THE MAJOR DECISION: LABOR MARKET IMPLICATIONS OF THE TIMING OF SPECIALIZATION IN COLLEGE

LUC BRIDET AND MARGARET LEIGHTON

Abstract. College students in the United States choose their major much later than their counterparts in Europe. American colleges also typically allow students to choose when they wish to make their major decision. In this paper we estimate the benefits of such a policy: specifically, whether additional years of multi-disciplinary education help students make a better choice of specialization, and at what cost in foregone specialized human capital. We first document that, in the cross section, students who choose their major later are more likely to change fields on the labor market. We then build and estimate a dynamic model of college education where the optimal timing of specialization reflects a tradeoff between discovering comparative advantage and acquiring occupation-specific skills. Multi-disciplinary education allows students to learn about their comparative advantage, while specialized education is more highly valued in occupations related to that field. Estimates suggest that delaying specialization is informative, although noisy. Working in the field of comparative advantage accounts for up to 20% of a well-matched worker's earnings. While education is transferable across fields with only a 10% penalty, workers who wish to change fields incur a large, one-time cost. We then use these estimates to compare the current college system to one which imposes specialization at college entry. In this counterfactual, the number of workers who switch fields drops from 24% to 20%; however, the share of workers who are not working in the field of their comparative advantage rises substantially, from 23% to 30%. Overall, expected earnings fall by 1.5%.

Date: October 16, 2015.

The results presented in this draft are preliminary. We thank our advisors, Christian Hellwig and Stéphane Straub. We have benefited from discussions with a great number of people, particularly Joseph Altonji, Melissa Tartari and Paul Rodríguez-Lesmes as well as colleagues and seminar participants at the Toulouse School of Economics, Yale, St Andrews, Princeton EconCon, the ENTER Jamboree, the Summer School on Socioeconomic Inequality and Mount Allison University. We are additionally grateful to Joseph Altonji and the National Center for Education Statistics staff for making it possible for us to access to the Baccalaureate and Beyond 93:03 dataset.

Affiliations: University of St Andrews. Contact: mal22@st-andrews.ac.uk; lb222@st-andrews.ac.uk.

1. Introduction

On the first day of college at an American university, many freshmen do not know what field they will concentrate on during their undergraduate studies. Four years later, newly-minted graduates enter the world of work with a new credential and a field of specialization. Despite the dire importance students ascribe to their choice of major, relatively little research has examined the process by which this decision is reached and the implications of constraining that process.

Education systems differ widely in how and when students are allowed to select a field. In many european universities the choice of major is made prior to enrollment and is difficult to adjust thereafter. Other countries, in contrast, are much more forgiving towards the undecided. It is well accepted in the United States that college is a time of self discovery: the exploration of different fields is encouraged and sometimes mandated. In the US majors can be chosen several years into college, and adjusted even later.

The impact of constraining the timing of specialization on eventual labor market outcomes is potentially large. If delayed specialization enables students to make better-informed decision about their field of specialization, the returns to education are affected through two channels. First, better-matched students are more likely to pursue careers in a field related to their studies, thus making better use of their specialized training. Second, workers who are in occupations which are well-suited to their innate talents are likely to be more productive. Education reforms which seek to increase the returns to college education must take account of such effects, particularly those reforms which would narrow the breadth of college education.

Does broad education help students discover their idiosyncratic talents? And if so, does the accuracy of this match translate into better labor market outcomes? This paper provides empirical answers to these questions. We first document the positive cross-sectional correlation between the timing of specialization and the probability of working in an occupation unrelated to college major. This finding suggests that selection into specialization needs to be accounted for, and motivates our structural model.

In our model, the optimal timing of specialization reflects a trade-off between identifying one's field of comparative advantage and acquiring specialized skills. Each agent is best suited to one field, but the identity of that field is initially unknown. Taking courses in many fields simultaneously provides agents with information about their comparative advantage. College course choices solve an optimal experimentation (bandit) problem: a student may choose multi-disciplinary education, where he acquires skills and receives information about his match to different fields; alternatively, he may choose specialized education, which conveys field-specific skills at a faster rate but does not provide such information.

Students update their beliefs about their comparative advantage by filtering a diffusion process in continuous time. Their course choices follow a stopping rule: students start by enrolling in multi-disciplinary education, where they learn about their field of comparative advantage. Once their confidence level is sufficiently high – the belief that they belong in either field being sufficiently close to 1 – they specialize. This bandit problem is not stationary: over time, as agents remaining in the mixed-education stream acquire skills, the value of their foregone wages rises and the expected length of additional specialized studies diminishes. This reduction in the length of specialized education depresses the value of current information and makes the agents less willing to experiment. Agents optimally lower the confidence level they require in order to specialize.

This property of optimal experimentation has important cross-sectional implications. Agents whose beliefs process drifts the fastest, and who therefore specialize early, are more likely to specialize in the field of their comparative advantage. Agents whose beliefs process remains close to their prior longer take their specialization decision on the basis of weaker information. Late specializers are therefore more likely to choose wrongly. Since these same individuals spent most of their studies in the multi-disciplinary stream, their education is also relatively more transferable. The model therefore delivers our reduced-form result: compared to early specializers, late specializers are more likely to change occupations.

To separate the contribution of self-selection from the lock-in effect of early specialization, the model is estimated structurally. Using data from a panel of college graduates, we estimate the parameters of the model through simulated method of moments. Detailed transcript data allow us to construct a proxy for the timing of specialization based on the course mix a student chooses in each period. We then simulate the model, selecting parameters to match the observed timing of specialization, occupation field choice, and wages.

Our estimated parameter values reveal that the benefits of flexible specialization are large. Time spent in mixed-discipline studies is informative, although imperfectly so. One year of exploratory college courses is as informative about comparative advantage as the entire pre-college period; nevertheless, despite considerable time spent acquiring information, only 53% of students major in the field of their comparative advantage. Furthermore, being type-matched to one's occupation is well-rewarded – workers employed in the field of their comparative advantage earn a 20% wage premium. While our estimates indicate a large, one-time switching cost equivalent to 1.5 years of wages, field-specific education is highly transferable: ten years after graduation, out-of-field education is remunerated at 90% the rate of field-related schooling. While the parameters are estimated simultaneously, variation in the timing of specialization and occupation choice appears to drive identification of the precision of signals, both prior to and during college. These values in turn, along with the earnings moments, pin down the parameters governing returns and switching costs.

These parameter values allow us to compare the current college system to a Europe-style counterfactual where specialization is imposed at college entry. While such a policy will be welfare-reducing by construction, imposing a timing of specialization may be a necessary practical or cost-saving measure on the part of an education system. Our estimates suggest that such a policy would have non-trivial consequences. We predict a modest change (a reduction from 24% to 20%) in the proportion of agents that pursue careers outside of their field of specialization, as the lock-in effect of early specialization counteracts the poorer information students have at the start of college. The change in the allocation of individuals across occupations is more substantial. In the early-specialization counterfactual, 70% of workers are employed in the field of their ex-ante comparative advantage, down from 77% in the benchmark. The average welfare cost of such a policy would be equivalent to reducing education stocks by half a year of specialized studies, or to reducing expected discounted lifetime earnings by 1.5%.

This paper contributes to the small literature on the timing of specialization in higher education by introducing a model of endogenous timing choice. Malamud (2010, 2011) assumes, just like we do, that broad education is informative about horizontal match characteristics. He compares labor market outcomes of early-specializing students (in England) to late-specializing students (in Scotland). Comparing these two cohorts, which face specialization imposed at different times, he finds that late specializers are less likely to choose occupations unrelated to their studies (2011 paper). Our model suggests that these results are reversed in a context where the timing of specialization is chosen by each student; a prediction which is borne out in our data. Flexible specialization times are typical of American colleges; our model therefore allows us to explore how the timing of specialization affects labor market outcomes in the US context.

Bordon and Fu (2013) estimate an equilibrium admissions model to explore the impact of unbundling college choice and major choice in Chile, where the current system requires students to apply to a college-major pair. The authors estimate the impact of alternate systems on college retention and peer quality, finding that more flexible policies are welfare improving so long as the relative returns to specialized education are not too high. Our model, which does not include college choice, allows us to estimate the returns to specialized education and the information value of unspecialized studies simultaneously.

This work also relates to the literature that integrates information revelation into models of college major choice. Altonji (1993) introduces a sequential model of college education where both aptitude and completion probabilities are ex-ante uncertain. Subsequent papers generally adhered to Altonji's approach of sorting majors by difficulty. Arcidiacono (2004) estimates a rich structural model with four majors, using grades as signals of ability and allowing college students to change majors, or drop out, between the early and later periods of college. While his results show that students do sort across majors based on ability, wage differencials between majors – large as they are – are not sufficient to explain this sorting. Taking advantage of the expectations data in their Berea Panel Study, Stinebrickner and Stinebrickner (2014), find that students switch out of majors in math and science as a result of learning that they perform less well in those subjects than they had anticipated. These findings are supported by Arcidiacono et al. (2012), Arcidiacono et al. (2013b) and Ost (2010).[1] We diverge from these papers by focusing on horizontal type discovery. We find that purely horizontal considerations matter a great deal: within our subset of majors – all of which are relatively high earning – a graduate working in the field of his comparative advantage earns significantly higher wages than an individual with the same education who is not well matched to his occupation.

Two recent papers extend the literature in this direction. Kinsler and Pavan (forthcoming) model ability as a two-dimensional vector – loosely corresponding to math and verbal skills – which is only fully revealed on the labor market. Silos and Smith (forthcoming) allow college students to choose their investment in three skills – quantitative, humanities and social sciences – which are ultimately rewarded to different degrees in different occupations. Residual uncertainty about agents' match to different occupations is resolved after a probationary period working in that particular job. In contrast to these papers, we are concerned with how education choices can help resolve this uncertainty prior to the labor market. This emphasis allows us to speak directly to education policy, unpacking the returns to higher education and informing the debate on college reform.

Finally, this paper speaks to the literature on the returns to education breadth. Dolton and Vignoles (2002) include secondary school course diversity in earnings equations. In their UK data, breadth of courses, at A-level or O-level, are shown to have insignificant effects on earnings. Using data from a European post-secondary graduate survey, Heijke and Meng (2006) find that graduates from programs that provide both academic and discipline-specific competencies produce less performant workers. There are many reasons to expect that selection into broad studies is not random, making a causal interpretation of these findings problematic.[2] Our paper contributes to this literature by exploring one specific channel through which broad curricula could be beneficial: by facilitating better field choices. In doing

[1]We do not treat college attrition in this paper; however, it remains an important focus of this literature. See Stinebrickner and Stinebrickner (forthcoming), Arcidiacono et al. (2013a) and Trachter (forthcoming), among others.

[2]To the best of our knowledge, no study has estimated the returns to breadth in education in a setting where education breadth is plausibly exogenous. Joensen and Nielsen (2009) come perhaps the closest, although the change in breadth they consider is quite marginal: taking advantage of a policy experiment in Denmark, which allowed students to take advanced math without taking advanced physics (taking advanced chemistry instead), the authors estimate the returns to advanced coursework in math.

so, we shed light on why broad curricula – even if they improve field choice – might be unrelated to labor market outcomes in the cross section.

In the following section, we introduce the data and describe how we determine the timing of specialization. Section 3 presents the model: first a two-type, continuous time version, and then the N-type, discrete time model we use in our simulations. The estimation procedure and the results are outlined in Section 4, while Section 5 discusses the policy experiments. Section 6 describes robustness exercises and extensions and Section 7 concludes.

2. Data

2.1. **Overview.** The data we use in this paper come from the restricted version of the Baccalaureate & Beyond 93:03 dataset.[3] This 10-year panel follows approximately 10,000 students who earned a bachelor's degree[4] from an American institution in the 1992-1993 academic year. Three follow-ups are carried out (one, four and ten years later), during which labor market and further education variables are collected.

Two features make this dataset ideal for our purposes: the policy context and the level of detail in the education and labor market variables. We require data in which there is some variation in the timing of specialization among students who graduate with the same major and degree. Given that college students in the US have considerable discretion over their course choices (and indeed, take relatively few courses in their major field), they have some scope to choose when they choose their major. This motivates our use of US data.

Our more restrictive requirement is that the panel include data on all courses taken in college, and the date at which these courses were taken. Timing of specialization, although central to our story, is not a well-established variable[5]

[3]Referred to in the following as B&B93:03. Dataset sponsored by the National Center for Education Statistics, U.S. Department of Education (for more information, see Wine et al. (2005)).

[4]A bachelor's degree is a 4-year undergraduate degree.

[5]The closest measurable variable might be the declaration of a major, as used in Bradley (2012)'s study of major choice during recessions. For a number of reasons this is not fully satisfactory for our purposes: first, different universities may require students to declare a major at different times, and this information is not in our data; second, if students are required to declare a major in order to, for example, register for their second year courses, this declaration is not necessarily an active form of specialization. Given that it is generally easy to change majors, and that major choice does not usually constrain course choices (at least early in college), the declaration of a

and we rely on the sequence of course choices to derive a proxy. Few datasets include both labor market outcomes and detailed course data.

We make several important sample restrictions. First, since we estimate information revelation in college, we retain only those students who graduated between the ages of 21 and 23:[6] older students are likely to have acquired information through other means, such as by working. Second, as our model supposes that students anticipate specific college courses to be useful on the labor market, we restrict our analysis to applied fields.[7] Finally, due to differences in pre-college ability across majors, as well the associated earnings differences on the labor market, we retain only the high-earning quantitative majors in our estimation.[8] Details of the sample construction can be found in Appendix A.

2.2. **Timing of specialization.** We derive a proxy for specialization using the mix of courses a student chooses in each period of school. In contrast with some panels that follow students through college,[9] our data does not include self-reported college major at multiple periods of time. Instead, we observe the entire sequence of course choices, as they appear on the final college transcript. Our specialization proxy in an indicator variable based on the share of credits chosen in the eventual major field of study. The intuition behind this is that students have a stronger incentive to take courses in the field in which they will eventually work. Once a student has settled on a particular field, the expected return from taking related courses rises.

Table 1 lists the total credits earned by students in each major, along with the share of those credits taken in that field. American bachelor degrees are evidently quite broad: even with coarse major categories, the share of credits taken in-major rarely exceeds 50%. We therefore do not expect students to commit themselves

major could be little more than a statement about the field a students thinks it is *most likely* she will pursue. Finally, this information is not available in the B&B93:03 dataset.

[6]These students in general took between 3 and 6 years to complete their degree.

[7]We are by no means the first to make a distinction between majors based on their links to the labor market. To cite two of many examples, Saniter and Siedler (2014) distinguish between fields with a strong versus a weak labor market orientation in German data, based on whether or not the field of study leads to a particular profession. Using Canadian data, Finnie (2002) finds differences in the early labor market outcomes of graduates from 'applied' fields versus those with majors in 'softer' subjects.

[8]These differences are explored in depth by Arcidiacono (2004); the distribution of SAT or ACT scores across majors for our sample is given in Figure 11 in Appendix A. We retain those majors with more mass in the upper two quartiles than in the lower two.

[9]For instance the NLS72, used by Arcidiacono (2004).

full-time to their major once they have specialized[10] – only that courses will be chosen differently after specialization than before.

TABLE 1. Credits: mean total and specialization by major

Major	**Total credits** (SD)	**In-major share**
Science	127 (13)	0.51
Engineering	135 (12)	0.43
Business & Econ	126 (11)	0.43

Source: B&B93:03, sample restrictions described in section A.1. Note: a bachelor's degree requires 120 credits.

We derive the timing of specialization as follows. Courses are first associated with a major:[11] this allows us to calculate the total credits earned in each field for each academic year.[12] The timing of specialization is then defined as the term in which the share of credits taken in the eventual major field of study exceeds a given threshold.

To choose an appropriate threshold, we must take into account two factors: how strictly to define specialization, and how to treat differences across majors. If the threshold is too slack, it is difficult to justify that we are capturing a genuine change in course-choosing behavior; on the other hand, given the low number of credits students take in their major field, a high threshold results in many students 'never' specializing, despite the fact that they successfully graduate with a major.[13]

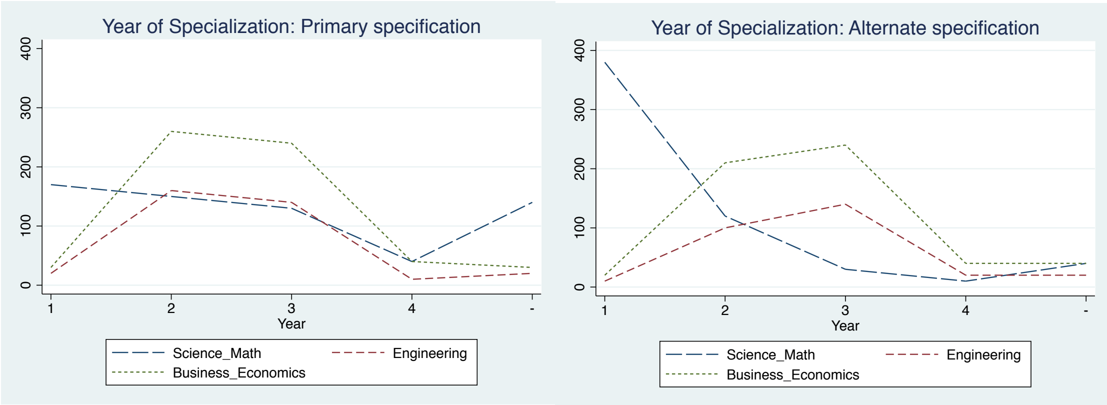
The major categories retained for analysis – while common and intuitive – are not necessarily aggregated in a similar fashion. Consider, for example, a major in science and a major in engineering. The field of science includes biology, physics, chemistry and math, and so courses in all these fields will be coded as science courses. In contrast, the field of engineering is relative narrow, since it includes

[10]There could be many reasons for this: it may not be possible to take a full load of courses in that field, due to missing course prerequisites or simply a shortage of courses at a given level; it may not be desirable to do so, particularly if advanced level-courses are more difficult than introductory courses in other fields; it may not even be permitted within the confines of the bachelor degree program, as many institutions impose a minimum number of courses to be taken in fields different from one's major.

[11]In the dataset, there are approximately 1000 unique course codes. We attribute each of these to one of 14 coarse major categories, several of which are later aggregated (see Table 19 in the Appendix).

[12]Courses vary in how long and intensive they are. The dataset includes a conversion metric for each course, translating the credit units attributed by the degree-granting institution into standard credit equivalents based a 120-credit degree.

[13]Graphical representations of sample retention and timing of specialization for increasingly strict thresholds are presented in Appendix A.3.



Major-specific threshold 40% threshold

Figure 1: Timing of specialization: major-specific threshold vs. constant threshold

only course which are catalogued as 'engineering'. To address this issue, we adopt a data-driven approach to defining the threshold. Specifically, we choose a threshold for each major such that 50% of students who graduate in that field specialize by the end of their second year. Students who never reach the threshold, but who nevertheless graduate with a major in that field, are assumed to have specialized at the very end of college.

