from specialization. The second term is positive because later tracking is associated with lower depreciation of vocational skills; the third term is positive because later tracking reduces the noise in the selection process; the last two terms capture the changes in the conditional distribution of θ as τ varies and can take either sign. In the absence of noise, $\mu = 0$ and (25) boils down to

$$-\beta\left\{[\lambda E\left[\theta|\theta \geq \theta^*\right] + (1-\lambda)E\left[\theta|\theta < \theta^*\right]]\right\} + (1-\lambda)\delta g = 0 \tag{28}$$

Without skill depreciation the left hand side is negative and optimal τ is equal to zero: in the absence of noise and depreciation, the positive effects of specialization prevails and tracking starts from the beginning of the schooling period. On the other hand, in the absence of peer effects ($\beta = 0$) the left hand side is positive and the optimal tracking time is $\tau = 1$ (no tracks).

Assuming that the second order conditions for the maximum hold, we can establish the following

Lemma 2: $\theta^* \in (-\infty, \infty)$ *and the selection threshold is finite*

Proof. Since $N_G = 1 - \Phi(\theta^*)$ and $N_V = \Phi(\theta^*)$, the threshold needs to be a finite number to guarantee positive output. ■

Proposition 1: *Tracking can be optimal*

Proof. See Appendix[12] ■

Corollary 1: *If tracking is optimal,* $\tau^* \in (0, 1)$ *when the noise parameter* μ *is sufficiently large*

Proof. See Appendix ■

Proposition 2: *When an interior solution* (τ, θ^*) *exists, the effect of an acceleration in the rate of TFP growth* g *on the optimal tracking time* τ *is positive.*

Proof. See Appendix. ■

Proposition 1 is key because it establishes conditions for an efficient solution with $\tau < 1$ to exist. Corollary 1 shows that, if the noise of the test is large enough and Proposition 1 holds, an internal solution for tracking time τ exists. An acceleration of growth increases the depreciation of skills provided by vocational schools. The optimal government response consists of delaying stratification. Unfortunately, because of the complexity of (25), the two propositions and the corollary are the only analytical results that can be derived from the model. Therefore, we turn to calibration and illustrate the properties of the model by focusing on the German system of early tracking.

3. Calibration

Stratification by ability in Germany starts at age 10, when pupils are allocated to the H track (*Gymnasium)* or to the L track (*Hauptschule* and *Realschule).* While education in the H track is general, the L track leads in most cases to vocational education and training (see Schnepf, 2002). The calibration of the model requires that we assign numerical values to the parameters β, δ and

[12] We are indebted to Richard Romano for suggesting the Proposition and especially for providing a proof.

λ. Starting with β, we need to recognize that most available empirical evidence on the size of peer effects is based on US data. In a recent survey of the US empirical literature, Hoxby, 2001, reports that the estimated value of β ranges between 0.15 and 0.4. We assume that these estimates can also be applied to Germany and take a conservative view by setting $\beta = 0.2$.

Next consider parameter δ. We start from the working assumption that average working life during the second period lasts 30 years and take from Nickell and Layard, 1999, the 1976 to 1992 average annual rate of total factor productivity growth in the private sector in Germany, which is equal to 0.0191. We use ECHP data for Germany[13] for the period 1994-2000 and identify G tasks with professionals, technicians and clerks and V tasks with craft workers and plant and machine operators. We select male workers aged 25 to 59 employed full time in the private sector and fit for each occupational group the following Mincerian equation

$$\ln w = \alpha + \beta X + \gamma AGE + \eta AGE^2 + u$$

where w is the hourly wage, X a vector of standard controls and AGE is individual age. The fitted regression is used to predict the age wage profile at ages 29 and 59 respectively. Defining $Z_j^i = \bar{\gamma} AGE_j^i + \bar{\eta} AGE_j^{i2}$ as the fitted wage for age $i, i = 29, 59$, and occupational group $j, j = H, L$, the ratio

$$\frac{\frac{Z_L^{59}}{Z_L^{29}}}{\frac{Z_H^{59}}{Z_H^{29}}} = \omega$$

can be considered as a proxy of the depreciation of L skills after 30 years in the labor market, relative to H skills. Our estimates suggest that $\omega = 0.862$. The value of δ must be such that the relative value of human capital in L tracks after 30 years of use is equal to ω. Therefore we estimate δ by solving

$$(1 - 0.0191\delta)^{29} = 0.862$$

which yields $\delta = 0.267$. Since one single period in the model corresponds to 30 years of working life, it is not appropriate to use in the calibrations the