Figure 1 shows the distribution of the timing of specialization, for both our primary specification and for a universal 40% concentration threshold.[14] The two specifications give a similar picture for Engineering and Business & Economics majors, although the timing of specialization is slightly earlier using the major-specific threshold. Science & Math majors display two important differences. First, a 40% threshold is quite low for this group of majors: a vast majority of students meet this threshold in their first year.[15] Second, even using a major-specific threshold, students who complete a major in Science & Math tend to specialize earlier than students in other majors. This feature is robust to all of the timing of specialization specifications we have developed (see Appendix A.3 for a discussion of these approaches).

2.3. **Reduced-form evidence.** Table 2 presents a reduced form regression of the timing of specialization on the probability or working in an occupation related to

[14]This approach attributes specialization to the year in which the concentration of major-related courses meets or exceeds 40%.

[15]As discussed above, this is not surprising. A student taking 5 courses a semester must only take 2 of these in science, math or computer science to reach a 40% threshold.

one's field of study, ten years after college. While Malamud (2011) finds that early-specializing English students are more likely to change fields on the labor market than late-specializing Scottish students, our data display the opposite result: late specializers are more likely to work in a field unrelated to their studies. Restricting to our core sample of quantitative graduates, delaying specialization by one year is associated with a 1.5% decrease in the probability of working in an occupation related to one's major.

Our cross-sectional findings are echoed elsewhere. Using US data, Silos and Smith (forthcoming) also find that students with less specialized education stocks are more likely to switch occupations. Their concept of specialization, the closeness of a student's skill bundle to the average skill bundle of a given occupation, is different from ours; however, the concept of hedging through skill diversification is closely related. In a similar vein, Borghans and Golsteyn (2007) estimate a model of occupation changes where human capital is imperfectly transferable. Using Dutch data, in which almost 30% of graduates were working in an unrelated field 3 years after college, they find that higher skill transferability is associated with a greater probability that a graduate who regrets his field of study will switch to an occupation in a different field.

TABLE 2. Match probability

	Probability of working in related field		
	All majors	Quantitative	Non-quantitative
Timing	-0.0376***	-0.0154**	-0.0244**
	(0.000)	(0.048)	(0.027)
Controls	X	X	X
R^2	0.381	0.527	0.771
adj. R^2	0.375	0.522	0.765
Sample size	2110	1560	550

Source: B&B93:03, sample restrictions described in section A.1. P-values in parentheses; * $p < 0.10$, ** $p < 0.05$, ***. Sample sizes rounded to the nearest 10. Occupation is observed 10 years after graduation. Timing is the primary timing of specialization variable (see Section 2.2); match refers to the relation between field of study and field of work. Controls are major and occupation dummies.

Table 3 presents results of a regression of log income on the timing of specialization, controlling for field of study and occupation. The coefficient on timing of specialization in the main regression is small and not significant. While the contexts under consideration are quite different, this result is consistent with Dolton and Vignoles (2002)'s finding that curriculum breadth is unrelated to earnings.[16]

The absence of a trend in the cross sections hides interesting subgroup effects. When the sample is restricted to those individuals who switched fields on the labor market, timing becomes positive and significant: late specialization is associated with higher wages. If we consider only those individuals who are working in the field of their major, the coefficient on timing is negative but not significant.

TABLE 3. Log income

	Log income		
	All	Matched	Not matched
Timing	0.00734	-0.0131	0.0323*
	(0.527)	(0.411)	(0.063)
Controls	X	X	X
R^2	0.189	0.191	0.219
adj. R^2	0.182	0.185	0.203
Sample size	2100	1220	880

Source: B&B93:03, sample restrictions described in section A.1. P-values in parentheses; * $p < 0.10$, ** $p < 0.05$, ***. Sample sizes rounded to the nearest 10. Timing is the primary timing of specialization variable (see Section 2.2); match refers to the relation between field of study and field of work. Controls are major and occupation dummies.

Can these results be consistent with multi-disciplinary studies helping individuals discover their comparative advantage? A naive reading of Table 2 could conclude the opposite: after all, more time spent in broad education is associated with a higher probability of changing fields. If the timing of specialization is endogenous, however, the issue of selection looms large: those who are quite confident about their comparative advantage may opt out of multi-disciplinary studies at an early stage, especially if specialized education is more highly rewarded in their intended field of work. This selection pressure is intensified by the fact that the opportunity cost of studies is rises the longer a students spends in school. We are therefore unable to conclude on the learning value of multi-disciplinary education from cross-sectional regressions on the probability of switching fields.

The results in Tables 2 and 3 motivate us to develop a model where the timing of specialization is endogenous to the student's confidence about his comparative

[16]To see why, note that since late specializers will spend more of their degree taking a broad range of courses, the timing of specialization is correlated with college course breadth: late specializers chose broader curricula.

advantage. In Table 2 we observe a positive correlation between the timing of specialization and the probability of working in an occupation related to field of study. Table 3 shows that, among those workers who change fields, late specializers have higher earnings than early specializers. These results are consistent with the imperfect transferability of skills across fields: late specializers have a more portable skill set, and are therefore able to change fields more easily. They are also consistent with the gradual and imperfect revelation of information about students' comparative advantage, through a process of selection described in detail below. These two channels are indistinguishable in reduced-form evidence: to identify them separately, we will need to estimate the model structurally.

3. Model

3.1. **2-Type Continuous Time Model.** In this section we describe a two-field, continuous-time version of the model, which enables use to arrive quickly at the optimal policy and to introduce the parameters of interest. Restricting to two types helps convey the intuition and makes the learning process most transparent. The model that we estimate – with N-types and discrete periods – is described in Section 3.2.

3.1.1. *Agents.* There are two fields of work ($f \in \{s, a\}$), with two corresponding subjects taught at school, so human capital is a two-dimensional state variable $e = (e_S, e_A)$. Agents are born at date $t = 0$ with human capital $(e_S, e_A) = (0, 0)$, have an infinite lifetime and discount the future at rate $r > 0$. They have one of two possible comparative advantages (type $\theta \in \{S, A\}$).

3.1.2. *Choice set and education tenures.* At each point in time, agents can enter the labor force, follow specialized studies in subject S or in subject A, or follow multi-subject studies M.[17] The laws of motion of human capital are respectively $(\dot{e_S}, \dot{e_A}) \in \{(0,0), (1,0), (0,1), (1/2, 1/2)\}$. Since we assume that uncertainty is resolved upon entry on the labor market, we do not allow agents to return to school once they have started working.

[17]The idea that education has both general and specialized segments is echoed by Altonji et al. (2012), who explicitly model college as two decision periods: one where the student takes many courses, and a second where they choose a major.

3.1.3. *Information acquisition.* The type θ is initially unknown to the agent, who enters date 0 with a prior belief $p_0 = \mathbb{P}[\theta = S|\mathcal{F}_0]$. The prior p_0 may be correlated with θ, and thus reflects information acquired prior to higher education.[18] Mixed-discipline education has an informational benefit: provided that he is engaged in multi-subject education for a short time interval dt, the agent observes an informative signal about his type, modeled as a diffusion $\tilde{Y}$ with type-dependent drift. [19] Filtering this observation allows him to update his estimate $p(t) = \mathbb{P}[\theta = S|\mathcal{F}_t]$ following Bayes' rule. The learning technology is characterized by a single signal-to-noise ratio ϕ, such that the agent correctly forecasts his own belief as a pure drift-less diffusion (Beliefs derived from Bayes' rule are always martingales):

$$\mathrm{d}p(t) = p(t)\left(1 - p(t)\right)\phi \mathrm{d}\tilde{W}(t) \tag{1}$$

The volatility of beliefs, roughly equivalent to to the speed of learning, is highest the closer the agent is to indecision ($p(t) = 1/2$) and the higher the signal-to-noise ratio. Learning is informative, so conditionally on type the belief process of $\theta = S$ types drifts upwards (similarly type A-agents' beliefs drift downwards):

$$dp(t) = \left(1 - p(t)\right)^2 p(t)\phi^2 \mathrm{d}t + p(t)\left(1 - p(t)\right)\phi \mathrm{d}\tilde{W}(t) \tag{2}$$

A more informative signal helps the belief converge towards the truth faster by increasing the drift of the process, but also raises its volatility.

This signal technology is characterized by gradual and incomplete learning: unlike in Poisson bandit models, agents never learn their type completely but instead continuously and gradually update their beliefs. An important property of this learning technology is that agents do not necessarily acquire better information over time.[20]

[18] $\mathcal{F}_t$ denotes time-t filtrations which summarize the agent's information accumulated up to date t. In the binary case, a single scalar $\mathbb{P}[\theta = S]$ fully characterizes the beliefs of the agent at any point in time. With N fields, a belief is identified with an element of the N-dimensional simplex.

[19] This standard model of gradual and continuous learning is used by Felli and Harris (1996); Moscarini and Smith (2001), among others.

[20] By contrast, filtering a sequence of Gaussian signals with unknown mean results in a Gaussian posterior belief about the mean, and the variance of the belief decreases deterministically. In that sense, agents necessarily become better informed over time, whereas in our setup, once an agent returns to a previously-held belief, it is as if any information received in the meantime had been wasted.

3.1.4. *Payoffs.* Payoffs accrue to agents from two sources: wages and benefits while working, and flow payoffs in school. During their education tenure, agents earn constant, possibly negative flow payoffs z that reflect tuition and overall enjoyment of studies. At the end of the education tenure, agents choose a field of work which may differ from their field of specialization. We assume that the labor market rewards within-field education more than out-of-field education and to describe labor market returns, we first define an effective stock of skills for each field:

$$\epsilon_S = e_S + \beta e_A,\ \epsilon_A = e_A + \beta e_S. \tag{3}$$

While a unit of human capital in subject s contributes one unit to ϵ_S, an equivalent investment in subject a contributes only β to the stock of skills applicable to field s. We call β a transferability parameter and assume $\beta \leq 1$: skills acquired in one field are only partially transferable. The flow wage when working in sector f is given by:

$$R(\epsilon_f) + \mathbb{1}_{\theta=f} P\ , \tag{4}$$

where $R(.)$ is an increasing and concave function and has the dimension of flow utility. Returns are therefore increasing in within-field skills ($R'(\epsilon_f)$ at the margin) and out-of-field skills ($\beta R'(\epsilon_f)$ at the margin), with an additive premium P for working in the field of one's comparative advantage ($\theta = f$).

3.1.5. *Optimal policy and testable implications.* The optimal policy is characterized by optimal stopping times. Agents begin in multi-subject education, specialize their studies at time $\tau_1 \geq 0$, and proceed to the labor market at time $\tau_2 \geq \tau_1$.

It can be optimal for agents to choose $\tau_1 = 0$ if the signal is totally uninformative ($\phi = 0$) and more generally, experimentation may not be worthwhile. The length of the specialization period may also vanish ($\tau_2 = \tau_1$) if the return function does not depend on the level of effective human capital ($R'(\epsilon) = 0$), so that agents transfer directly from multi-subject education into the labor market. Figure 2 represents a regular case in which $\tau_1 > 0$ and $\tau_2 - \tau_1 > 0$.

The agent's decisions are presented recursively over the following paragraphs. Starting from the choice of occupation, we discuss the determination of the time

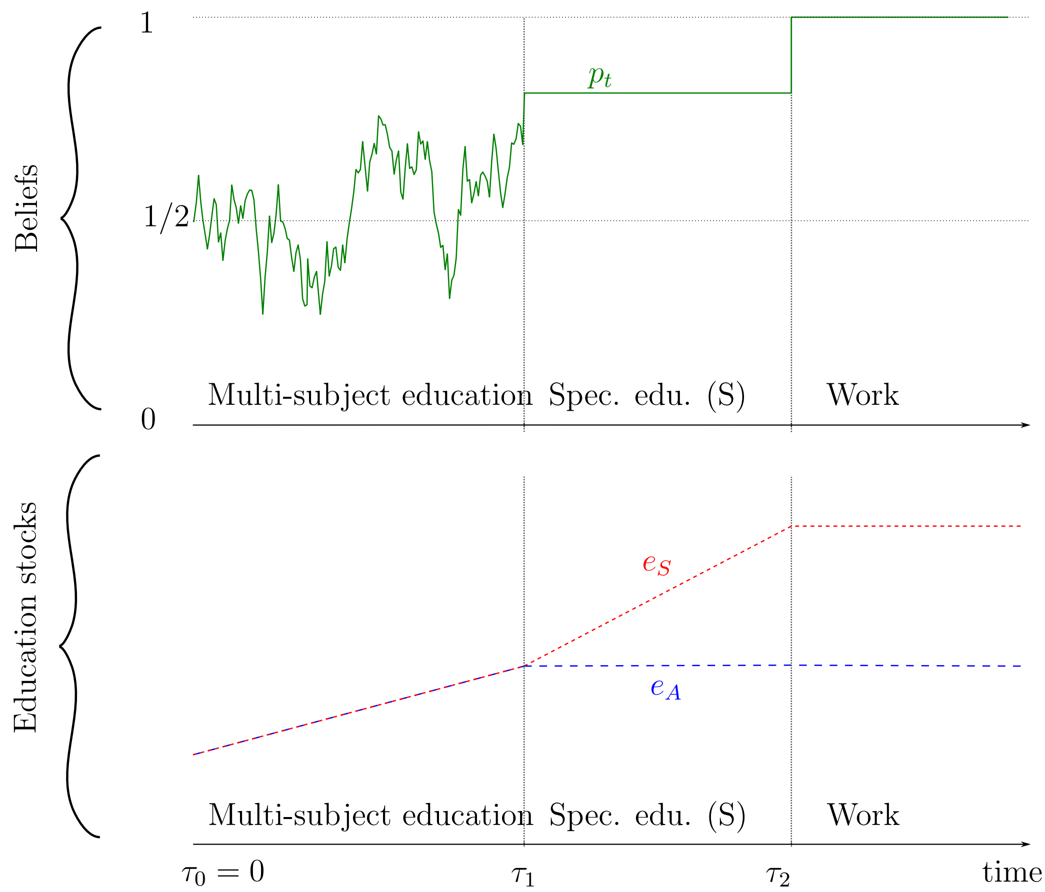


Figure 2: **Sample education tenure.** *An agent of type* $\theta = S$ *with prior* $1/2$ *starts by following multi-subject education, both education stocks increase at rate* $1/2$ *and the belief process* $p(t) = \mathbb{P}[\theta = S|\mathcal{F}_t]$ *obeys equation* (2) *(hence drifts upwards). At* $t = \tau_1$ *the belief is sufficiently high that the agent exits into specialized education in subject* s*, where* e_S *increases at rate* 1 *while* e_A *stays constant. The agent does not update his belief about* θ *until entry in the labor market at* $t = \tau_2$*, when uncertainty resolves.*

spent in specialized studies, $\tau_2 - \tau_1$ and ultimately, the optimal timing of specialization, τ_1.

Determination of τ_2. The agent's type is revealed upon entry into the labor market, after which he may switch fields. In this exposition, the switching of fields comes at no additional cost. Consider an agent who has spent H periods specializing in subject S,[21] and is endowed with human capital stocks $(e_S, e_A) = (H + \tau_1/2, \tau_1/2)$ (where τ_1 is the length of the mixed-studies tenure that predated specialization).

[21]Symmetric expressions obtain for subject A, with the important caveat that the confidence threshold is then $1 - p$.

The agent's expected payoff at the end of his education tenure is Y_s, such that:

$$\begin{aligned} & rY_s\left(p, e_S, e_A\right) = \\ & \qquad p\left(R_s\left(e_S+\beta e_A\right)+P\right) \\ & \qquad +(1-p) Max\left\{R_s\left(e_S+\beta e_A\right), R_a\left(e_A+\beta e_S\right)+P\right\}. \end{aligned} \tag{5}$$

Indeed, with probability p the agent is truly of type S, in which case he earns the premium P when working in field S. With probability $1-p$, he is in fact type A: he may then choose either to remain employed in field s and forgo the matching premium, or to switch fields. If he switches fields he does earn the premium; however, he suffers a transferability penalty since in his case $e_S+\beta e_A > \beta e_S + e_A$. Equation 5 allows us to define the optimal length of specialized studies, which solves the deterministic program:

$$V_S\left(p, e_S, e_A\right) = \underset{\{H \geq 0\}}{Max} \int_0^H \exp\{-rt\} z dt + \exp\{-rH\} Y_s\left(p, e_S+H, e_A\right). \tag{6}$$

It follows that a student who specialized in subject S will continue in that field until the marginal value of studies (flow value and increment in future earnings) falls below the opportunity cost of remaining at school (the flow-equivalent value of working). This is formalized in the following first-order condition:

$$z + \frac{\partial Y_s\left(p, e_S+H, e_A\right)}{\partial e_S} \leq rY_s\left(p, e_S+H, e_A\right), \text{ with equality if } H > 0. \tag{7}$$

Since this tradeoff is known at the time of specialization, the duration of specialized studies ($H = \tau_2 - \tau_1$) is deterministic, as no uncertainty emerges during that time. Finally, since either field can be chosen as specialization, we can define a value of specializing $V = \max(V_A, V_S)$.

Determination of τ_1. Due to uncertainty about the beliefs process $p(t)$, τ_1 is not deterministic. While human capital accumulates steadily during multi-subject education (education stocks at time t are $(t/2, t/2)$), beliefs evolve stochastically. The expected value writes as:

$$\mathbb{E}\left[\int_0^{\tau_1} \exp\{-rt\} z dt + \exp\{-r\tau_1\} V\left(p(\tau_1), \tau_1/2, \tau_1/2\right)\right] \tag{8}$$

The expectation is taken over paths of the beliefs process $p(t)$ and τ_1 is chosen optimally.

For each date t we can define two boundary beliefs: one close to certainty in type-S (call it $p_s(t)$), and one close to certainty in type-A ($p_a(t)$). If, at time t, the agent's beliefs exceed the boundary in either direction, it is optimal for him to specialize. τ_1 is then the first random time such that either $p(t) \geq p_s(t)$ or $p(t) \leq p_a(t)$. Figure 3 illustrates the boundaries, overlaid with a simulated belief path.

$$\tau_1 = Min \left\{ t \geq 0, p(t) \geq p_s(t) \text{ or } p(t) \leq p_a(t) \right\} \tag{9}$$

Also pictured are the densities of exit times: the optimal specialization policy induces a distribution of exit times, with cumulative distribution $F_{ET}(t, \theta) = \mathbb{P}\left[\tau_1 \leq t | \theta\right]$. Similarly, by conditioning on the field of specialization and the agent's type, we can define a distribution of correct specialization times. For example, for $\theta = S$ agents, we can define $F_{CET}(t, \theta) = \mathbb{P}\left[\tau_1 \leq t \cap p(\tau_1) = p_s(\tau_1) | \theta = S\right]$. The limit of $F_{CET}(t, \theta)$ as t grows large is the proportion of agents who specialize in the appropriate field.

3.1.6. *Properties of the optimal policy.* If we impose symmetry and assume that the returns to effective skills are linear, we can show that the specialization boundaries are bounded away from 0 and converge to 0 in finite time, and simulations indicate that optimal boundaries are monotonic. In Appendix B we illustrate the properties of simulated optimal policies and show how parameter changes affect the empirical predictions of the model; which shed light on the sources of identification in Section 4.

Monotonic boundary beliefs – with $p_s(t)$ decreasing and $p_a(t)$ increasing – have important empirical implications. They imply that early and late specializers have two important differences: not only do they accumulate different stocks of human capital, they also exit education with different probabilities of having specialized in the field of their comparative advantage. While all agents have the same information technology and use the same decision rules, differences in the idiosyncratic noise cause some agents' belief processes to drift rapidly towards the boundary, prompting them to specialize early relative to their peers. Consider two students

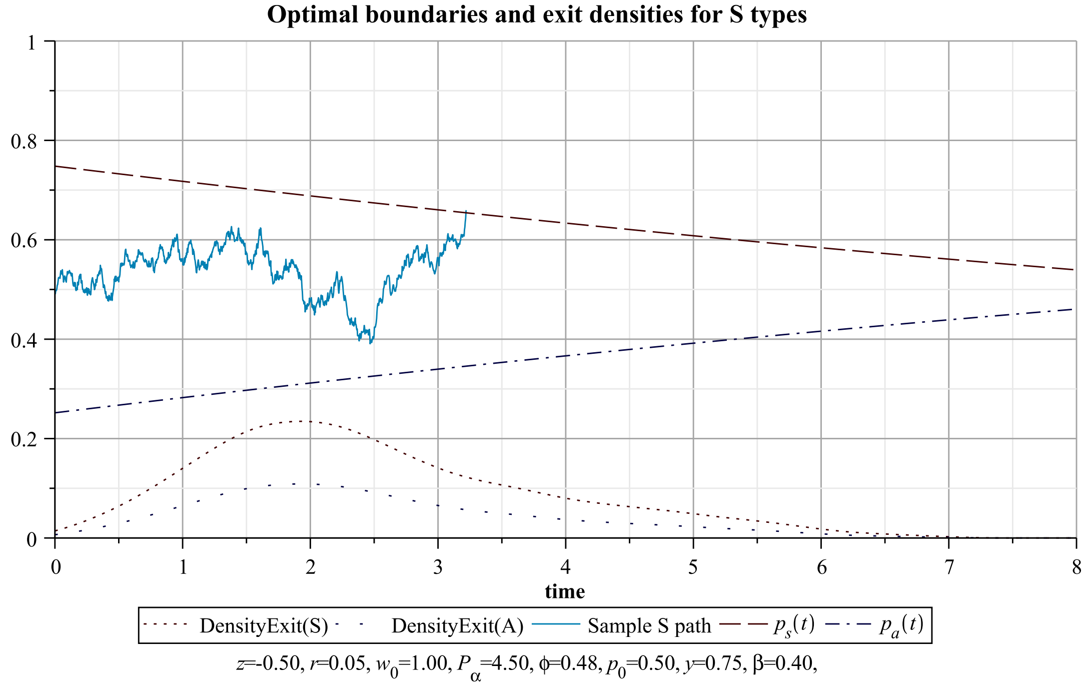


Figure 3: Optimal Boundaries and density of specialization times. *The agent starts specialized education once the belief process $p(t)$ escapes the interval* $(p_a(t), p_s(t))$. This particular sample path leads to specialization time $\tau_1 \approx 3.15$ and corresponds to correct specialization (in subject S).

that choose specialize in field S, the first at time t and the second at time T (with $t < T$). The first student specializes at a point where his belief satisfies $p(t) = p_s(t)$ while the second one's cutoff is reached as $p(T) = p_s(T)$. Since p_s is a decreasing function, we have $p_s(t) > p_s(T)$: the first student specializes with a higher level of confidence than the second one. Furthermore, the belief level $p_s(t)$ is also statistically the cross-sectional proportion of types S among agents specializing at time t. Fast learners thus are more likely to specialize *correctly* and be type-matched to their initial job field than slow learners.

Figure 3 illustrates one realized belief path of a single S-type student, overlaid with the optimal boundaries. Near $t = 0$, the agent requires his subjective probability of having comparative advantage S to be above 0.71 or below 0.29 in order to specialize into fields S or A, respectively. If he has not specialized by time $t = 7$, however, much smaller deviations from $1/2$ would be sufficient to trigger specialization: beliefs above 0.55 or below 0.45 would suffice.

3.2. **Estimated N-type discrete time model.** To bring the model to data, we modify it in several ways that provide a better fit with the observations and respond to the computational challenges brought about by the rich informational structure. First, we estimate the college phase of the model in discrete time, which enables the use of standard dynamic programming techniques. Second, we consider a choice between three fields as opposed to two in the continuous time model, to accommodate the three high-ability applied majors that we retain for estimation. The modifications these changes entail are highlighted below.

3.2.1. *Computational challenges.* The curse of dimensionality affects bandit models particularly, as beliefs enter as multi-dimensional state variables in the optimization and the number of decision nodes increases very quickly, both with the number of periods and the with the dimensionality of beliefs (beliefs have $N-1$ degrees of freedom). This difficulty can be alleviated with the use of Gittins indices and recent papers[22] manage to accommodate binary state variables; however, non-stationary problems remain intrinsically challenging. Our formulation of the learning process postulates a sequential sampling of discrete signals (three per year of mixed education), which has the benefit of not necessitating further approximations. The discrete model limits the number of nodes at which the value function is estimated, first because the number of possible beliefs in the period following a given node is equal to the number of fields, but also because the ordering of signals is irrelevant, so several belief paths lead to the same belief. We retain the continuous-time formulation once agents specialize, which enables us to obtain explicit value functions in the specialization phase, and therefore simplifies and speeds up computation appreciably.[23]

3.2.2. *Agents.* There are N fields of work (fields $f \in \{f_1, f_2, \ldots, f_N\}$), with N corresponding subjects taught at school, so human capital is an $N-$ dimensional state variable $e = (e_1, e_2, ..., e_N)$. Agents are born at date $t = 0$ with human capital $(0, 0, \ldots, 0)$, have a finite lifetime T, and discount the future at a constant rate $\delta < 1$. They are endowed with an unknown comparative advantage in one of the N fields (type $\theta \in \{1, 2, \ldots, N\}$). [24]

[22]See Papageorgiou (2014); Eeckhout and Weng (2011).

[23]Discount factors are adjusted to make preferences consistent.

[24]The model does not allow for heterogeneity among students other than their comparative advantage and the beliefs they hold about their comparative advantage. By limiting the number

3.2.3. *Choice set and education tenures.* Agents choose their education path as described in Section 3.1.2. For each period they are enrolled in multi-subject studies, agents now acquire $1/N$ units of education in each field.

3.2.4. *Information acquisition.* Agents begin higher education with an $N-$dimensional, type-dependent prior, $p_{0,\theta} = [P_1, P_2, \ldots, P_N]$. At the end of each period of multi-subject studies, the agent receive a noisy signal about his type.[25] Specifically, he observes a signal $\sigma \in 1, 2, ..., N$. With probability ρ, the signal corresponds to his type ($\sigma = \theta$), while with probability $1 - \rho$ the signal is misleading ($\sigma \neq \theta$) and correspond to any of the $N - 1$ other types. The agent updates his beliefs before choosing his education stream for the subsequent period. For example, if the agent holding belief p observes $\sigma = 1$, his updated belief vector will be given by:

(10)
$$p' = \left[\frac{\rho P_1}{\rho P_1 + \gamma(1 - P_1)}, \frac{\gamma P_2}{\rho P_1 + \gamma(1 - P_1)}, ..., \frac{\gamma P_N}{\rho P_1 + \gamma(1 - P_1)}\right], \text{ with } \gamma = \frac{1 - \rho}{N - 1}.$$

3.2.5. *Payoffs.* Payoffs and occupational choice are as described in Section 3.1.4. In the N-type case, the effective stock of skills is defined as follows:

$$\forall n = 1 \ldots N, \, \epsilon_n = e_n + \beta \sum_{m \neq n} e_m. \tag{11}$$

There is no distinction across fields: all out-of-field education is treated symmetrically, as are all fields outside of an agent's comparative advantage.[26]

---

of majors considered, and by considering only successful college graduates, we narrow the span of ability within our sample; however, we acknowledge that vertical ability differences, not captured in our model, do remain. Table 23 in Appendix A.4 gives the correlation of the timing of specialization with several observable characteristics.

[25]A growing body of research explores how grades effect students' beliefs and course choices (see Arcidiacono (2004), Zafar (2011), Main and Ost (2014) and Stinebrickner and Stinebrickner (2014, forthcoming)). In this paper we remain agnostic about the source of the signals that students receive. While grades no doubt play a role, other unmeasurable factors also influence students' academic paths.

[26]We treat all 'out-of-field' education symmetrically. At the high school level, there is little evidence that any subjects are more universally rewarded than others (see Altonji (1995)). Math is something of an exception, though the evidence is sparse. Joensen and Nielsen (2009) find a strong causal effect of advanced math courses on later earnings in Denmark; elsewhere, the effect is small (Morin (2013), Canadian data), or only present in some groups (Levine and Zimmerman (1995), female college graduates in the US).

Flow returns (theoretical counterparts of log earnings) are given by:

$$y_f(\theta, (e_f, e_{-f})) = R(\epsilon_f) + \mathbb{1}_{\theta=f} P. \tag{12}$$

We briefly review the optimal behavior of agents and introduce the notation necessary to describe moment equations. We start by describing the last decision node, the decision to change fields, then we describe the optimal length of specialization and eventually the optimal experimentation phase.

3.2.6. *The decision to change fields.* We assume that type θ is perfectly revealed upon entry but agents face a cost c of switching fields, a shortcut for the more realistic gradual realization of type mismatch and associated foregone experience and possible retraining.[27] This simplistic assumption also correspond to the limited observations of agents' early careers in the data. Agents who have stayed in mixed education until period t and in specialized education in field 1 for H years have education stock $(t/N + H, t/N, \ldots, t/N)$. Initially type-mismatched agents observe their type and the draw from the cost distribution c and receive the following value if they switch to their field of comparative advantage $\theta \neq 1$:

$$J_{sw}(t+H, e, c) = -c + \int_0^{T-t-H} \exp\{-rt\} \left[R(\epsilon_\theta) + P\right] dt \tag{13}$$

If they remain in their field of specialization (field 1), they obtain value

$$J_{st}(t+H, e, c) = \int_0^{T-t-H} \exp\{-rt\} R(\epsilon_1)\, dt \tag{14}$$

They choose to switch fields provided that $J_{sw}(t+H, e, c) \geq J_{st}(t+H, e, c)$ or equivalently, whenever c is lower than a cutoff value $\hat{c}(H)$.

From an ex ante perspective, upon starting work after training for a length H, initially type-mismatched agents obtain value

$$\bar{J}(t, e, H) = \int_{c_0}^{\hat{c}(H)} f(c)\, J_{sw}(t+H, e, c)\, dc + (1 - F(\hat{c}(H)))\, J_{st}(t+H, e, c), \tag{15}$$

[27]We do not model the labor market explicitly. This raises concerns that field-specific labor market fluctuations could affect our results, either by drawing in large numbers of students during booms or forcing graduates into other fields during crashes (in addition, Altonji et al. (2013) document that the returns to individual majors are affected differently by recessions). While we cannot rule this out, the sensitivity of our estimates should be reduced by the coarse aggregation of majors and occupations: large categories mean that those who cannot find work exactly corresponding to their major are likely to land a job in the broad field it is associated with. Furthermore, existing evidence suggests that the elasticity of major choice to market conditions, while positive, is relatively small: see Blom (2012) and Beffy et al. (2012).

where f and F are the *pdf* and *cdf* of the truncated exponential distribution. Agent who are type-matched to field 1 receive value

$$J_{tm}(t+H,e) = \int_0^{T-t-H} \exp\{-rt\}\left[R(\epsilon_1)+P\right]dt \tag{16}$$

3.2.7. *The optimal length of specialization.* At time 1, upon beginning specialization in field 1, the agent expects to be type-matched to field one with probability p_1 (the first entry in the beliefs vector) and type-unmatched with probability $1-p_1$. Given that the education stock and time are linked by the relation $e = (t/N, t/N, \ldots, t/N)$, we omit the explicit dependance on e. The expected value of specializing in field 1 writes as:

$$\begin{aligned} V_1(t,p) = & \underset{\{H\geq 0, H\leq T\}}{Max} \int_0^H \exp\{-rt\}\, z dt \\ & + \exp\{-rH\}\left[(1-p_1)\bar{J}(t,e(H)) + p_1 J_{tm}(t+H,e(H))\right] \end{aligned} \tag{17}$$

Upon finding the maximizer $H^*(t,p)$, we can define the expected wage conditional on changing fields.

$$ew_{sw,1}(t,p) = R\left(\beta(t+H^*(t,p)) + (1-\beta)t/N\right) + P \tag{18}$$

The expected wage conditional on staying in the given field is the weighted average of the wage of initially-matched and initially-unmatched agents. The total probability A of carrying on in field 1 is the sum of the probability of being properly matched $A_1 = p_1$ and the contribution of initially unmatched agents who remain in the field, $A_2 = (1-p_1)\left(1 - F\left(\hat{c}(H^*(t,p))\right)\right)$. Both groups receive as earnings $x_1 = R\left(\beta(t+H^*(t,p)) + (1-\beta)(t/N + H^*(t,p)\right)$ while type-matched agents receive the premium P.

$$ew_{st,1}(t,p) = (A_1(x_1+P) + A_2 x_1)(A_1+A_2)^{-1} \tag{19}$$

3.2.8. *Optimal experimentation and beliefs histories.* Continuing in mixed-education enables agents to periodically receive signals about their field of comparative advantage, while increasing their stock of skills in all fields. In the discrete-time formulation, call δ the discount rate that can be compounded into the calibrated annual rate .[28] Recalling that a period is a third of a year, the Bellman equation

[28]There are three periods per year, so δ^3 is the annual discount rate.

that defines the value of mixed-education reads:

$$V_0(t, p, e) = 1/3\, z + \delta\, \mathbb{E}\, [V(t + 1/3, p', e')] \tag{20}$$

Where V is the maximum of V_0 and the V_i reflects optimal behavior from $t + 1/3$ and $e' = e + 1/3 \times (1/N, \ldots, 1/N)$. The expectation is taken over future values of p', the bayesian update obtained as in equation (10).

The last object we define is the density of exit times, necessary for the definition of moments. Call $h_t(p)$ the N-dimensional vector such that the i-th entry reflects the mass of agents of type i holding belief p at time t, having never specialized before time t. Starting from a mass 1 of agents at date -1, we distribute them across types according to the vector p_{-1} chosen so as to reflect the empirical distribution of majors, so $h_{-1}(p_{-1}) = p_{-1}$. Next, agents observe one signal which agrees with their type with probability ρ_0 and update their beliefs according to formula (10), replacing ρ with ρ_0.

This procedure generates the time-0 beliefs $p_{0,i}$ that correspond to the Bayesian update of p_{-1} upon observing signal i. To update the distribution, observe that type-1 agents receive signal 1 with probability ρ_0 while type-i, $i \neq 1$ agents receive signal 1 with probability $\gamma_0 = (N-1)^{-1}(1-\rho_0)$. The mass of agents of type 1 holding belief $p_{0,1}$ is therefore $h_0(p_{0,1})_1 = \rho_0 \times (p_{-1})_1$, where $(p_{-1})_1$ is the first entry of vector p_{-1}. The mass of agents of type $i > 1$ holding the same belief is $h_0(p_{0,1})_i = \gamma_0 \times (p_{-1})_i$, while the total mass of agents holding belief $p_{0,1}$ at time 0 is $\sum_{i=1}^{N} h_0(p_{0,1})_i$.

All possible belief points can be split between experimentation and specialization nodes. Say $p \in Ex(t)$ if agents holding belief p at time t choose to carry on in mixed-education and $p \in Sp_i(t)$ if they choose to specialize in field i at time t. While experimentation prevails, we can update vector h iteratively using the same procedure as above, except parameter ρ is used in updates instead of parameter ρ_0.

4. Estimation

4.1. **Methodology.** We simulate the model described in Section 3.2. Parameter values are obtained through a combination of external calibration and simulated method of moments. The model is estimated using a subset of three majors: sciences & math, engineering, and business & economics.[29]

The parameters of the model are described in Table 4. While we have chosen to aggregate courses by year in the data, we estimate the model with shorter periods: each year is represented by three periods, with 12 periods being the maximum duration of non-specialized studies. This allows for a greater heterogeneity of beliefs and smoothness of decision nodes, reducing the granularity introduced through the discretization of the model.

For computational reasons, we impose that the cost of switching fields is drawn from an exponential distribution truncated[30] above a cutoff $c_0 > 0$ and we add the mean of the cost distribution (C_λ) to the list of estimated parameters. Since there is no mass on negative realizations of the cost, type-matched agents never find it advantageous to switch fields. Furthermore, the optimal stopping property and symmetry guarantee that type-unmatched agents do not find it optimal to switch to a field different from their comparative advantage.

4.2. **Moments.** The moments we use to estimate the model relate to four observables: the timing of specialization, the field of specialization, the relation between the field of work and the field of studies, and wages. Empirically, these correspond to four sets of moments. The first is the proportion of students specializing in each major, in each year (shown previously in Figure 1). These proportions fully describe the first two observables: that is, the timing of specialization and the field of specialization. The second is the probability of working in a field related to one's field of studies, conditional on major and timing of specialization. The third and fourth sets of moments are wages. We calculate two average wages for each major–timing of specialization cell: wages for those who are working in the field of their major, and wages of those who have switched to a different field.

[29]The inclusion of detailed major categories remains a challenge throughout the literature. In empirical work, Kinsler and Pavan (forthcoming) retain three majors (science, business and other), Arcidiacono (2004) uses four (natural sciences, business, education and social science/humanities/other), Stange (2013) includes three (business, engineering and nursing), while Altonji (1993)'s conceptual model has only two (math or science, and humanities). Our choice of majors is constrained by both computational power and cell size. While education majors are sufficiently numerous to be included, we restrict our primary sample to the more homogenous set of quantitative majors.

[30]The truncation enables to put probability mass on relatively high cost values without requiring a very low decay rate. The assumption of exponential distribution enables us to compute explicit continuation values without requiring dynamic programming in the specialized phase.

Table 4. Parameters

Parameter	Description
Calibrated	
p_{-1}	Prior belief: set to reflect the empirical distribution of majors*
δ	Annual discount factor: set to 0.96%
T	Working lifetime: set to 20 years
z	Flow cost of education: set to approximately -15% of average wage
Estimated	
ρ	Precision of learning
ρ_0	Precision of the initial signal
β	Transferability
P	Matching premium
$R_f(\epsilon_f)$	Returns to education** with $R_f = R \; \forall \; f$***
C_λ	Expected cost of switching fields

*This assumption is innocuous when the returns across fields are identical and there are no differences in flow utility from studying different subjects.

**Returns to effective education – the sum of in-field education and β times out-of-field education – are assumed concave. We constrain this function to be a cubic polynomial with a value of one and a zero derivative at the maximum effective education level (the level of education corresponding to immediate specialization in any field for 4 years), and to be equal to zero at zero education. We estimate one remaining curvature parameter from the data.

***Returns do vary across fields, both in the cross-section (Carnevale et al. (2012) present recent evidence from the United States, Finnie (2002) from Canada, to name just a few) and in studies controlling for selection (e.g. Chevalier (2012), Kinsler and Pavan (forthcoming)).