[13] The ECHP data - release 2003 - are available at the Department of Economics, University of Padova, contract 4/99.

annual rate of productivity growth, which refers to a single year. We define the average rate of technical progress over 30 years, g_{30}, as the rate which produces in a single year of depreciation the average value of human capital over 30 years of working life and solve

$$1 - 0.267 g_{30} = \frac{1 + (1 - 0.267 * 0.0191) + ...(1 - 0.267 * 0.0191)^{29}}{30}$$

which yields $g_{30} = 0.264$.

With a Cobb Douglas production function, λ is the share on the total wage bill of the wages paid out to workers in G jobs. Therefore

$$\lambda = \frac{W_G N_G}{WN}$$

We use the 2000 wave of ECHP and estimate that the value of λ for Germany is 0.625.

With these values of the key parameters in hand, we illustrate in Figures 1 to 4 how the optimal tracking time τ and the optimal selection threshold θ^* adjust to variations in the peer effect β and in the noise parameter μ. In Figures 1 and 2 we plot the optimal values of τ and θ^* by keeping β constant and by allowing μ to vary between 0 and 3. In Figures 3 and 4 we set instead μ to 0.495, the value which would produce as an internal solution for τ the observed value, and allow β to vary between 0 and 1.

Figure 1 shows that, as μ increases from zero, the optimal value of τ also increases and converges fairly rapidly to its upper value, where schools are fully comprehensive. Figure 2 shows that the increase in τ as μ rises is accompanied by a reduction in the optimal threshold θ^*. Finally, Figures 3 and 4 show that an increase in the size of the peer effect, given the noise in the test, reduces the optimal tracking time and increases the selection threshold.

In particular, it takes a value of the peer effect equal at least to 0.5 to make tracking from the start optimal. These figures suggest that optimal τ and θ^* tend to move in opposite directions: later tracking is associated to less selective

tests for access to H tracks and consequently to a higher share of students in these tracks. Therefore the two policy instruments turn out to be substitutes in the maximization of total net output.

The calibration of β and λ leaves two endogenous variables, τ and θ^*, and an additional parameter, μ, which measures the relative variance of the noise in the test with respect to the variance of true talent, α. Clearly, it is very difficult to pin down μ. Rather than trying to do this, we assume that the actual value of τ in Germany is equal to the optimal value and solve (25)-(26) for θ^* and μ. Since tracking time is likely to be persistent and vary slowly over time, we feel that this working hypothesis is reasonable.

The actual value of τ for Germany is 0.31 and is computed as the ratio of the total years of schooling spent in a comprehensive system, before selection takes place, - 4 years - to the total years of schooling from primary school to upper secondary education - 13 years. The corresponding value of μ turns out to be 0.495. The value of the selection threshold and the percentage of students enrolled in H tracks associated to the assigned parameters and to the actual value of τ are 0.812 and 0.221 respectively. The latter value is very close to the percentage of high school graduates from general tracks reported by the OECD for Germany in 1995 (0.23)[14], which suggests that our calibration baseline is not far from observed values.

Next, we turn to simulations and consider the following experiments: a) a 25 percent decline in the rate of productivity growth, a proxy of the rate of technical progress g_{30}, which corresponds to the decrease experienced by (West) Germany between the early 1980s and the late 1990s (see Gust and Marquez, 2002); b) a 10 percent increase in the relative demand shift parameter λ, a good approximation of the increase in the actual wage bill share of non-production workers between 1970 and 1990 (see Berman and Machin, 2000); c) a 10 percent increase in the peer effect β; d) a 10 percent increase in the noise parameter μ. The results are reported in Table 1[15]. The figures in the table are percentage

[14] OECD; *Education at a Glance*, Paris, 1997.

[15] In each simulation we solve explicitly for θ^* and perform a detailed grid search for τ to

deviations from the baseline solution described above.

Table 1. Simulation results. Percentage deviations from the baseline.

		θ^*	τ	Eh_H	Eh_L	N_G
-25%	g_{30}	0.70	-16.10	0.10	0.80	0
$+10\%$	λ	-13.10	-12.90	-10.60	-6.50	18.10
$+10\%$	β	2.80	-29.03	3.30	-0.20	-0.50
$+10\%$	μ	-2.10	38.71	-2.80	0.50	1.40

The optimal tracking time τ is affected negatively by the decline in the rate of productivity growth g_{30}, as predicted by Proposition 1, and by the relative demand shift toward more general and versatile jobs, measured by λ. More in detail, we find that a 25% reduction in g_{30} triggers a 16.1% decline in the optimal tracking time. We also find that a 10% increase in λ reduces tracking time by 12.9%. If we simulate the combined effect of g and λ on τ, we obtain that the optimal tracking time should decline by 22.6%.

Starting from 4 years of comprehensive school before selection into tracks, which corresponds to the German situation in the early 1970s, these simulations imply that the optimal tracking time should have been anticipated further by the end of the century to about 3 years of comprehensive school in order to accommodate the slowdown of productivity growth and the relative demand shift toward more general and versatile jobs. In practice, however, during this period "..reforms have attempted to narrow the gap between the Hauptschule and the other tracks through prolongation of compulsory education from eight to nine years and by introducing additional subjects into the curriculum..." (Muller, Steinmann and Ell, 1998, p.145). These reforms can be interpreted as a prolongation of the comprehensive period and as a delay of the tracking period.

We see two ways to reconcile our simulations with the observed trends in German school design. The most natural way is to argue that either the size

find the pair which maximizes total net output.

of peer effects has declined or the noise in the selection process has increased, perhaps as a consequence of the substantial inflow of immigrants. As shown in Table 1, the efficient tracking time τ is very sensitive to changes in these two parameters. The other way is to interpret the current trends as deviations from the efficient policy, driven perhaps by distributional and equity concerns.

If the observed equilibrium is a political equilibrium driven by majority voting, tracking can be delayed if the majority of students are in the vocational track. The pressure of majority could also affect the optimal selection threshold, because a higher threshold improves both peer groups, lowering the human capital only of those who are forced in the lower track[16].

Our simulations also show that the relative share of graduates from general tracks, which depends on the strictness of the selection criterion θ^*, is marginally affected by changes in g_{30} but varies significantly with changes in λ. In particular, a 10% increase in λ is expected to reduce significantly the admission threshold and to increase by 18.1% the share of H graduates. We conclude from this that the widespread academic drift, which characterizes both Germany and other developed countries, can be interpreted as the response of school design to the relative demand shift toward more general and versatile skills[17].

Table 1 also reports the impact of each simulation on the expected individual human capital in each track. We find that a 10 percent increase in parameter λ leads to a significant reduction in the expected human capital associated to either track. Since upskilling increases the relative size of the academic track, individuals with relatively lower ability are admitted to this track, which reduces average human capital. Similarly, the lower track loses the individuals with highest ability and ends up with lower average human capital. Relative wages can go either way, because the higher value of λ is compensated by the increase of N_G.

[16] We are grateful to Richard Romano for suggesting these points to us.
[17] Academic drift is discussed in detail by Green, Wolf and Leney, 1999.

Conclusions

We have presented a simple model of endogenous tracking in secondary schools. In the model, tracking has two features: the time spent in separate tracks and the relative size of each track, which depends on the difficulty of the admission test. Optimal tracking is the outcome of the trade-off between the advantages of specialization and the costs of early selection and skill obsolescence. We calibrate the model for Germany and simulate how endogenous school design should vary with the significant changes in the rate of technical progress and in the relative demand for skilled and versatile jobs which occurred in Germany during the last twenty years of the century.

Our simulations suggest that the relative share of graduates from general schools should have significantly increased, which confirm the existing evidence on academic drift in secondary schools. They also suggest that tracking time should have been anticipated by close to 23 percentage points, which is not what has happened in Germany since 1970. We speculate that either other key parameters have changed in the required direction - a reduction in the size of peer effects and / or an increase in the noise of the selection test - or that the observed policies have deviated from efficiency considerations, perhaps because of distributional concerns.

Our simple model can be best viewed as a first step in the modeling of endogenous tracking. To be simple, a model requires assumptions. Some of the assumptions used in the paper can be removed or modified in future research. For instance, we have assumed a frictionless labor market and no uncertainty about the future allocation of general and vocational tasks. Removing this assumptions would allow us to consider the possibility of mismatch between the supply and the demand of skills, and to discuss important issues such as over-education. Moreover, when the future is uncertain and labor market frictions do not allow an instantaneous adjustment of demand and supply, an additional reason to delay tracking is the option value of waiting.