The vector of moments therefore has 60 entries: 4 moment types, 5 years[31] and 3 fields. To formalize notation, let MT be the set of theoretical moments, each specific to a year-field cell. The first block of moments concerns timing: entry $y + 3 \times (f - 1)$ is the proportion of agents specializing in field f and year y:

$$MT_{y+3\times(f-1)} = \sum_{t=3(y-1)}^{3y-1} \sum_{p \in Sp_f(t)} \sum_{i=1}^{N} (h_t(p))_i. \tag{21}$$

The innermost summation symbol denotes the fact that we are summing across agent types, the intermediate one corresponds to the pooling of belief nodes leading to specialization in field f in period t, and the outermost summation aggregates 3 periods, to bring the unit of observation from a period up to a year. Since all agents eventually specialize, the sum of the first 15 entries equals the total population.

[31]Recall from Section 2.2 that, although all students in our sample graduated with a major in one of the three fields, some never reach the threshold of specialization. We interpret this as *very late* specialization; that is, specialization just prior to entering the labor market.

The second block of moments measures the number of horizontally matched agents: entry $15 + y + 3 \times (f - 1)$ is the proportion of agents who specialize in field f and year y, eventually carry on working in field f. Each such entry is the product of the corresponding timing entry and the year-field cell's average probability of remaining in the chosen field:

$$MT_{15+y+3\times(f-1)} = \sum_{t=3(y-1)}^{3y-1} \sum_{p\in Sp_f(t)} \sum_{i=1}^{N}$$

$$(h_t(p))_i \times [p_f + (1-p_f)(1 - F(\hat{c}(H^*(t,p))))] . \tag{22}$$

These first two blocks of moments are proportions which are directly comparable to their empirical counterparts. The next two blocks represent earnings: entry $30 + y + 3 \times (f - 1)$ is the average earnings of agents who specialize in field f and year y and eventually carry on working in field f, while entry $45 + y + 3 \times (f - 1)$ is the average earnings of agents who specialize in field f and year y but work in a field different from f. These vectors are computed according to section 3.2.7 and averaged using the relevant proportions.

4.3. **Distance criterion.** In order to relate the earnings moments to their theoretical counterparts, particularly regarding orders of magnitude and dispersion, we first standardize them: standard deviations of income are therefore compared to standard deviations of returns in the model.

Our parameter estimates are the values which minimize the weighted difference between the empirical and theoretical versions of the moments described above. The model is simulated for every element s in the set S of parameters on a defined grid, and the estimates are the parameter values $\hat{\Theta}$ that satisfy:

$$\hat{\Theta} = argmin_{\Theta_s}(M_E - M_T(\Theta_s))'W(M_E - M_T(\Theta_s)), i = s...S; \tag{23}$$

where M_E are the empirical moments, $M_T(\Theta_s)$ their simulated counterparts using parameter values Θ_s, and W is a weighting matrix. There are two issues with the determination of matrix W. The first one has to do with commensurability of measurements. Empirical and theoretical income measurement units must be made comparable, and then adjusted so that they are of similar magnitude to proportion-based moments. This is required so that no one set of moments dwarfs another, or the distance function would emphasize them disproportionately. To address both difficulties, we standardize sets of moments block-wise. That is, each block of 15 entries is adjusted linearly so as to have mean 0 and standard deviation 1.[32] Finally, we constrain our estimates to generate a pattern of majors consistent with the observed distribution – and thereby consistent with the prior. We do this by eliminating parameter values that cause the predicted share in each major to deviate from the observed share by more than 25%.

The weighting matrix used for the primary estimation is a diagonal matrix of the empirical proportions of the sample in each major-timing of specialization cell, that is, the first 15 entries of the M_E vector, repeated 4 times. We also compute distance criteria that put a lower weight on income moments, to reflect the higher uncertainties associated with the modelling of earnings. Distance functions using the Identity matrix for weights, as well as alternative measurements of the timing variable and selected subsamples, are explored in Section 6.1.2.

4.4. **Estimates.** The values of the six estimated parameters are given in Table 5. These results support the existence of learning in general education and also imply that specialized education is imperfectly transferable across fields. Specifically, we find that each year in multi-subject information gives a signal with precision-level 0.39. This value can be compared to a pure noise 'signal', which in our 3-type case would have a precision of 0.33. Conversely, if a single period of broad studies revealed type with certainty, the precision of the signal would be 1. Our estimated signal is therefore informative, but still noisy.

Another way to understand the estimated precision level is by considering the *expected entropy reduction*, a pure measure of informativeness.[33] We find that the initial distribution of majors is associated with an entropy of 1.58, which the

[32] Populations (cell sizes) are used as weights in the standardization.

[33] If variable θ can take any of N values, write $q_k = \mathbb{P}[\theta = k]$. Shannon entropy measures the uncertainty associated with belief vector q and is defined as

$$E(q) = -\sum_{j=1}^{N} q_j log_2(q_j), \text{ with } 0 \times log_2(0) = 0 \text{ by convention} \tag{24}$$

To define expected entropy reduction, suppose the agent starts a given period with belief $p \in \Delta_N$. The signal σ has n possible realizations and with probability ω_i, the bayesian update of p following observation σ_i takes value g_i. We can thus define the expected entropy of the posterior, leading to the following definition

$$I(\sigma, p) = E(q) - \sum_{i=1}^{n} \omega_i E(g_i) \geq 0 \tag{25}$$

pre-college signal reduces by 0.075. Each period of multi-disciplinary college (a third of a year) reduces entropy by a further 0.027. Students therefore acquire as much information in one year of broad college courses as they did in the entire pre-college period ($3 * 0.027$).[34]

The parameter β is estimated at 0.90; this implies that out-of-field education is remunerated at 90% of the level of education related to one's field of work. An individual suffers a modest loss of human capital when choosing to work in a field different from his major – more substantial if he specialized early. This loss of human capital is compensated for by a large premium to working in the field of one's comparative advantage. Our estimated matching premium is 0.20: those who are type-matched to their field of work earn 20% higher wages than similarly-skilled individuals who do not. Finally, students incur large one-time costs when they switch fields, equivalent to 1.59 years of income.

Table 5. Estimated parameter values

Parameter	Definition	Estimate	Discussion
ρ	Precision of learning	0.39	Compare to an uninformative signal: $\rho = 0.33$
ρ_0	High school signal	0.49	The precision of beliefs at college entry
β	Transferability	0.90	Out-of-field education is remunerated at 90% of in-field education
P	Matching premium	20	20% percent of earnings are due to type-match with occupation
$R_f(\epsilon_f)$	Return function	23.5	Curvature parameter for the returns function (no intuitive interpretation)
C_λ	Expected switching penalty	1.59	Corresponds to $\sim$1.5 years of income

4.5. **Model fit.** We explore model fit in two ways. First, we present graphically the relative and absolute deviations of the 60 theoretical moments from their empirical counterparts. The left panel of Figure 4 shows the contribution each

In investment problems, expected entropy reduction is shown by Cabrales et al. (2013) to be the unique parameter-independent complete ordering of information structures that agrees with investors' willingness to pay. It is therefore a valid measure of the informativeness of a signal, particularly when it comes to comparisons.

[34]Neither signal reduces the absolute value of entropy by a large amount, but entropy is a concave function of beliefs and decreases fast near the edges of the simplex, so absolute variations near the middle of the simplex are small.

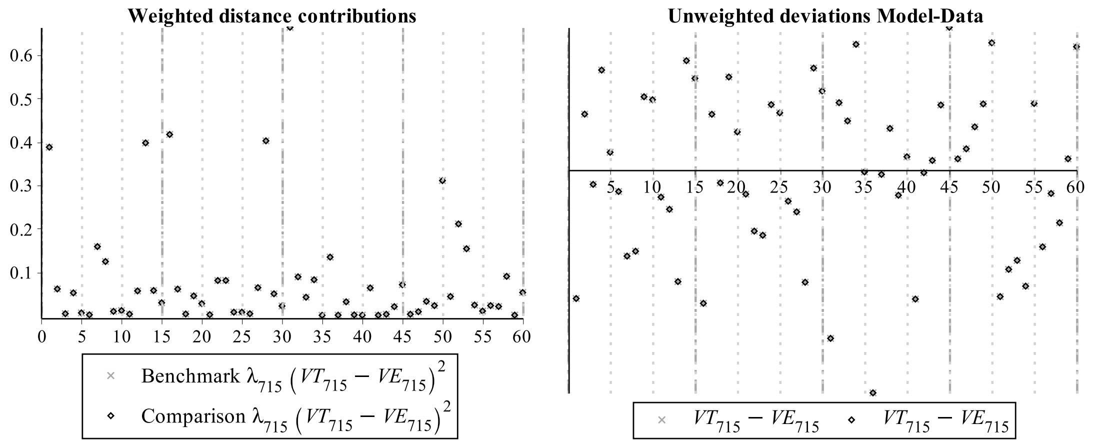
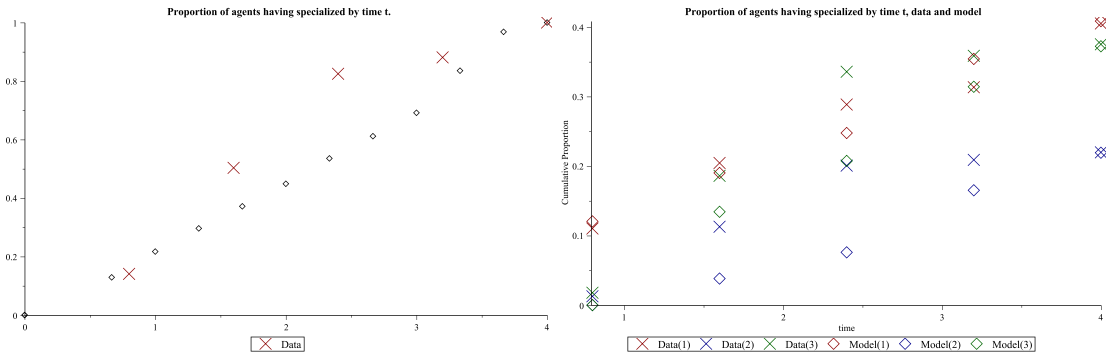


Figure 4: Relative (L) and absolute (R) deviations: model vs data

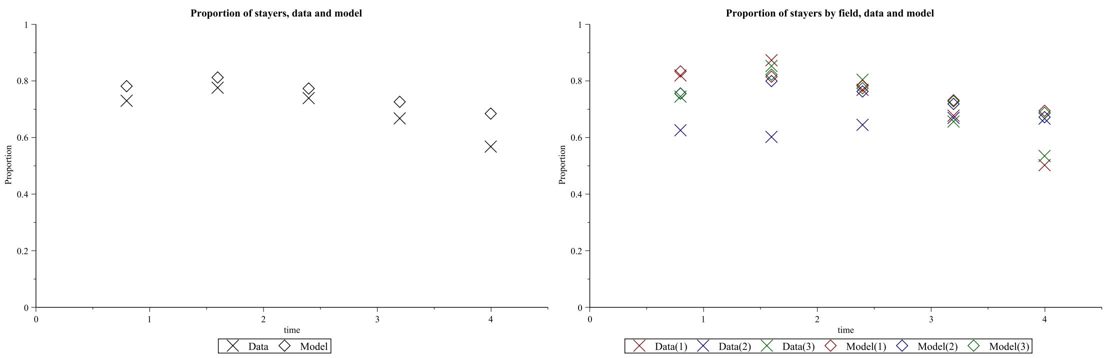


Aggregated By major

Figure 5: Model vs. data: specialization times

moment makes to the distance function (moments are numbered from 1-60, as described in Section 4.2). The absolute deviations behind these contributions are shown in the right panel of Figure 4.

To understand the implications of these differences, we next examine each block of moments individually. Figure 5 shows predicted and observed patterns of specialization. In the left panel, all majors are aggregated: black diamonds represent predicted specialization at different belief nodes, while red crosses plot the empirical counterpart. The right panel displays the same data, broken down by major (majors are (1) Science & Math, (2) Engineering, (3) Business & Economics). The model predicts that students will specialize slightly later, on average, that they do in the data. When broken down by major, we can see this arises primarily from a failure to match the mass of specialization by engineering students in periods 2 and 3, although the model under-predicts mid-term specialization overall as well.



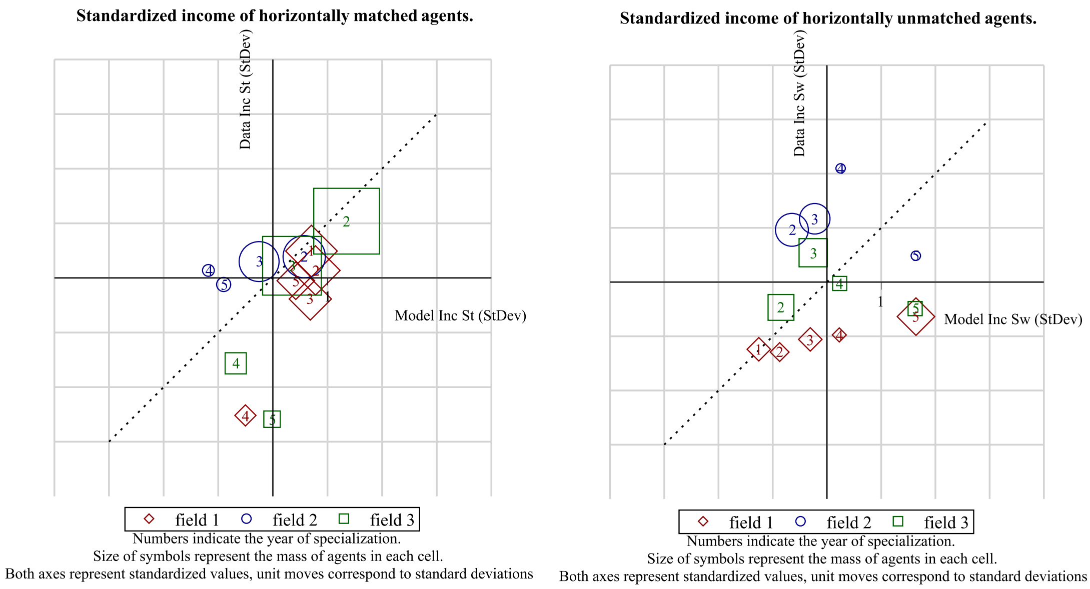
Aggregated — By major

Figure 6: Model vs. data: probability of switching

The probability of working in the field of studies, given the timing of specialization, is shown in Figure 6. On aggregate (left panel), staying is over-predicted by the model. The disaggregated comparison (right panel) suggests that the low propensity of engineers to remain in engineering is driving the divergence between the model and the data. This is compounded for late specializers by a low probability of staying for the other majors as well.

The two sets of wage moments are presented in Figure 7, with stayers in the left panel, and switchers on the right. The graphs present standard deviations from mean income for both the model, on the horizontal axis, and the data, plotted on the vertical axis. If the model perfectly predicted income differences, the plotted observations should be arranged along the 45-degree line. Figure 7 shows that the model predicts income quite well for workers who remain in the field of their major. While there are some off-diagonal observations, these are generally small masses of individuals.

This is not so much the case for the income of switchers: in the data, engineers who switch fields earn systematically higher incomes, while science and math majors who switch fields earn low incomes. The model, which predicts lower incomes for early specializers who switch, matches empirical wages for early-specializing scientists and business & economics majors, but performs poorly elsewhere. This may be partly due to our coarse treatment of occupations outside of the field of study: in the model, all occupations unrelated to the field of study are treated symmetrically. Table 18 in Appendix A.2 lists the occupations of graduates from each major. A full 43% of all engineering graduates who switch fields are employed in business and management (a category which includes high-paying occupations



Stayers Switchers

Figure 7: Model vs. data: wages of stayers (L) and switchers (R)

in finance), whereas science and math graduates who are working in other fields are spread out across a greater diversity of occupations, including 29% employed in the low-paying field of education.

4.6. **Identification.** Although all parameters are estimated simultaneously, we can give some intuition about the identification process. We do so using two approaches. First, we compute illustrative simulated comparative statics for a linear-symmetric case of the model. These results are reported and discussed in Appendix C. Second, we constrain the parameters of our simulation one at a time, and re-estimate the remaining parameters. To accomplish this, we take the grid of parameter values that we used to estimate the model and impose the value of one parameter at a time. We then select the set of parameter values which, while respecting our imposed constraint, minimizes the primary distance criterion. In order to compare high and low values of the parameter in question, we do this for the highest and lowest value that parameter takes on our grid.These experiments suggest that the timing of specialization and switching behavior are primarily responsible for identifying the precision of learning, both prior to and during college. These parameters, along with the earnings moments, in turn pin down the parameters governing returns and switching costs.

The resulting parameter estimates are given in Table 6, and will be discussed in more detail in Section 6.1.1. We focus here on how these experimental variations affect the distance criterion, and the implications this has for identification. Table 7 gives the distance criterion for each experiment, broken down into the contribution of each block of moments. These block fall naturally into two groups: the first two capture the behavior of agents (major & timing of specialization, and the share of each major-timing cell who stayed in their field of study on the labor market), while the second two are wage moments (for stayers, and for switchers). The contribution of each of the 60 individual moments to the distance criterion is presented graphically in Appendix D.

4.6.1. *Learning parameters.* The first two rows of Table 7 impose the precision of signals received during multi-disciplinary studies. Imposing an imprecise signal actually improves the match with the empirical wage moments; however, it does so by worsening the match with the behavioral moments considerably. Imposing a high precision of learning has a smaller effect on the distance criterion, with the largest deviation from the benchmark coming from the wages of stayers. Overall, it appears that variations in the precision of the college signals during affects the distance criterion primarily through the behavior moments.

Imposing a highly informative pre-college signal impacts the distance criterion in a similar way as did the imposition of an *un-informative* college signal, and vice-versa for a highly informative pre-college signal: notice the symmetry between the first and second pair of rows in Table 7. This suggests that the precision of the pre-college signal is also being pinned down by the behavioral moments, although in this case the variation in the wages of stayers is also quite important.

4.6.2. *Return function parameters.* The next four rows of Table 7 present variations in the transferability of specialized education and the matching premium. Based on the distance criteria alone, it appears that these two parameters are relatively unimportant. In the case of the comparative advantage premium, P, the distance criteria for the low and high values are almost identical, and the contribution of each block of moments barely changes. In both cases, the distance criterion is relatively evenly contributed to from each block of moments.

A look ahead to the distance-minimize parameter values estimated under each constraint, listed in Table 6, suggests that these parameters matter a great deal. This is particularly the case for the estimation of the switching cost and the transferability of specialized education, which appear to move together: low switching costs coexisting with high transferability, for an 'easy mobility' alternative, and vice-versa for a 'tough mobility' alternative. The fact that these two very different alternatives have such similar distance criteria suggests that the identification of these parameters is not as strong as the others.[35]

4.6.3. *Switching costs.* The switching cost parameter appears to be driven by both the behavioral moments and the wage moments (see the final rows of Table 7). While the differences in each case are modest, with the experiment imposing high switching costs matching three of the four moments better than that with low switching costs, variation in this parameter appears to affect all four sets of moments to a similar degree.

TABLE 6. Estimated parameters when constraining one parameter at a time (actual estimates obtain by dividing by 100)

Contraint	ρ	ρ_0	P	$R_f(\epsilon_f)$	β	C_λ
Benchmark	**39**	**49**	**20**	**23.5**	**90**	**1.59**
Low ρ	**37**	45	24.5	23.5	90	2.97
High ρ	**43**	47	17	23.5	66	5.99
Low ρ_0	40	**39**	15.5	23	70	5.68
High ρ_0	39	**51**	24.5	23	90	1.2
Low P	40	45	**15.5**	23.5	82	4.72
High P	39	49	**24.5**	23	90	2.94
Low β	41	49	18.5	23.5	**66**	6.82
High β	39	49	20	23.5	**90**	1.59
Low C_λ	39	47	15.5	22.5	90	**1.11**
High C_λ	40	43	24.5	23.5	66	**8.75**

5. Policy simulation: the costs of imposing early specialization

Calls to reform college education in the US regularly accuse bachelor degrees of being too broad and weakly linked to the labor market. What would happen if students were forced to specialize at college entry? Using the parameter values estimated above, we can predict the impact of such a policy. Specifically, we

[35]We do not present experimental variation in the curvature of the returns function. While this parameter is important, it is not a primary focus of our study: we estimate it because we lack any reasonable outside calibration.

TABLE 7. Contribution of each block of moments to the distance criterion, under different constraints

Constraint	Block 1	Block 2	Block 3	Block 4	TOTAL
Benchmark	**0.67**	**0.53**	**0.54**	**0.97**	**2.72**
Low ρ	1.43	1.41	0.34	0.96	4.15
High ρ	0.70	0.62	0.84	0.86	3.01
Low ρ_0	0.75	0.60	1.09	0.88	3.32
High ρ_0	1.49	1.38	0.50	0.97	4.34
Low P	0.75	0.61	0.49	0.90	2.76
High P	0.72	0.61	0.46	0.97	2.77
Low β	0.71	0.71	0.63	0.86	2.91
High β	0.67	0.53	0.54	0.97	2.72
Low C_λ	0.97	0.82	0.64	0.96	3.39
High C_λ	0.86	0.75	0.72	0.83	3.15

Moment blocks: (1) timing of specialization and major, (2) stayers, by timing and major, (3) wages of stayers, (4) wages of switchers.

consider a policy where students must specialize after receiving a single college signal: this corresponds to spending 1/3 of a year in mixed-discipline studies prior to specializing.

We focus on two outcomes, summarized in Table 8. The first, which is observable in the data, is the fraction of students who choose an occupation in a field different from their major.[36] In our baseline simulation we find that 47% of workers are type-mis-matched to their field when they reach the labor market; nearly half of these workers, 24% overall, switch occupations and thus end up working in a field unrelated to their field of studies.[37] In the counterfactual, students do not learn about their type in college, and 50% graduate in a field different than their comparative advantage. These students specialized early, acquiring a large stock of human capital in their major field. The high transferability and large comparative advantage premium nevertheless induce many students to seek out their preferred field: 20% change fields when their type is revealed.

The second outcome is the number of agents who are working in a field different from that of their comparative advantage. In the baseline simulation, 53% of

[36] The large number of young people working in fields unrelated to their field of study has been studied in a number of countries. See, for instance, Finnie (2001) in Canada, McGuinness and Sloane (2011) in the UK, Badillo-Amador et al. (2005) in Spain, Bender and Heywood (2011) for scientists in the US.

[37] Although our criteria for matching occupations and majors is based on coarse categories, the level of horizontal mis-match in our data is similar to that found through other methods. Using the 1993 Survey of College Graduates, Robst (2007) finds that 20% of respondents – across all ages and majors – report that their work is 'not related' to their degree field.

students correctly discover their type during college: all of these are correctly type-matched on the labor market. In addition, 24% switch fields, leaving 23% of the population mis-matched to their field of work. Under the counterfactual policy, only 50% of students graduate in the field best suited to them; 30% overall remain type-mismatched to their occupation.

TABLE 8. Policy Experiment

	Initially matched	Change fields	Remain type-mismatched
Baseline	53%	24%	23%
Counterfactual	50%	20%	30%

This counterfactual experiment highlights the deep implications of our results for education policy. Imposing early field choice actually improves the correspondence between field of study and field of work: 17% fewer students choose an occupation outside their field of study. This apparent improvement masks a significant worsening of the allocation of individuals to occupations that suit them best: early specialization increases type-occupation mis-match by 30%. Our estimates suggests that the average individual cost of this policy is equivalent to the return earned on 0.45 of a year of occupation-related specialized studies, or approximate 3% of wages.

How does a student's expected value evolve as function of imposed timing of specialization? Figure 8 shows the ex-ante expected value for a range of mandated specialization policies. Note that the policy under consideration is specialization imposed *at or before* the date on the x-axis; prior to mandated specialization, students may opt in to specialization at any time. The left-most observation corresponds to the policy described above. The relationship between the mandated specialization time and the expected value is almost linear: while any constraint makes students on average worse off, the time at which specialization is imposed has a large impact.

We do not draw conclusions on whether or not early specialization is an efficient policy choice. There are two reasons for this. First, we do not have data on the relative costs of broad and specialized education. Anecdotal evidence suggests that the breadth of courses and flexibility of course choices at American universities presents non-trivial administrative challenges: early-specialization is often associated with simpler, homogenous course schedules. Second, we make the

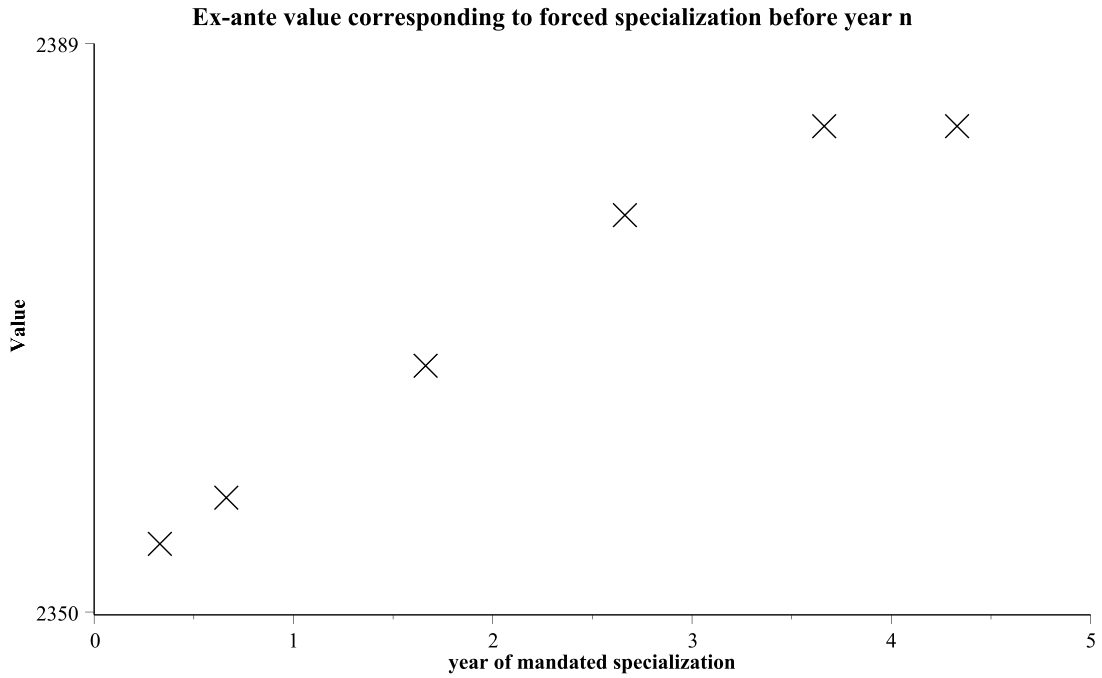


Figure 8: Ex-ante expected value under different specialization regimes

standard assumption that students know more about themselves than the social planner, and that they both process information optimally and make optimal experimentation decisions. It follows from this assumption that *any* constraint on course choices is at least weakly welfare decreasing. While our counterfactual will therefore make students worse off by construction, the magnitude of the effects we find can inform policies which take both the costs and benefits of allowing flexible course choices into account.

6. Robustness Checks and Extensions

6.1. **Robustness checks.** To explore the sensitivity of our estimates to individual characteristic which are outside our model, we perform three types of experiments. First, we constrain our parameter values one at a time, and estimate the remaining parameters. Second, we estimate the model parameters using different distance functions. Third, we estimate the model using different subsamples of the data.[38]

[38]For each subsample, the timing of specialization variable is attributed as it is in the primary estimation; however, the moments are adjusted to reflect the different population under consideration. The calibrated parameters are held constant across the subsample estimations, while the estimated parameters are allowed to vary. This means that the prior belief, which is calibrated to the empirical distribution of graduates in the full sample, is maintained for each subsample estimation. For this reason, we do not report estimates for subsamples with very different patterns of specialization from the full sample: doing so would violate the fixed-point assumption behind our prior beliefs.

While the interpretation of these parameter estimates is quite limited, they shed some light on the sensitivity of our results to different specifications. Overall, the picture is encouraging. The informativeness of mixed-discipline education varies little across the estimations; however, transferability of education and the matching premium are more volatile.

6.1.1. *Constraining parameters.* In Section 4.6 we introduced a series of experiments where we constrain the value of one parameter and estimate the remaining five. In addition to shedding light on identification, these experiments allow us to explore the robustness of our parameter estimates.

The first four experiments, reported in Table 6, concern the precision of learning before and during college. We first constrain the precision of learning during multidisciplinary studies to be low. The resulting parameter estimates, with respect to our baseline specification, have a lower level of pre-college information, but higher switching costs and a higher comparative advantage premium. If learning is imprecise, there must be high returns to making a correct match in order to justify observed behavior - and even then, this set of parameters matching the behavioral moments quite poorly (see Table 7. When imposing a high precision of learning, on the other hand, we find low transferability, high switching costs and a low premium. If learning happens quickly, students must expect that readjustment on the labor market is very difficult, otherwise they would not spend so much time acquiring information.

Constraining the precision of pre-college learning has the reverse effect: imposing a precise pre-college signal leads to estimates with high transferability and low switching costs, while imposing a highly noisy college signal leads to estimates with a high switching costs and low transferability. That the precision of college signals is also different in the two sets of parameters, with a precise pre-college signal associated with a less-precise college signal and vice versa, could partially explain this result. Notice, however, that precise pre-college signal condition – with a loose labor market – misses the behavioral moments quite badly.

Comparing the low and high transferability experiments[39] echoes the comparisons above by suggesting the existence of easy-mobility and tough-mobility alternatives. Imposing low transferability leads to a set of parameters with a *high*

[39]High transferability also corresponds to the benchmark.

switching cost, while high transferability co-exists with relatively low switching costs. Note also that the college signal is quite precise under the low transferability constraint. Variations in the comparative advantage premium perform similarly to variations in transferability with, as discussed in Section 4.6, a negligible difference in the distance function between the high premium and low premium conditions.

The final two rows of Table 6 compare parameter values when we impose a high cost of switching, and when we impose a low cost. As we have seen previously, to justify a low switching cost the transferability needs to be high, while the opposite is true for a high switching cost. In keeping with previous findings, the easy-mobility alternative is associated with less precise college signals and a stronger pre-college signal, although modestly in both cases.

These experiments suggest the existence of an alternate set of estimates, which may not be too distant from our best-fit parameters, with lower transferability, high average switching costs, and more precise college signals than our current benchmark.

6.1.2. *Alternate distance criteria.* In addition to our primary distance criterion, we consider several other distance functions. We consider six variants in two families of distance functions: the first, in keeping with our primary specification, weights each moment by the fraction of the sample found in the corresponding major-timing of specialization cell. This approach puts more weight on cells that are heavily populated. The second family of distance functions is not weighted,[40] but is otherwise identical to the first.

Table 9 presents distance-minimizing parameter estimates for the weighed distance functions, while Table 10 displays the unweighted equivalents. The primary estimates are listed in the first column of Table 9, for comparison. The distance functions differ in the moments which are targeted, and whether or not the moments have been standardized. These moments are, by column: (I) major & timing of specialization, probability of switching and wages, all standardized; (II) major & timing of specialization, probability of staying, standardized; (III) major & timing of specialization, probability of staying, not standardized; (IV) cumulative density of major & timing of specialization (as opposed to cell shares),

[40]The unweighted distance function uses the Identity matrix as the weighting matrix.

probability of match as proportion of cell (as opposed to share of population); (V) cumulative density of major & timing of specialization; (VI) probability of match as proportion of cell.

TABLE 9. Distance-minimizing parameters using alternate criteria - weighted

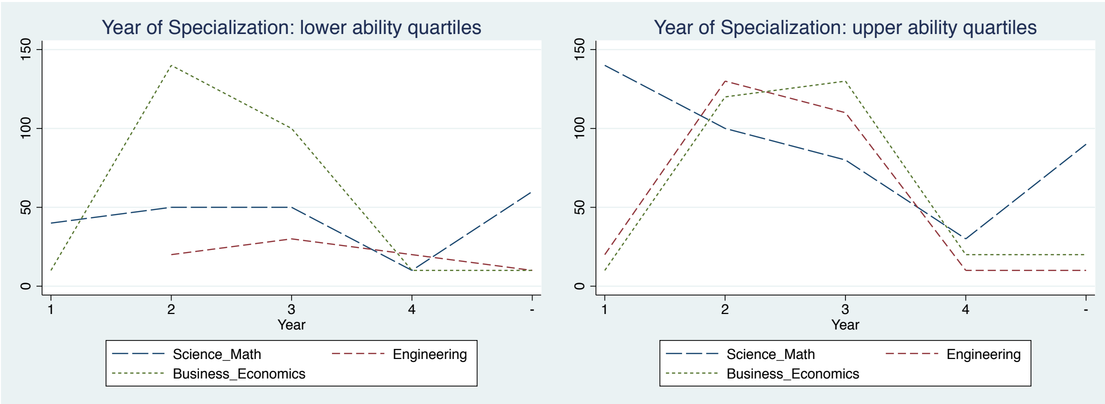
		Distance criteria					
Param.	**Definition**	**I**	**II**	**III**	**IV**	**V**	**VI**
ρ	Precision of learning	0.39	0.42	0.42	0.40	0.42	0.41
ρ_0	High school signal	0.49	0.43	0.43	0.51	0.51	0.47
β	Transferability	0.90	0.66	0.66	0.74	0.66	0.66
P	Matching premium	0.20	0.155	0.155	0.20	0.155	0.155
$R_f(\epsilon_f)$	Return function	0.235	0.225	0.225	23	0.235	0.225
C_λ	Expected switching penalty	1.59	1.65	1.65	1.16	1.72	0.67

TABLE 10. Distance-minimizing parameters using alternate criteria - not weighted

		Distance criteria					
Param.	**Definition**	**I**	**II**	**III**	**IV**	**V**	**VI**
ρ	Precision of learning	0.38	0.41	0.41	0.39	0.42	0.40
ρ_0	High school signal	0.45	0.41	0.41	0.49	0.51	0.45
β	Transferability	0.82	0.66	0.66	0.66	0.66	0.66
P	Matching premium	0.18.5	0.155	0.155	0.20	0.155	0.155
$R_f(\epsilon_f)$	Return function	0.235	0.235	0.23	0.23	0.235	0.225
C_λ	Expected switching penalty	2.10	1.72	1.43	1.03	1.72	0.67

6.1.3. *Ability.* Figure 9 shows the distribution of timing of specialization for graduates who scored in the upper and lower ability quartiles on their college entrance ACT or SAT exam. The quantitative majors we have retained attract relatively high-ability students; approximately 1/3 of the sample falls into the lower two quartiles, and this sparsity makes the specialization patterns between the two groups difficult to compare. Note that this is particularly the case for Science & Math and Engineering, while Business & Economics students are more evenly distributed across ability groups. For the latter, while lower ability students tend to specialize slightly earlier than their higher ability peers, the two curves are overall quite similar.

Table 11 presents two sets of parameter estimates: our primary estimates for comparison alongside values that best fit the subsample of individuals in the higher



Lower ability Upper ability

Figure 9: Timing of specialization: lower and upper SAT/ACT quartiles

ability bracket. Note that we do not report parameter estimates for the lower ability sample: while our entire sample contains roughly equal shares of students in each major (slanted towards science, the largest category), the subsample of lower-ability graduates is heavily dominated by Business & Economics majors. The high-ability subsample estimates are very similar to the full sample with respect to learning; transferability of education and the matching premium are both a little lower.

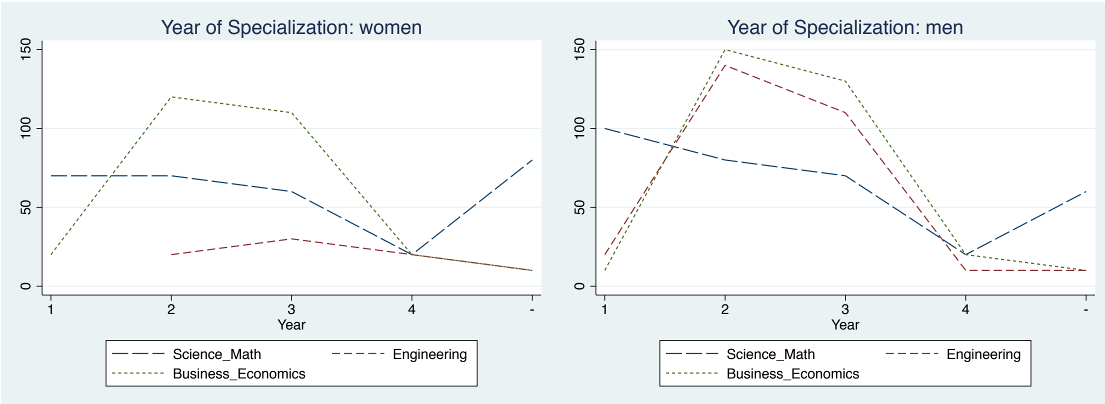
Table 11. Restricted sample: students from upper SAT/ACT quartiles

Parameter	**Definition**	**Baseline**	**Upper Qts**
ρ	Precision of learning	0.39	0.41
ρ_0	High school signal	0.49	0.49
β	Transferability	0.90	0.70
P	Matching premium	0.20	0.17
$R_f(\epsilon_f)$	Return function	0.235	0.235
C_λ	Expected switching penalty	1.59	3.3

6.1.4. *Gender.* Does the learning value or transferability of education depend on gender? There is considerable evidence that major choice itself varies across genders.[41] Furthermore, gender-correlated differences in expected labor market attachment could influence the importance of specialized skills vs. information about ones comparative advantage. Bronson (2014) highlights differential penalties in labor supply reductions as one reason women avoid high-paying majors.[42]

[41] See Montmarquette et al. (2002), Kirkeboen (2012), Holzer and Dunlop (2013), Turner and Bowen (1999), Dickson (2010)

[42] Walker and Zhu (2011) also find that returns to majors vary across genders.



Women Men

Figure 10: Timing of specialization: women and men

Gender-correlated differences in risk aversion (De Paola and Gioia (2011)) and competitiveness and overconfidence (Reuben et al. (2013)) can also play a role.

Figure 10 shows the timing of specialization and major choice for men and women. As above, we do not report parameter estimates for women due to the very small share of women majoring in Engineering. A comparison of parameter estimates for men vs. the full sample (see Table 12) shows slightly greater divergence than the ability subsample: learning is faster – both in high school and in mixed-discipline studies – while transferability of education and the matching premium are lower. Importantly, however, the tradeoff of interest remains pertinent: with low transferability of education, field-related education earns a large premium; however, multi-disciplinary studies are informative.