We have focused in the paper on efficiency issues, and have rstricted attention

to policies which maximize net output. It would be interesting to contrast this approach with an approach based on the concept of political equilibrium. We have speculated in the paper that the pressure of majority voting could affect significantly both the tracking time and the selection threshold.

Finally, we have restricted attention to secondary schools and ignored college choice. The introduction of college in the model complicates things in a number of ways: first of all, we need to consider that a relevant percentage of students do not enrol in college. Second, the sequence of comprehensive and stratified education typical of compulsory education can be modified in an interesting way, because students from vocational tracks can enrol in colleges providing general education and thereby reduce their specialization.

We plan to consider these and other extensions of the model in future research.

Appendix

- Proof of Remark 1:

$$\frac{\partial m_h}{\partial \theta^*} = \frac{1}{1+b^2}\frac{\phi(\theta^*)}{1-\Phi(\theta^*)}\left[E\left[\theta|\theta \geq \theta^*\right] - \theta^*\right]$$

  is positive because the expression within brackets is positive. Similarly

$$\frac{\partial m_l}{\partial \theta^*} = \frac{1}{1+b^2}\frac{\phi(\theta^*)}{\Phi(\theta^*)}\left[\theta^* - E\left[\theta|\theta \leq \theta^*\right]\right]$$

  is also positive.

- Proof of Lemma 1: the expression $[\lambda m_h + (1-\lambda)m_l] > 0$ can be written as

$$\lambda \int_{\theta^*} \theta f(\theta)d\theta \int^{\theta^*} f(\theta)d\theta + (1-\lambda)\int^{\theta^*} \theta f(\theta)d\theta \int_{\theta^*} f(\theta)d\theta > 0$$

Adding and subtracting from the left hand side of the above expression $\lambda \int^{\theta^*} \theta f(\theta)d\theta \int^{\theta^*} f(\theta)d\theta$ and using the facts that $E(\theta) = 0$ and $m_l < 0$, we can rewrite it as follows

$$(1-\lambda)\left[1-\Phi(\theta^*)\right] < \lambda\Phi(\theta^*)$$

which corresponds to (27) in the main text.

- Proof of Proposition 1. Assume that $\tau \Longrightarrow 1$ in the limit (no tracking). As τ converges to 1 $w_G = w_V$ and

$$\frac{\Phi(\theta^*)}{1-\Phi(\theta^*)} = \frac{1-\lambda}{\lambda}$$

so that (26) turns positive and the selection threshold increases while remaining a finite number. Next, it is convenient to re-write

$$\lambda m_h + (1-\lambda)m_l = \frac{E\left[\theta|\theta \leq \theta^*\right]}{1+b^2}\left[1-\lambda-\lambda\frac{\Phi(\theta^*)}{1-\Phi(\theta^*)}\right]$$

by using

$$E\left[\theta|\theta \geq \theta^*\right] = \frac{E(\theta) - \Phi(\theta^*)E\left[\theta|\theta \leq \theta^*\right]}{1-\Phi(\theta^*)}$$

and $E(\theta) = 0$. It follows that in the vicinity of $\tau = 1$ the moving average $\lambda m_h + (1-\lambda)m_l$ is equal to zero, which makes sense in the absence of tracking. Recall that in the vicinity of $\tau = 1$ the variance b^2 tends to zero, and the distribution of observed ability θ converges to the distribution of true ability α, which is independent of b and τ. Therefore in the vicinity of $\tau = 1$ the first order condition (25) boils down to

$$\chi_\tau = (1-\lambda)\delta g$$

Notice that $(1-\lambda)\delta g$ must be small for the approximation in (12) to be correct. Finally write total net output in implicit form as

$$\chi = \chi(\tau, \theta^*(\tau))$$

Then optimal timing in the vicinity of $\tau = 1$ is given by

$$\frac{\partial \chi}{\partial \tau}_{|\tau \simeq 1} = \chi_\tau + \chi_{\theta^*} \frac{\partial \theta^*}{\partial \tau}$$

The first element on the left hand side is positive but small. The second element has two parts: the first part is positive because of (26) and $\frac{\Phi(\theta^*)}{1-\Phi(\theta^*)} = \frac{1-\lambda}{\lambda}$; the second part is negative because $\chi_{\theta^*\tau} = -\beta\left[\lambda \frac{\partial m_h}{\partial \theta^*} + (1-\lambda)\frac{\partial m_l}{\partial \theta^*}\right]$. Since χ_τ is positive but small and $\chi_{\theta^*}\frac{\partial \theta^*}{\partial \tau}$ is negative, we can have that

$$\frac{\partial \chi}{\partial \tau}_{|\tau \simeq 1} < 0$$

which guarantees that tracking ($\tau < 1$) can be optimal.