Table 12. Restricted sample: men only

Parameter	**Definition**	**Baseline**	**Men**
ρ	Precision of learning	0.39	0.40
ρ_0	High school signal	0.49	0.52
β	Transferability	0.90	0.86
P	Matching premium	0.20	0.238
$R_f(\epsilon_f)$	Return function	0.235	0.235
C_λ	Expected switching penalty	1.59	1.63

6.2. **Extensions.**

6.2.1. *Relation to Mincer equation specifications.* Our specification, along with the assumption of logarithmic utility, implies that log earnings are a concave function of years of schooling. Assume that the mapping from earnings to flow payoffs is logarithmic and write $w_f = \exp y_f$ for earnings:

$$w_f = \exp\{(R(\epsilon_f)) + \mathbb{1}_{\theta=f}P\} \tag{26}$$

$$\approx \exp\{(R(\epsilon_f))\}(1 + \mathbb{1}_{\theta=f}P) \tag{27}$$

This justifies our interpretation of P as a proportional earnings premium for type-matched agents. When agents apply a logarithmic utility mapping to earnings, we recover the specification in (12).

Equation (12) relates to the *schooling* component of a Mincer equation.[43] Two important features of our specification are at odds with Mincer equations:[44] while Mincer equations use years of schooling as a covariate, we use the effective stock of skills ϵ. Second, a standard Mincer equation has the logarithm of income depend linearly on years of schooling. For comparability, we can use the best linear approximation (in the sense of minimizing quadratic distance) to our estimated returns function, using ϵ as the covariate, which leads us to retain the value:

$$R(\epsilon_f) = 0.01 + 0.051\epsilon_f. \tag{28}$$

Ignoring informational benefits and field switches, an additional year of schooling increases log earnings by 0.051, corresponding to a 5.1% increase in earnings.

This estimate can be refined in light of our results: taking into account imperfect transferability, $\beta \times 5.1\%$ is a lower bound on the return to schooling. Since about one quarter of agents end up switching fields, $(1/4\beta + 3/4) \times 5.1\%$ gives us an estimate of the average return to specialized schooling. The informational benefits (which increase the probability that a premium will be earned) imply that these are underestimates of the total return to education.

6.2.2. *Overeducation.* While not the focus of this study, our model of higher education has implications for over-education.[45] According to the model, those who choose to work in a field unrelated to their studies will have a smaller stock of specialized education than their fellow graduates who did not change fields. While the

[43]Since our sample contains students of the same age, all of whom attain a bachelor's degree, there is little observable variability in experience, and no observable employment record that would enable tenure observations. Accordingly, our theoretical specification omits experience and tenure effects, leaving only the years of schooling component.

[44]See, for example, Heckman et al. (2006).

[45]See McGuinness (2006) for a review.

allocation of tasks within an occupation grouping is outside our model, it is natural to suppose these individuals will be hired into less-advanced posts than their peers who majored in the occupation-related field. In keeping with Kim et al. (2012)'s study of Korean college graduates, we therefore anticipate a positive correlation between horizontal and vertical occupation-education mis-match.

To investigate this, we take advantage of an additional variable in the data: the (self-reported) education level required by the respondent's most recent occupation. We recode these responses into a binary over-education variable, equal to 1 if the occupation requires less than a bachelor's degree, or if the occupation requires a bachelor's degree and the respondent has earned a master's degree or more.

Table 13 presents results of a regression of occupation-education match on over-education. As expected, we find a positive relationship between over-education and horizontal education-occupation mismatch. Controlling for field of study and occupation, we find that mis-matched workers are approximately 11% more likely to be overeducated. The effect is stronger when restricting to non-quantitative majors: students graduating in these fields are 18% more likely to be overeducated if they have switched to an occupation unrelated to their field of study.

Table 13. Probability of overeducation

	Probability of overeducation		
	All majors	Quantitative	Non-quantitative
Match	-0.108***	-0.0967***	-0.183**
	(0.000)	(0.004)	(0.050)
Controls	X	X	X
R^2	0.097	0.100	0.107
adj. R^2	0.089	0.092	0.083
Sample size	2110	1560	550

Source: B&B93:03, sample restrictions described in section A.1. P-values in parentheses; ** $p < 0.05$, *** $p < 0.01$. Sample sizes rounded to the nearest 10. Match is a dummy variable equal to 1 if the field of studies is the same as the field of work. Controls are major and occupation dummies.

6.2.3. *Early career labor market rigidities.* While our estimated parameter values are specific to the context under investigation – American college students majoring in one of three quantitative, applied fields – the question we address is not. How would an education system characterized by flexible specialization times perform in different countries? One important mediating factor is the flexibility of the domestic labor market, particularly with respect to early career occupation changes. While we have not modeled the labor market explicitly, the one-time cost incurred by workers who change fields reflects the stickiness of occupation categories, above and beyond the transferability of skills.

Suppose the education system we have modeled was adopted by a country with different labor market conditions. How would the probability of changing fields, and of comparative advantage-occupation mismatch, adjust? To explore these questions, we carry out two experiments. Starting from our baseline parameter estimates, we vary the expected cost of switching fields (C_λ). Maintaining all other parameters at their estimated levels, we then compute counter-factual labor market outcomes.

Table 14 presents the result of this experiment, with the expected cost of switching set at the high end to 2.02 years of income, and at the low end to 1.17 years of income.[46] As expected, a higher cost of switching fields reduces the probability of changing fields, and decreases the probability of working in the field of comparative advantage. Reducing the expected switching cost produces a symmetric effect. Interestingly, the proportion of students who specialize in the field of comparative advantage is not affected:[47] this suggests that students do not significantly adapt their timing of specialization in light of a change in expected switching costs, but they do adjust their occupation choices.

Table 14. Policy Experiment

	E(cost)	Initially matched	Change fields	Remain type-mismatched
Baseline	**1.52**	**53%**	**24%**	**23%**
High cost	2.02	53%	20%	27%
Low cost	1.17	53%	29%	18%

7. Conclusion

Does a broad education help people orient themselves towards occupations which are well-suited to them? In the case of post-secondary education, we find evidence that it does. The parameter values we estimate are consistent with a genuine exploration-exploitation tradeoff: broad studies provide information, but

[46]These values are chosen as they represent one step up and one step down on our parameter grid, and are roughly symmetrical increments around our benchmark value.
[47]There is in fact a very small effect, which is not robust to rounding.

specialized studies are more valuable on-the-job. Furthermore, when given the freedom of choice students choose to their timing of specialization in a way consistent with optimal stopping behavior.

While the parameter values we estimate lend support to our model, they also highlight features of the economic environment which are often overlooked. First of all, having explicitly modelled individual heterogeneity as a comparative advantage, we estimate the importance this has on the labor market. The return to working in a field related to one's comparative advantage is large: our estimates put it at 20% of total wages. Secondly, we unpack the college premium along a new dimension: controlling for degree and major, does the timing of specialization matter? The fact that it does suggests that the degree of specialization in college, along with the individual heterogeneity, should be accounted for more carefully when calculating returns to higher education.

Our results point to the importance that education policy has in shaping the labor market returns to education. Since we have not modelled education provision, it is beyond the scope of this paper to assess whether the resulting welfare loss is efficient. However, as the policy experiment in Section 5 illustrates, imposing early specialization is costly to students – and that cost is largely hidden from view.

References

Altonji, J. G. (1993): "The Demand for and Return to Education When Education Outcomes are Uncertain," *Journal of Labor Economics*, 11, 1–37.

——— (1995): "The effects of high school curriculum on education and labor market outcomes," *The Journal of Human Resources*, 30, 409–438.

Altonji, J. G., E. Blom, and C. Meghir (2012): "Heterogeneity in human capital investments: High school curriculum, college major, and careers," *Annual Review of Economics*, 4, 185–223.

Altonji, J. G., L. B. Kahn, and J. D. Speer (2013): "Cashier or consultant? Entry labor market conditions, field of study, and career success," working paper.

Arcidiacono, P. (2004): "Ability sorting and the returns to college major," *Journal of Econometrics*, 121, 343–375.

Arcidiacono, P., E. Aucejo, A. Maurel, and T. Ransom (2013a): "College attrition and the dynamics of information revelation," working paper.

Arcidiacono, P., E. M. Aucejo, and V. J. Hotz (2013b): "University differences in the graduation of minorities in STEM fields: Evidence from California," working paper.

Arcidiacono, P., E. M. Aucejo, and K. Spenner (2012): "What Happens After Enrollment? An Analysis of the Time Path of Racial Differences in GPA and Major Choice," *IZA Journal of Labor Economics*, 1(1), 1–24.

Badillo-Amador, L., A. Garca-Snchez, and L. Vila (2005): "Mismatches in the Spanish Labor Market: Education vs. Competence Match," *International Advances in Economic Research*, 11, 93–109.

Beffy, M., D. Fougère, and A. Maurel (2012): "Choosing the field of study in postsecondary education: Do expected earnings matter?" *Review of Economics and Statistics*, 94, 334–347.

Bender, K. A. and J. S. Heywood (2011): "Educational Mismatch and the Careers of Scientists," *Education Economics*, 19, 253–274.

Blom, E. (2012): "Labor market determinants of college major," working paper.

Bordon, P. and C. Fu (2013): "College-Major Choice to College-Then-Major Choice," working paper.

Borghans, L. and B. H. H. Golsteyn (2007): "Skill transferability, regret and mobility," *Applied Economics*, 39, 1663–1677.

Bradley, E. S. (2012): "The Effect of the Business Cycle on Freshman Major Choice," working paper.

Bronson, M. A. (2014): "Degrees Are Forever: Marriage, Educational Investment, and Lifecycle Labor Decisions of Men and Women," working paper.

Cabrales, A., O. Gossner, and R. Serrano (2013): "Entropy and the Value of Information for Investors," *American Economic Review*, 103, 360–77.

Carnevale, A. P., B. Cheah, and J. Strohl (2012): "Hard Times, College Majors, Unemployment and Earnings: Not All College Degrees Are Created Equal." *Georgetown University Center on Education and the Workforce.*

Chevalier, A. (2012): "To Be or Not to Be... a Scientist?" working paper.

De Paola, M. and F. Gioia (2011): "Risk Aversion And Major Choice: Evidence From Italian Students," working paper, University of Calabria.

Dickson, L. (2010): "Race and gender differences in college major choice," *The Annals of the American Academy of Political and Social Science*, 627, 108–124.

Dolton, P. and A. Vignoles (2002): "Is a broader curriculum better?" *Economics of Education Review*, 21, 415–429.

Eeckhout, J. and X. Weng (2011): "Assortative Learning," working paper, Department of Economics, University of Pennsylvania.

Felli, L. and C. Harris (1996): "Learning, Wage Dynamics, and Firm-Specific Human Capital," *Journal of Political Economy*, 104, 838–68.

Finnie, R. (2001): "Graduates' earnings and the job skills-education match," *Education Quarterly Review: Statistics Canada*, 7, 7–21.

——— (2002): "Early Labour Market Outcomes of Recent Canadian University Graduates by Discipline: A Longitudinal, Cross-Cohort Analysis," Statistics Canada Analytical Studies Branch.

Heckman, J., L. Lochner, and P. Todd (2006): "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond," in *Handbook of the Economics of Education*, ed. by E. A. Hanushek and F. Welch, Elsevier.

Heijke, H. and C. Meng (2006): "The effects of higher education programme characteristics on allocation and performance of the graduates: A European view," working paper.

Holzer, H. J. and E. Dunlop (2013): "Just the Facts, Ma'am: Postsecondary Education and Labor Market Outcomes in the US," working paper.

Joensen, J. S. and H. S. Nielsen (2009): "Is there a Causal Effect of High School Math on Labor Market Outcomes?" *Journal of Human Resources*, 44, 171–198.

Kim, H.-K., S. C. Ahn, and J. Kim (2012): "Vertical and Horizontal Education-Job Mismatches in the Korean Youth Labor Market : A Quantile Regression Approach," working paper 1201, Research Institute for Market Economy, Sogang University.

Kinsler, J. and R. Pavan (forthcoming): "The Specificity of General Human Capital: Evidence from College Major Choice," *Journal of Labor Economics.*

Kirkeboen, L. J. (2012): "Preferences for lifetime earnings, earnings risk and nonpecuniary attributes in choice of higher education," working paper.

Levine, P. B. and D. J. Zimmerman (1995): "The benefit of additional high-school math and science classes for young men and women," *Journal of Business & Economic Statistics*, 13, 137–149.

Main, J. and B. Ost (2014): "The Impact of Letter Grades on Student Course Selection and Major Choice: Evidence from a Regression-Discontinuity Design," *The Journal of Economic Education*, 1, 1–10.

Malamud, O. (2010): "Breadth versus Depth: The Timing of Specialization in Higher Education," *Labour*, 24, 359–390.

——— (2011): "Discovering One's Talent: Learning from Academic Specialization," *Industrial & Labor Relations Review*, 64, 375–405.

McGuinness, S. (2006): "Overeducation in the Labour Market," *Journal of Economic Surveys*, 20, 387–418.

McGuinness, S. and P. J. Sloane (2011): "Labour market mismatch among UK graduates: An analysis using REFLEX data," *Economics of Education Review*, 30, 130–145.

Montmarquette, C., K. Cannings, and S. Mahseredjian (2002): "How do young people choose college majors?" *Economics of Education Review*, 2002, 543–556.

Morin, L.-P. (2013): "Estimating the benefit of high school for universitybound students: evidence of subjectspecific human capital accumulation," *Canadian Journal of Economics/Revue canadienne d'économique*, 46, 441–468.

Moscarini, G. and L. Smith (2001): "The Optimal Level of Experimentation," *Econometrica*, 69, pp. 1629–1644.

Ost, B. (2010): "The role of peers and grades in determining major persistence in the sciences," *Economics of Education Review*, 29, 923–934.

Papageorgiou, T. (2014): "Worker Sorting and Agglomeration Economies (Online Appendix)," working paper.

Reuben, E., M. Wiswall, and B. Zafar (2013): "Preferences and biases in educational choices and labor market expectations: shrinking the black box of gender," working paper.

Robst, J. (2007): "Education and job match: The relatedness of college major and work," *Economics of Education Review*, 26, 397–407.

Saniter, N. and T. Siedler (2014): "Door Opener or Waste of Time? The Effects of Student Internships on Labor Market Outcomes," working paper.

Silos, P. and E. Smith (forthcoming): "Human capital portfolios," *Review of Economic Dynamics*.

Stange, K. (2013): "Differential Pricing in Undergraduate Education: Effects on Degree Production by Field," working paper.

Stinebrickner, R. and T. Stinebrickner (2014): "A major in science? Initial beliefs and final outcomes for college major and dropout," *The Review of Economic Studies*, 81, 426–472.

Stinebrickner, T. R. and R. Stinebrickner (forthcoming): "Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model," *Journal of Labor Economics*.

Trachter, N. (forthcoming): "Stepping Stone and Option Value in a Model of Postsecondary Education," *Quantitative Economics*.

Turner, S. E. and W. G. Bowen (1999): "Choice of Major: The Changing (Unchanging) Gender Gap," *Industrial & Labor Relations Review*, 52, 289–313.

Walker, I. and Y. Zhu (2011): "Differences by degree: Evidence of the net financial rates of return to undergraduate study for England and Wales," *Economics of Education Review*, 30, 1177–1186.

Wine, J. S., M. B. Caminole, S. Wheeless, K. Dudley, J. Franklin, and K. Perry (2005): "1993/03 Baccalaureate and Beyond Longitudinal Study (B&B:93/03)," *National Center for Education Statistics*, 1–882.

Zafar, B. (2011): "How do college students form expectations?" *Journal of Labor Economics*, 29, 301–348.

Appendix A. Data

A.1. **Sample.** The B&B 93:03 sample is based on the 1993 National Postsecondary Student Aid Study,[48] restricted to students who were identified as baccalaureate recipients in the 1992-1993 school year. We use the restricted-access version of this dataset to generate the moments we match in our simulation. Since the simulation exercise was not done within the secure data environment, all data used in the estimation – as well as descriptive statistics reported elsewhere in the paper – have to meet disclosure restrictions. This requirement imposes two substantial limitations for our purposes: all sample sizes and frequency counts must be rounded to the nearest 10, and no values can be reported for cells with fewer than 3 individuals. While the estimation procedure uses percentage frequencies rather than count data, the second restriction results in some data loss: this is particularly the case for wage data, which we measure separately for those who switch fields and those who do not, within each major-timing of specialization cell.

While college graduates are likely to be more homogenous in ability than other groups (for instance, college entrants), they are nevertheless a disparate collection of individuals. We restrict the sample in a number of ways, both due to data quality and for conceptual reasons. The entire sample includes 10,980 individuals. We first restrict to individuals who are between 21 and 23 years of age at college graduation, reducing the sample to 7,090. Many of these students transferred institutions some years into their studies. Unfortunately, detailed course data is only available from the degree-granting university; in many cases, transferred courses are noted on the final transcripts, but without a date and often without a course-specific credit value. We retain as many transfer students as we can: essentially, those who transferred after one year or less, and whose transferred courses are identified.[49] This remains a costly restriction, reducing the sample to

[48]The NPSAS is a nationally representative sample of students (and institutions) at all levels of post-secondary education, at all types of institutions.

[49]We do not observe how long a student spent at a different institution. In practice, we allow students to have up to 45 transfer credits, if these transfer credits are or can be associated to a list of transferred courses. Students with more than 45 transfer credits are dropped. Students with between 20 and 45 transfer credits are retained: when applicable, these students are assumed to have 1 year of transferred courses. If a student has more than 20 transfer credits which are not associated to a list of transferred courses, that student is dropped. Students with 20 transfer credits or less are considered to not have taken an additional year, but to have earned these credits in other ways. These transfer credits (whether attributed or otherwise) are coded as part of their first year of studies.

5,750. Cleaning credit values and removing individuals with excessively high or low credit counts, as well as those with study gaps of a year or more, reduces the sample to 5,260. Finally, we restrict to those individuals who were followed up in 2003, leaving a sample of 4,170.[50]

A.1.1. *Majors.* While college may prepare students for work in many different ways, our model is constructed with applied majors in mind. The tradeoff between between skills and information only has bite when students anticipate that additional course material in their eventual field of work will bring additional returns. Many majors have weak links to the labor market; while this does not mean that they do not yield returns, optimal course choices in such majors may very well follow a different path from those in applied fields. We therefore restrict our sample to students with applied majors: these include six fields and 2160 individuals, subdivided into quantitative fields (1580) and non-quantitative fields (580).[51] While our estimation is done using only the quantitative graduates, we present statistics for all six majors here. The allocation across majors is given in Table 15.

Table 15. Count of observations by major

Major	Count	*Quantitative*	*Non-quantitative*
Science	630	X	
Engineering	350	X	
Business & Econ	600	X	
Education	430		X
Nursing	70		X
Social Wk & Protective	80		X
Total	2160	1580	580

Source: B&B93:03, sample restrictions described in section A.1. Counts rounded to the nearest 10 to respect disclosure restrictions.