- Proof of Corollary 1. We need to show that $\frac{\partial \chi}{\partial \tau}_{|\tau \simeq 0} = \chi_\tau + \chi_{\theta^*}\frac{\partial \theta^*}{\partial \tau} > 0$ in the vicinity of $\tau = 0$. We start by noticing that

$$\chi_\tau(\tau, \theta^*)|_{\tau=0} = (1-\lambda)\delta g + \frac{(2+\beta)\mu^2 - \beta}{1+\mu^2}\left[\lambda m_h + (1-\lambda)m_l\right]$$
$$- \left[1+\beta\right]\frac{\mu}{1+\mu^2}\left[\lambda \frac{\partial E\left[\theta|\theta \geq \theta^*\right]}{\partial b} + (1-\lambda)\frac{\partial E\left[\theta|\theta < \theta^*\right]}{\partial b}\right]$$
$$+ \left[\frac{1-\lambda}{\Phi} - \frac{\lambda}{1-\Phi}\right]\frac{\mu^2 \Phi}{1+\mu^2}\left(1 - \frac{E\left[\theta^2|\theta < \theta^*\right]}{1+\mu^2}\right)$$

because $b|_{\tau\simeq 0} = \mu$. If we allow μ to be large enough, the above expression tends to

$$\chi_\tau(\tau, \theta^*)|_{\tau\simeq 0} = (1-\lambda)\delta g + (2+\beta)\left[\lambda m_h + (1-\lambda)m_l\right] + \left[\frac{1-\lambda}{\Phi} - \frac{\lambda}{1-\Phi}\right]\Phi$$

Since when μ is large the second line in (26) vanishes, $\left[\frac{1-\lambda}{\Phi} - \frac{\lambda}{1-\Phi}\right]\Phi$ tends to zero and the above expression becomes positive. Similarly, $\chi_{\theta^*\tau}$ goes to zero as μ becomes large. Therefore $\frac{\partial\chi}{\partial\tau}|_{\tau\simeq 0} > 0$ when the noise of the test is sufficiently large.

- Proof of Proposition 2. Total differentiation of the first order conditions when μ is constant yields

$$\chi_{\tau\tau}\partial\tau + \chi_{\tau\theta^*}\partial\theta^* = -\chi_{\tau g}\partial g - \chi_{\tau\lambda}\partial\lambda \tag{29}$$

$$\chi_{\theta^*\tau}\partial\tau + \chi_{\theta^*\theta^*}\partial\theta^* = -\chi_{\theta^* g}\partial g - \chi_{\theta^*\lambda}\partial\lambda \tag{30}$$

so that by Cramer's rule we obtain

$$\frac{\partial\tau}{\partial g} = \frac{-\chi_{\tau g}\chi_{\theta^*\theta^*} + \chi_{\theta^* g}\chi_{\tau\theta^*}}{\Delta} \tag{31}$$

where

$$\Delta = \chi_{\tau\tau}\chi_{\theta^*\theta^*} - \chi_{\tau\theta^*}\chi_{\theta^*\tau} \tag{32}$$

is positive if the second order conditions for a maximum hold. The second order conditions also imply that $\chi_{\tau\tau} < 0$ and $\chi_{\theta^*\theta^*} < 0$. Moreover $\chi_{\theta^* g} = 0$ and $\chi_{\tau g} > 0$, which guarantee the result.

References

[1] Acemoglu, D., (2000), Technical Change, Inequality and the Labor Market, *NBER Working paper* no. 7800, Cambrisge, MA.

[2] Aghion, P., Caroli, E. and Penalosa, G., (1999), Inequality and economic Growth: the perspective of the new growth theories, *Journal of Economic Literature*, 1615-1660.