Our analysis abstracts from the vertical dimension of ability. Computational constraints require us to be parsimonious with parameters, and we are specifically interested in horizontal abilities. Restricting our analysis to individuals who

[50]In our analysis we make use of the 2003 occupation observation only. This refers to the occupation held most recently by the respondent at the time of the survey, and therefore is non-missing even for individuals who are unemployed at the time of the survey. For most individuals, however, this is the occupation held ten years after college graduation.

[51]While this figure represents a dramatic reduction of the original sample, it is worth noting that this is largely due to the vast heterogeneity in the college graduate population. Using a sister dataset, albeit with an even more heterogenous population, Silos and Smith (forthcoming) are required to make similarly harsh restrictions.

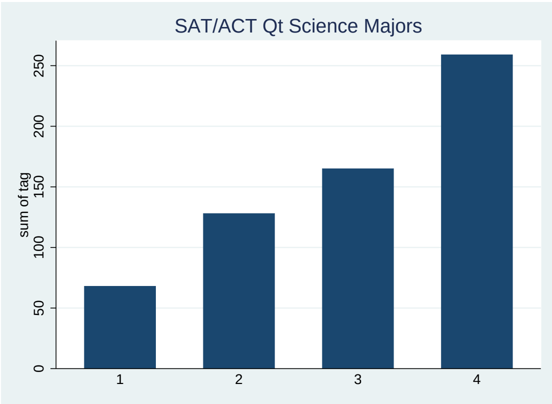
earned a bachelor degree between ages 21 and 23 narrows the distribution of ability within the sample. We remain concerned, however, that some skills may act as gatekeepers to certain fields, preventing lower-ability students from completing majors in those subjects even had they wished to.

Figure 11 shows the distribution of individuals across quartiles of SAT or ACT scores, by major. Clearly, some majors have a greater mass of high-ability students than others. This, combined with the small number of individuals choosing nursing and social work majors, motivates us to use only the three quantitative majors for our primary specification. While this restricts the interpretation of our results, it is plausible that this set of students is more homogenous that those enrolled in all six fields combined.

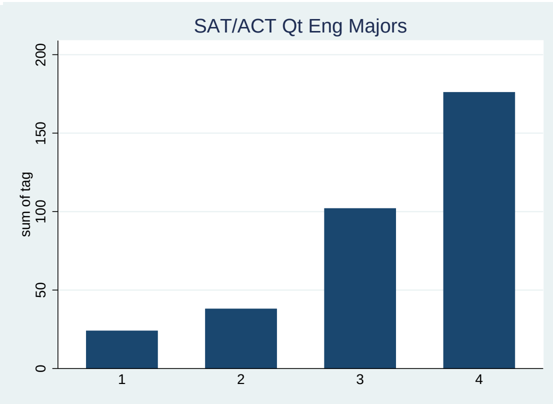
A.1.2. *Term length.* In principle, school terms are an intuitive and straight-forward concept, and relate naturally to the discrete-time version of our model. Classes are chosen at the start of the term and difficult to adjust once the term is underway; at the end of each term, enrollment for the next term – and the associated course selection – gets underway. In practice, however, the concept of a school term is difficult to pin down. In addition to the diversity of term structures (semesters, trimester, quarters, etc.), there are many students who enter university with some number of college credits. These may have been earned by exam (for instance, Advanced Placement courses), taken while in high school, or earned at a previous university and transferred to the degree-granting institution.[52] Even defining the start of the school year is not without difficulties: fall courses at one institution may start before summer courses have completed at another.

To mitigate these problems, we use the academic year as our period length; the remaining issues are dealt with in one of two ways. First, having restricted our sample to individuals who graduate between the ages of 21 and 23 – and therefore eliminating students who take an unusually long time to complete their degree – we abstract from term dates and divide a student's courses chronologically into four terms of equal credit value.

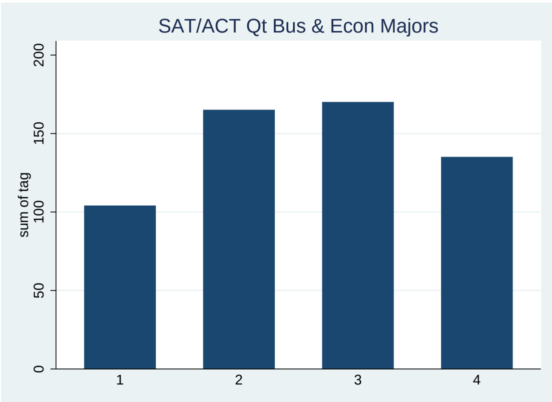
[52]In general, credits earned through exam are indicated as such on the transcript. Given that these are not actually classes, and are often earned based on prior education, they are not included in the analysis.



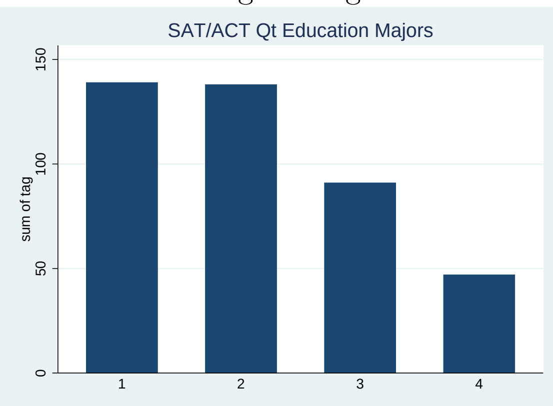
Science



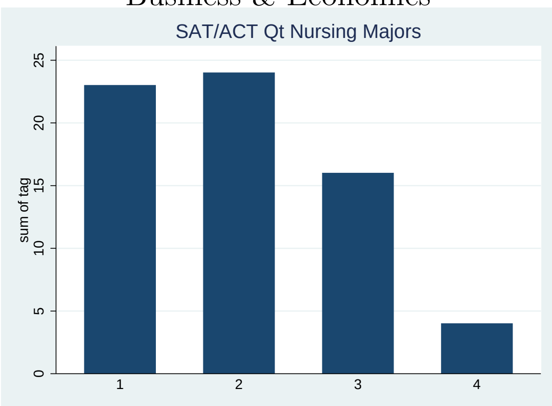
Engineering



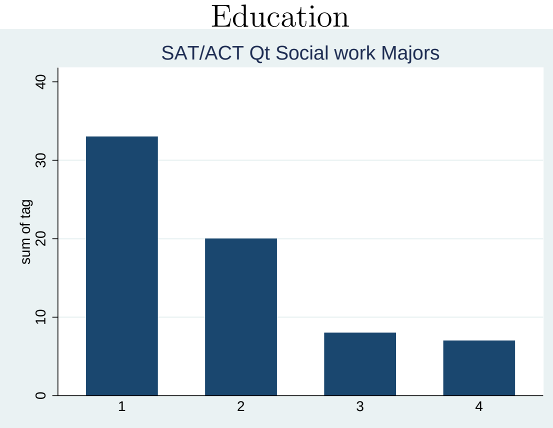
Business & Economics



Education



Nursing



Social work & protective services

Figure 11: SAT or ACT quartiles, by major

Our second approach accommodates diversity in time-to-degree by coding school years as faithfully as possible. We define the academic year as running from August to July, and attribute courses accordingly. To avoid creating spurious years of college (due to a late summer course, for instance), we recode any terms with 6 or fewer credits as belonging to the next or previous school term. Finally, to maintain a reasonably homogenous group, we restrict the sample to students who graduate in 4 or 5 years. Table 16 gives the distribution of time-to-degree for these students.

TABLE 16. Time to degree, by major

Major	4 years	5 years
Science	380	210
Engineering	140	170
Business & Econ	350	210
Education	200	180
Nursing	40	30
Social Wk & Protective	40	30
Total	1150	830

Source: B&B93:03, sample restrictions described in section A.1. The timing of specialization used is the primary specification, with time-to-degree computed using a true-term approach.

Table 17 presents correlations between two different timing variables computed using both true years and standardized years. The two approaches are strongly correlated. While the second approach would permit us to examine how the timing of specialization is related to time-to-degree, data sparsity becomes a pressing concern (note that this would require us to track of 4- and 5-year degree students separately). In addition, it is not clear how adequately our model captures the choice to pursue a 5th year of undergraduate study. For these reasons we use the first approach for our primary specification, standardizing the duration of a college degree to four years.

TABLE 17. Correlation between timing of specialization using true and standardized years

Timing variable	TY-con	TY-ret	4Y-con	4Y-ret
True years - concentration	1.0000			
True years - 90% retention	0.8277*	1.0000		
Four years - concentration	0.8288*	0.8190*	1.0000	
Four years - 90% retention	0.7512*	0.8298*	0.9009*	1.0000

Source: B&B93:03, sample restrictions described in section A.1. Star indicates significance at the 10% level. Thresholds are defined and explained in Section A.3.1.

A.2. **Matching occupations to majors.** One of the key outcomes we are interested in is whether individuals pursue a career in their field of studies, or whether they switch into a different field. This link is better defined for some fields than for others: some majors, such as engineering or education, have obvious careers associated to them. Other majors, including most of the humanities and social sciences, do not lead unambiguously to a certain occupation. Table 18 gives an overview of the occupations held by sample members, as a percentage of all graduates from each major.

TABLE 18. Share of major in each occupation

Occupation	Sci	Eng	Bus	Edu	Nur	Swp
Educators	14	3	5	74	0	14
Business/management	14	24	57	7	0	14
Engineering/Architecture	5	44	2	0	0	0
Computer science	11	9	5	0	0	0
Medical professions	27	0	2	2	100	14
Editors/writers/performers	2	0	2	2	0	0
Human/protective services/legal pro	3	3	7	2	0	43
Research/scientists/technical	14	9	2	2	0	0
Administrative/clerical/legal	2	0	3	2	0	14
Mechanics/laborers	2	3	2	2	0	0
Service industries	5	3	13	5	0	0
Other/military	2	3	2	0	0	0

Source: B&B93:03, sample restrictions described in section A.1. Note that columns may not sum to 100 due to rounding.

Matching occupations with majors is facilitated by the restriction to applied fields described in Appendix A.1. We base our major categories on 14 aggregated majors given in the data,[53] and derive a correspondence between these majors and 12 occupation categories. The aggregation of majors is given in Table 19, along with the corresponding occupations. Note that we select the sample based on major, making no restrictions with regard to occupation.

A.3. **Timing of specialization.**

A.3.1. *Alternate specifications.* In addition to the primary specification described in Section 2.2, we derive a number of alternate timing of specialization variables. These alternate approaches allow us to run robustness checks and explore the generality of our primary measure.

Figures 12 - 14 present the distribution of timing of specialization for four alternate specifications. Five different thresholds are represented for each specification,

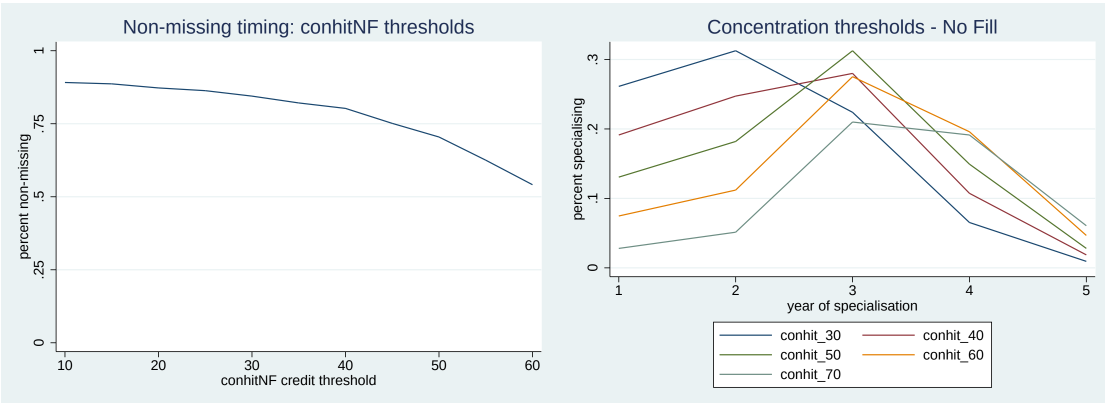
[53]We deviate from these categories by recoding economics with business, rather than with social sciences.

Table 19. Occupation-Major Correspondance

14 MAJORS	6 MATCH CATEGORIES	12 OCCUPATIONS
Biological/interdisc sciences Mathematics/physical sciences Computer science	Sciences	Research, Science & Technical Medical Professionals Computer science
Engineering/architecture	Engineering & Architecture	Engineering/architecture
Business Soc sciences (econ only)	Business & Management	Business and management
Education	Education	Educators
Health/nursing	Nursing	Medical professionals
Social work/protective serv	Human/Protective Services & Legal	Human/protective service/legal prof
	Unmatched occupations	Administrative/clerical/legal sup Mechanics, laborers Service industries Other, military Editors/writers/performers
Humanities/liberal studies Agriculture Social sciences (non-econ) Other Communications/journalism Health/other	*Excluded majors*	

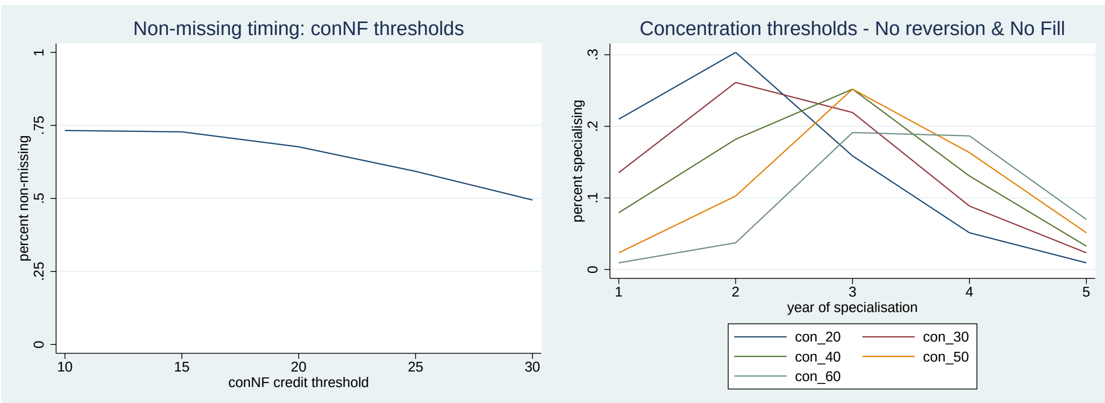
Source: authors' aggregation, based on B&B93:03 major and occupation codes. Note: 'other' majors are not considered matched with 'other' occupations.

as well as the attrition in the number of specializers as the threshold rises. Figures 12 and 13 are computed using true years (see section A.1.2). Both thresholds refer to a concentration of credits the student must reach or exceed; however, the thresholds in Figure 13 require that students remain above that threshold until graduation, while those in Figures 12 do not. Figure 14 presents the distribution of the timing of specialization for equivalent thresholds, using standardized 4-year college tenures. In this figure 'never'-specializers appear on the right as if they had specialized in Year 5 (hence their representation with one graph rather than two).



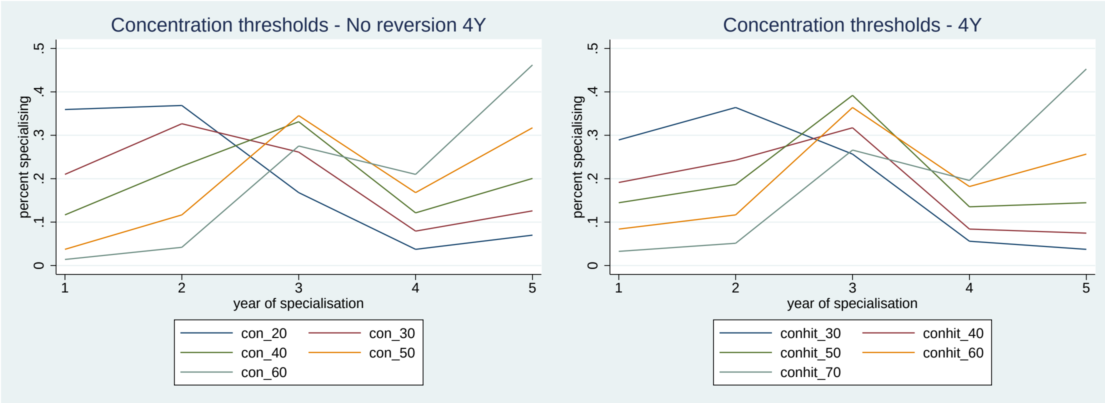
Count of specializers as threshold rises   Timing of Specialization

Figure 12: Concentration-hit thresholds, allowing reversion: true years



Count of specializers as threshold rises   Timing of Specialization

Figure 13: Concentration-hit thresholds, no with reversion: true years



Concentration-hit thresholds, no with reversion

Concentration-hit thresholds, allowing reversion

Figure 14: Timing of specialization using standardized years (note: specialization in Year 5 is equivalent to never specializing)

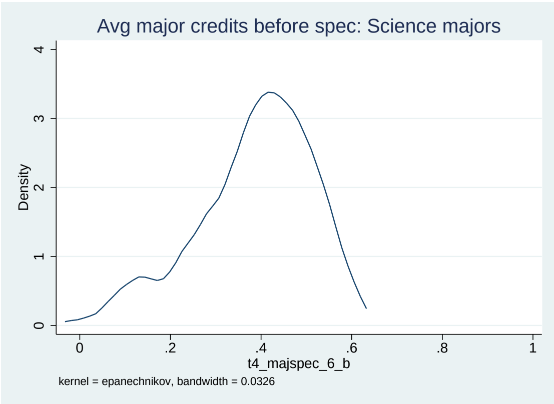
A.3.2. *Specialization and course choices.* We have identified specialization based on the concentration of courses a student takes during each school year. While we have no way to verify our specialization concept externally, we can make some basic checks.

First of all, our approach relies on students taking more major-field courses later in their degrees. Figures 15 and 16 show the distribution of the percent of credits taken in the major field, before and after specialization. While the specification itself could induce a modest difference in these distributions, it is encouraging to see that students are indeed taking few credits in their major prior to specialization.[54]

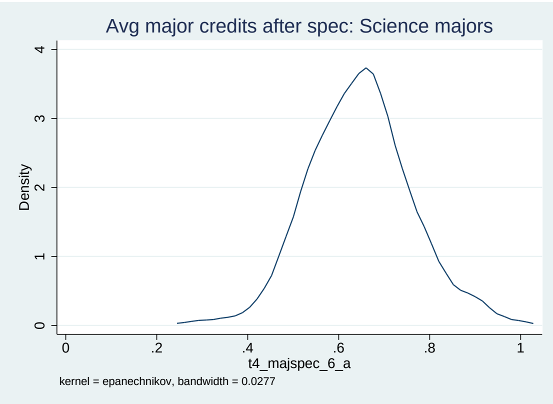
Next, our model supposes that students take a constant share of courses in their major field in each period of specialized studies, regardless of the timing of specialization. This does not permit late specializers to load up on major-specific courses in order to meet major requirements or catch up with their early-specializing peers. To check whether this assumption is reasonable we look at how total credits and credits in the major field vary with the timing of specialization. Table 20 gives the mean and standard deviation of total credits, for each major and timing of specialization. While there is some variation across majors, the average credit load is encouragingly flat with regards to the timing of specialization.[55]

[54]Science remains an outlier, with the mean credit share before specialization being quite large at 40%, while the mean credit share afterwards is a more typical 60%.