[3] Allen, J. and Barnsley, R., 1993, Streaming and Tiers: The Interaction of Ability, Maturity and Training in Systems with Age Dependent Recursive Selection, *Journal of Human Resources*, 649-659

[4] Anderson,B. and Moore,J., 1979, *Optimal Filtering*, Prentice-Hall,NJ, USA

[5] Argys, L. M., Rees, D. I., & Brewer, D. J. (1996). Detracking America's schools: Equity at zero cost? *Journal of Policy Analysis and Management*, 15(4), 623-645.

[6] Bedard, K., 1997, Educational Streaming, occupational Choices and the Distribution of Wages, mimeo, McMaster University.

[7] Berman, E. and Machin. S., 2000, Skill - biased technology transfer around the world, *Oxford Review of Economic Policy, 16, 3, 12-22*

[8] Betts, J., 1998, The Impact of educational Standards on the Level and Distribution of Earnings, *The American Economic Review*, 88, 1, 266-275.

[9] Bertocchi, G. and Spagat, M., 2003, The Evolution of Modern Education Systems, forthcoming in *Journal of Development Economics.*

[10] Brunello, G. and Giannini, M., 2004a, Stratified or Comprehensive? The Economic Efficiency of School Design, in *Scottish Journal of Political Economy*, 173-194

[11] Brunello, G. and Giannini, M., 2004b, Selective Schools, *Bulletin of Economic Research*, 3, 207-226.

[12] Epple, D. and Romano, R. (1998), Competition between Private and Public Schools, Vouchers and Peer Group Effects, *The American Economic Review*, 33-60.

[13] Epple, R., Newlon, E. and R. Romano, 2002, Ability tracking, school competition and the standardization of educational benefits, *NBER Working Paper* no. 7854.

[14] Galor, O. and Moav, O. (2000), Ability biased Technological Transition, Wage Inequality and Economic Growth, *Journal of Public Economics*, 83, 1-48.

[15] Gamoran, A. (1987), The stratification of high school learning opportunities, *Sociology of Education*, 60, 135-55.

[16] Goux, D. and Maurin, E., (1998), From Education to First Job: The French Case, in Shavit, Y. and Muller, W., *From School to Work*, Oxford University Press.

[17] 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.

[18] Gust, C. and Marquez, J., (2002), International comparisons of productivity growth: the role of information technology and regulatory practice, *International Finance Discussion Paper* #727, Board of Governors of the Federal Reserve System.

[19] Hanushek, E., Klain, J., Markman, J. and Rivkin, S., 2001, Does Peer Ability affect student achievement? *NBER Working Paper* 8502.

[20] Heath, A. and Cheung, S.Y., (1998), Education and Occupation in Britain, in Shavit, Y. and Muller, W., *From School to Work*, Oxford University Press.

[21] Hoxby, M.C., (2001), Peer Effects in the Classroom: Learning from Gender and Race Variation, *NBER Working Paper* 7867.

[22] Katz, L. and Murphy, K., 1992, Changes in relative wages, 1963-1987: supply and demand factors, *The Quarterly Journal of Economics*, 35-78.

[23] Krueger, D. and D.Kumar (2002), Skill specific rather than general education: a reason for US-Europe growth differences?, *NBER working paper* # 9458.

[24] Ishida, H., (1998), Educational Credentials and labor market Entry Outcomes in Japan, in Shavit, Y. and Muller, W., *From School to Work*, Oxford University Press.

[25] Lindbeck, A. and Snower, D., (2000), Multitask Learning and the Reorganization of Work, *Journal of Labor Economics*.

[26] Muller, W., Steinmann, S. and Ell, R., (1998), Education and Labor market Entry in Germany, in Shavit, Y. and Muller, W., *From School to Work*, Oxford University Press.

[27] Nickell, S. and Layard, R., (1999), Labor Market Institutions and economic Performance, in Ashenfelter, O. and Card, D., *Handbook of Labor Economics*, volume 3, North Holland, 3029-80.

[28] Oakes, J., (1994), Educational Matchmaking: Academic and Vocational Tracking in Comprehensive High Schools, RAND, Santa Monica, California.

[29] Shavit, Y. and Muller, W., (1998), *From School to Work*, Oxford University Press, Oxford.

[30] Schnepf, Sylke V., (2002), A sorting hat that falls? The transition from primary to secondary school in Germany, Innocenti Working paper # 92, Florence, Unicef.

[31] Zimmer, R. and Toma, E., (2000), Peer effects in private and public schools across countries, *Journal of Policy Analysis and Management*, 19, 1, 75-92.

[32] Zimmermann, C., (2003), Peer effects in academic outcomes: evidence from a natural experiment, *Review of Economics and Statistics*, 85, 9-23.

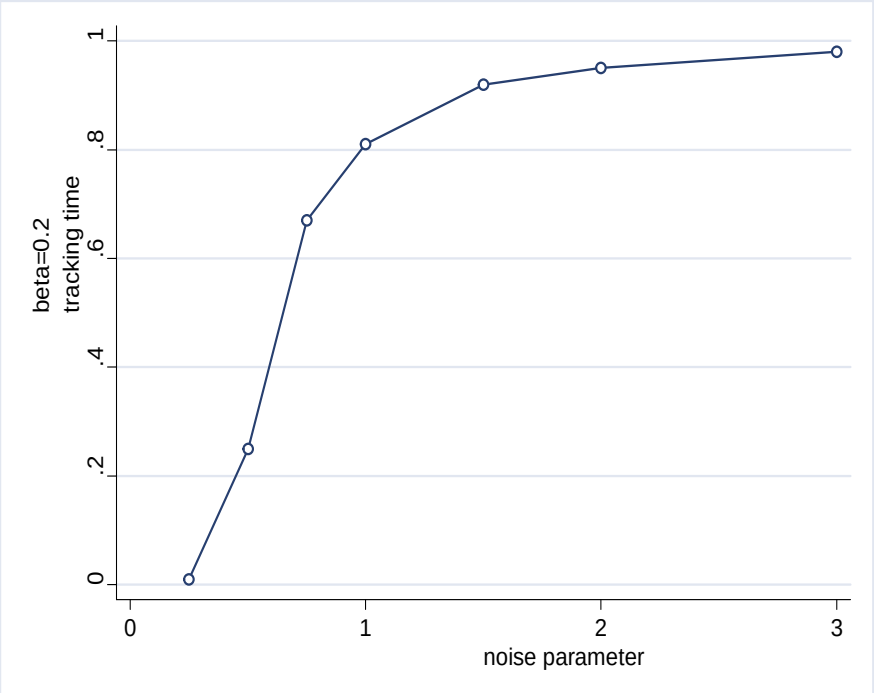

Figure 1: Changes in τ as μ increases

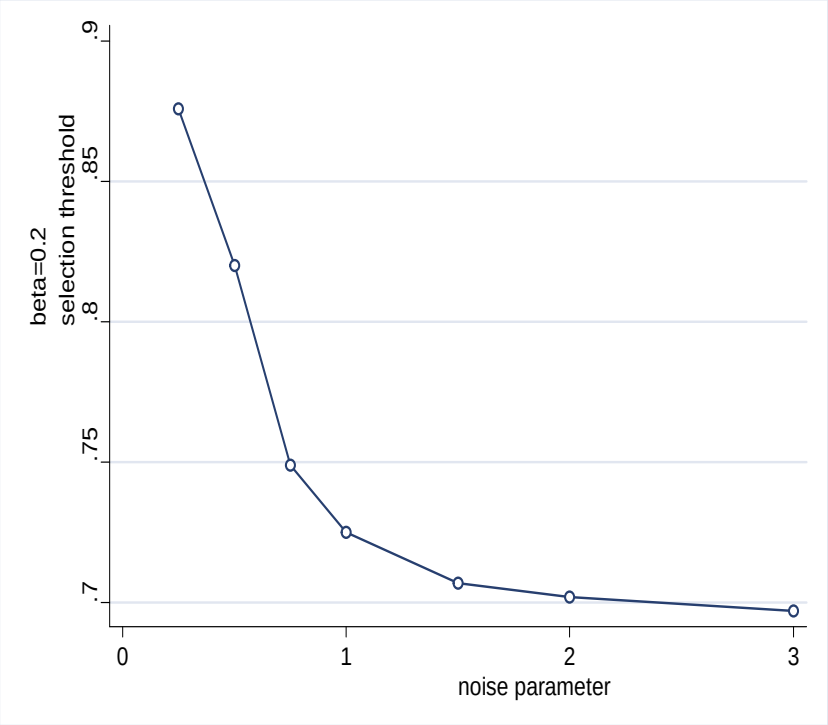

Figure 2: Changes in θ^* as μ increases

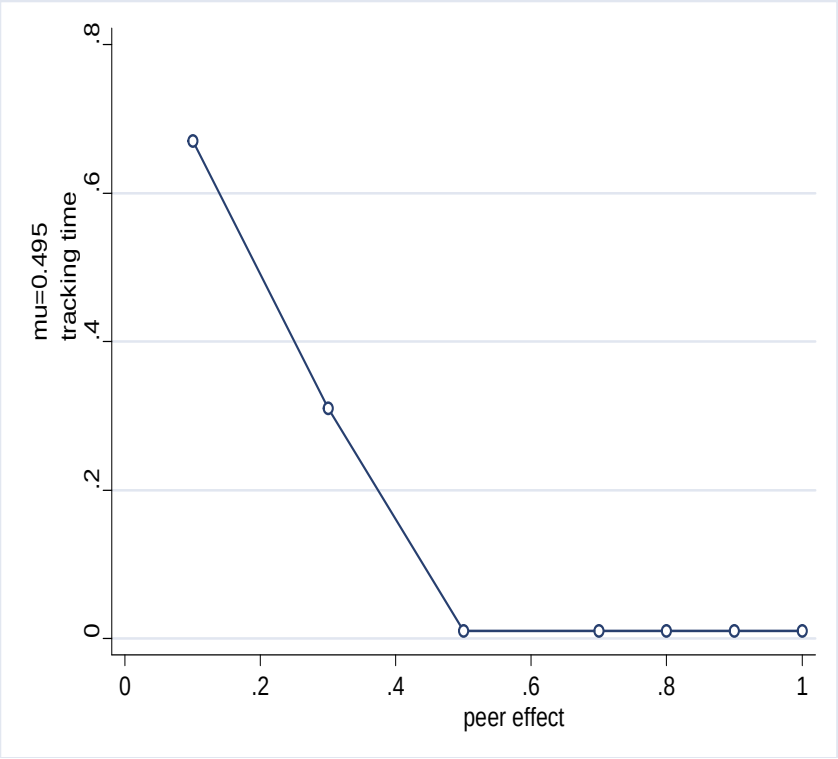

Figure 3: Changes in τ as β increases

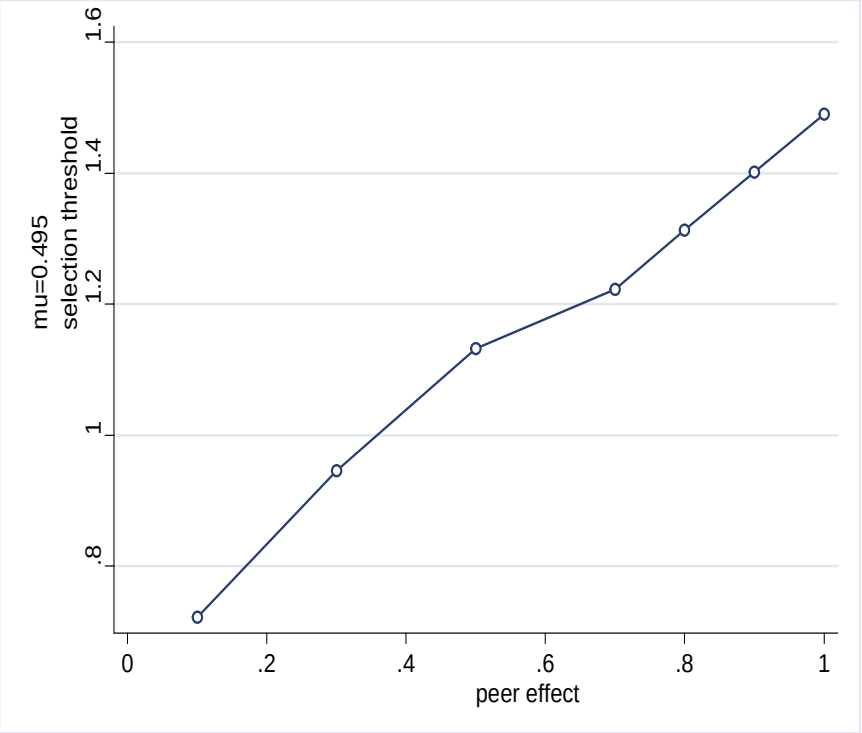

Figure 4: Changes in θ^* as β increases