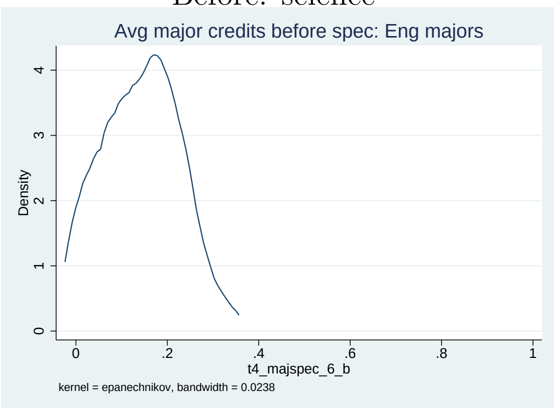
[55]The invariance of total credits to the timing of specialization suggests that the timing of specialization may not be that closely correlated with time-to-degree: late specializers do not



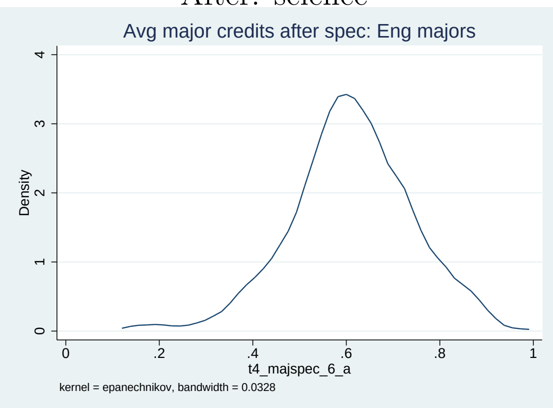
Before: science



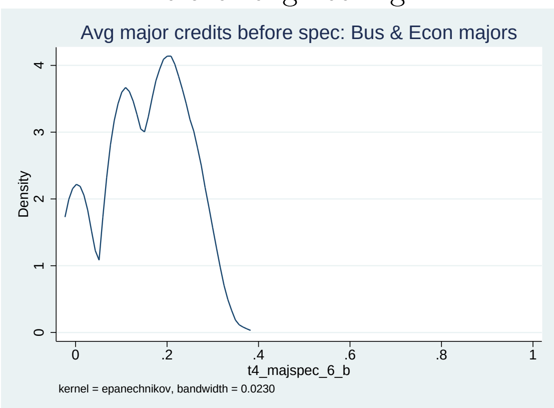
After: science



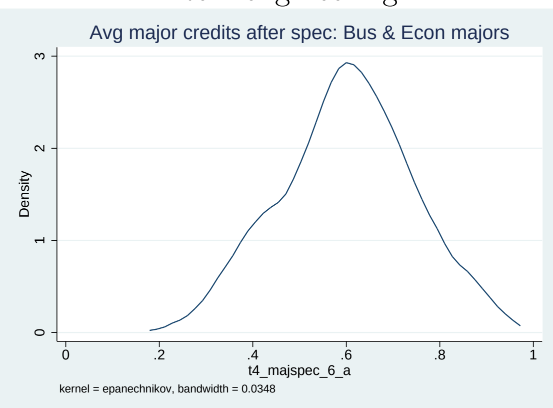
Before: engineering



After: engineering



Before: business & econ

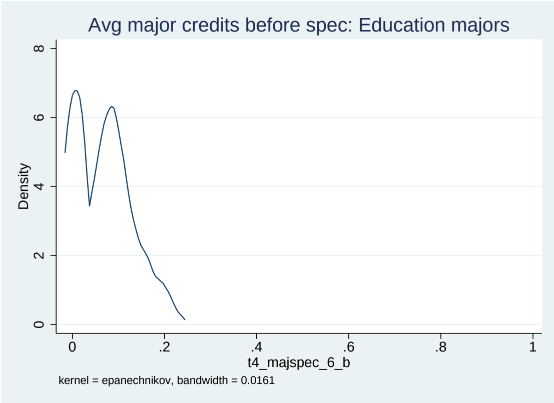


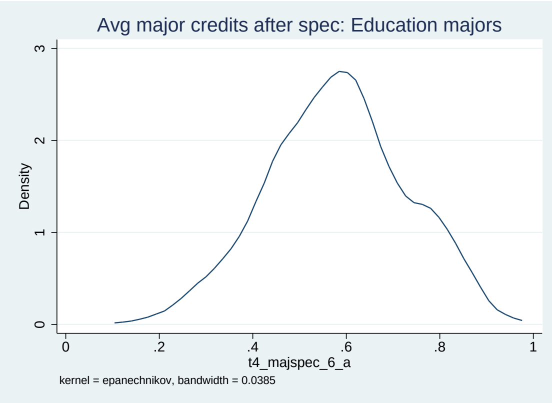
After: business & econ

Figure 15: In-major credit share before and after specialization: quantitative majors (primary specification)

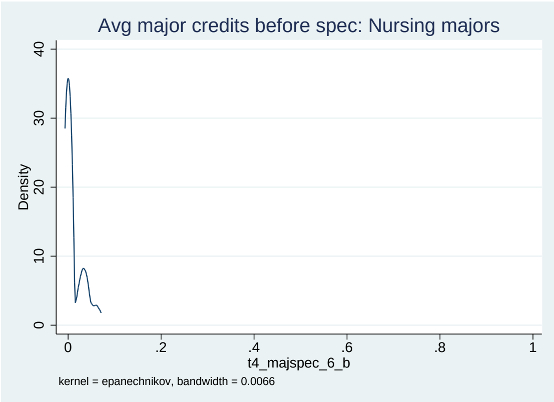
The mean number of in-major credits for each major and timing of specialization are shown in Table 21. Unlike for total credits, the number of in-major credits declines with later specialization, although clearly this is more true in some majors than in others. Late specializers may indeed try to 'catch up' by taking a heavier course load; however, early specializers still take more credits in their major field.

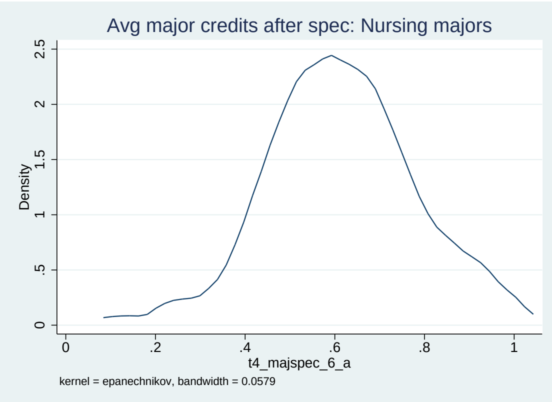
systematically accumulate an extra year of courses. The abstraction we make from time-to-degree is less striking in light of this.



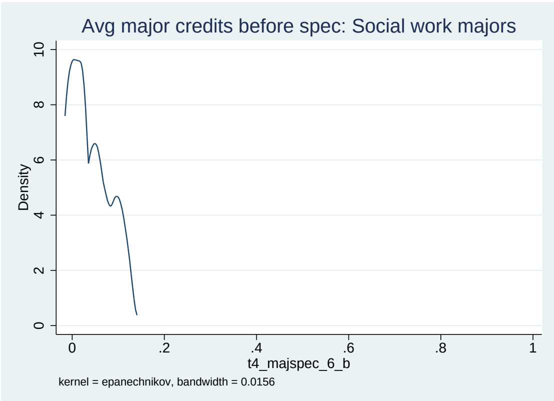


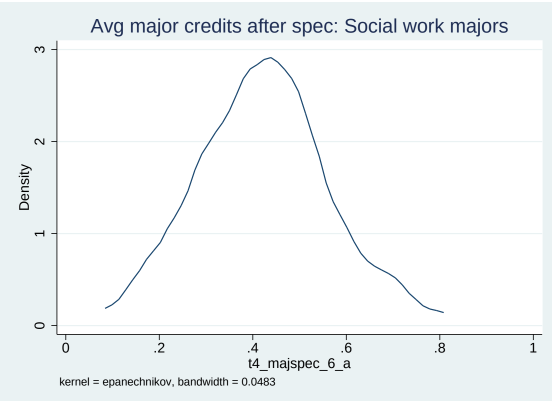
Before: education After: education





Before: nursing After: nursing





Before: social work & prot. After: social work & prot.

Figure 16: In-major credit share before and after specialization: non-quantitative majors (primary specification)

Finally, Table 22 shows the difference in the term-by-term share of courses taken in the major field, before and after specialization. From the table, it does not appear that late specializers experience a higher jump in course taking than early specializers. This suggests that any catch-up by later specializers is modest.

A.4. **Correlations on observables.** Table 23 presents the correlations of our main timing of specialization variable with several observable characteristics of

Table 20. Total credits, by major and timing of specialization

Major	Year 1	Year 2	Year 3	Year 4	Never
Science	126 (10)	127 (12)	127 (14)	129 (15)	127 (13)
Engineering	134 (12)	135 (11)	136 (12)	132 (17)	132 (8)
Business & Econ	125 (14)	126 (11)	126 (10)	126 (15)	125 (12)
Education	133 (13)	133 (12)	134 (13)	136 (15)	131 (15)
Nursing	125 (7)	128 (15)	138 (14)	-	-
Social Wk & Protective	129 (11)	122 (10)	128 (11)	126 (12)	-

Source: B&B93:03, sample restrictions described in section A.1. Standard deviations in parenthesis. A standard undergraduate degree is 120 credits.

Table 21. In-major credits, by major and timing of specialization

Major	Year 1	Year 2	Year 3	Year 4	Never
Science	77 (13)	74 (12)	70 (13)	65 (9)	39 (18)
Engineering	70 (13)	63 (12)	57 (11)	38 (7)	10 (9)
Business & Econ	62 (13)	60 (12)	52 (12)	38 (7)	25 (11)
Education	67 (17)	60 (14)	47 (13)	28 (7)	8 (6)
Nursing	59 (8)	55 (7)	49 (12)	-	-
Social Wk & Protective	41 (13)	37 (7)	32 (9)	25 (8)	-

Source: B&B93:03, sample restrictions described in section A.1. Standard deviations in parenthesis. A standard undergraduate degree is 120 credits.

Table 22. Major-field course share: *after* minus *before* specialization

Major	Year 2	Year 3	Year 4
Science	0.21	0.25	0.33
Engineering	0.46	0.5	0.33
Business & Econ	0.48	0.45	0.34
Education	0.51	0.54	0.47
Nursing	0.56	0.68	-
Social Wk & Protective	0.36	0.41	0.48

Source: B&B93:03, sample restrictions described in section A.1. Standard deviations in parenthesis. A standard undergraduate degree is 120 credits.

the students: family income quartile, father's and mother's education, academic ability prior to college (captured by SAT or ACT score quartiles, and also SAT math and verbal scores separately for those students who took the SAT), and gender. The correlations are small and in general not significant at the 10% level. Family income is the exception: higher incomes are associated with a later timing of specialization.

The absence of correlations may partly be an artifact of aggregation. Table 24 breaks the sample into students graduating with quantitative majors (science, engineering or business and economics), and those graduating with non-quantitative

TABLE 23. Correlations on Observables - all majors

	TIM	FaInc	FEdu	MEdu	S/Aq	SATV	SATM	G
TIMING of SPEC	1.000							
Family income	0.049*	1.000						
Father's ed	0.024	0.363*	1.000					
Mother's ed	0.046	0.313*	0.551*	1.000				
SAT/ACT quart	0.024	0.171*	0.224*	0.205*	1.000			
SAT verbal	0.055	-0.047	0.023	0.018	-0.013	1.000		
SAT math	0.035	-0.063*	-0.018	-0.019	0.002	0.672*	1.000	
Gender	-0.020	-0.065*	-0.070*	-0.038*	-0.178*	-0.007	-0.037	1.000

Source: B&B93:03, sample restrictions described in section A.1. Timing of specialization is the primary specification (see Section 2.2). Gender is increasing in femininity. Star indicates significance at the 10% level. Correlations are pairwise, starting from a maximum sample of 2160 observations.

majors (education, nursing or social work and protective services). While the correlations remain modest in size, some stronger patterns emerge. Quantitative graduates are more likely to specialize early if they have higher SAT or ACT scores, while the reverse is true for non-quantitative graduates. On the other hand, non-quantitative graduates are more likely to specialize early if they are women, while the reverse is true in the quantitative fields. The correlation between timing of specialization and family income disappears when considering quantitative majors alone, but strengthens slightly for non-quantitative majors. These correlations motivate our choice of split-sample robustness checks, presented in Section 6. In particular, we consider separately men and women, and upper and lower ability students. Given that the correlation with family income is not present in our primary sample, and is otherwise relatively small, we do not pursue it at this time.

Appendix B. Analytical results in the linear-symmetric model

B.1. **Linear-symmetric model.** To explore the impact that the parameters of the model have on the empirical outcomes we are interested in, we consider a simplified parametric version of the model presented in Section 3.1. We assume perfect symmetry between the fields ($R_s = R_a$) and a neutral prior, $p_0 = 1/2$. Furthermore, let the return function R be linear in the education stock and composed of a baseline wage level and a term proportional to the effective education

Table 24. Correlations on Observables - by type of major

	TIM	FaInc	FEdu	MEdu	S/Aq	SATV	SATM	G
Quantitative								
TIMING of SPEC	1.000							
Family income	-0.013	1.000						
Father's ed	0.015	0.340*	1.000					
Mother's ed	0.008	0.251*	0.546*	1.000				
SAT/ACT quart	-0.104*	0.131*	0.197*	0.192*	1.000			
SAT verbal	0.045	-0.108*	-0.041	-0.006	-0.049	1.000		
SAT math	0.017	-0.055	-0.100	-0.037	0.007	0.653*	1.000	
Gender	0.091*	-0.060*	-0.044*	-0.040	-0.153*	0.077	0.015	1.000
Non-quantitative								
TIMING of SPEC	1.000							
Family income	0.071*	1.000						
Father's ed	0.038	0.301*	1.000					
Mother's ed	0.033	0.245*	0.590*	1.000				
SAT/ACT quart	0.101*	0.110*	0.166*	0.155*	1.000			
SAT verbal	0.107	0.041	0.094	-0.134	0.048	1.000		
SAT math	0.088	-0.001	0.035	-0.080	-0.063	0.657*	1.000	
Gender	-0.149*	0.009	0.037	0.061	-0.007	-0.141	-0.005	1.000

Source: B&B93:03, sample restrictions described in section A.1. Timing of specialization is the primary specification (see Section 2.2). Gender is increasing in femininity. Star indicates significance at the 10% level. Correlations are pairwise, starting from maximum samples of 1580 (quantitative) and 580 (non-quantitative) observations.

stock

$$R(\epsilon_f) = w_0 + y\epsilon_f. \tag{29}$$

Under these assumptions, we can obtain simple and interpretable closed-form solutions for the optimal length of specialized schooling. Assuming that the agent with education stocks (e_S, e_A) anticipates to remain in field s (even if he learns that his comparative advantage is in field a), he will choose the value of H^* in order to achieve an optimal aggregate education stock ϵ_S. The optimal value H^* satisfies:

$$e_S + H^* + \beta e_A = \frac{1}{r} - \frac{w_0 + pP - z}{y_s}, \tag{30}$$

where z is the flow payoff from studies. The agent's optimal tenure in specialized education is $H = H^*$ if positive, and zero otherwise. A necessary condition for H^* to be positive[56] is that inequality (31) hold, which happens when the labor market

[56]Inequality (31) obtains from (30) by imposing $e_S = e_A = 0$ and $p = 0$ and imposing that the right-hand side be positive.

rewards training, flow payoffs of specialized education are large and the agent is patient.

$$P + 2w_0 < \frac{2y}{r} + 2z \tag{31}$$

We assume that condition (31) holds strictly. Under these conditions, the value of starting specialized studies in subject S at time t is given by:

$$V_{S,st}(p, t/2, t/2) = r^{-2}\exp\left\{ r\frac{(\beta+1)}{2}t + r\frac{pP}{y} + \frac{w_0 r - rz}{y_s} - 1 \right\} y + r^{-1}z, \tag{32}$$

where subscript st indicates the intention to stay in field s (as opposed to sw for switching). A similar formula can be obtained if the agent anticipates changing fields. Again assuming an interior choice of H^*, we have:

$$V_{S,sw}\left(p, t/2, t/2\right) =$$
$$r^{-1}z + r^{-2}y\left(p\left(1-\beta\right)+\beta\right)\exp\left\{ \frac{r\left(\beta+1\right)t}{2\left(p\left(1-\beta\right)+\beta\right)} + \frac{r\left(w_0+P-z\right)}{y\left(p\left(1-\beta\right)+\beta\right)} - 1 \right\}. \tag{33}$$

Using these formulas, we can show that some experimentation is worthwhile as long as the signal is informative, thereby guaranteeing $\tau_1 > 0$.

Proposition 1 (Minimum level of experimentation)**.** *In the linear-symmetric case, $p_s(0)$ is bounded away from $\frac{1}{2}$: the agent engages in mixed studies in the neighborhood of $t = 0$.*

We can further establish that mixed education becomes dominated in finite time: there exists a limit at which agents stop experimenting regardless of the path followed by the belief process.

Proposition 2 (Maximum level of experimentation)**.** *In the linear-symmetric case, there exists $t^* > 0$ such that $p_s(t^*) = p_a(t^*) = 1/2$: the agent never engages in mixed studies for $t \geq t^*$.*

These two propositions show that the boundary p_s moves from a position bounded away from $1/2$ at $t = 0$ down to $1/2$ before a fixed time t^*. While this is not inconsistent with the boundary being locally increasing, it must be decreasing on average. In numerical simulations, the optimal boundary is monotonically decreasing.

Appendix C. Simulations results in the linear-symmetric model

We illustrate the properties of the optimal solution by identifying the optimal boundary numerically for a given set of parameters in a linear-symmetric model with no switching cost. We show how parameter changes affect that optimal boundary, the flow of agents through the different education regimes as well as the level of confidence at which their initial specialization takes place. This justifies the parameter selection for our estimation exercise.

C.0.1. *Optimal boundaries and exit.* Figure 17 (reproduced from Section 3.1) displays the optimal belief boundaries, overlaid with a sample belief path of a type S agent. Near $t = 0$, agents require strong beliefs in order to specialize: above 0.75 (to specialize in S) or below 0.25 (to specialize in A). If they have not specialized by time $t = 8$, a very small deviation of p_t away from 1/2 is enough to trigger specialization. The density of specialization times is also displayed in the figure, for specialization into S or A separately.[57] By integrating these densities, we find that approximately 68 percent of type S agents eventually specialize correctly in field S – the field of their comparative advantage, while 32 percent specialize in field A and are therefore initially mismatched.

Upon specialization, agents already know the probability with which they will switch fields. The duration of specialized studies is then determined in part by whether the agent will choose to pursue his comparative advantage, should he discover on the labor market that his comparative advantage is in the other field. Figure 18 shows the probability of switching fields as a function of the timing of specialization (solid red line), and shows that the earliest specializers do not switch fields. The Figure also displays the length of specialized studies for these same individuals (dashed blue line), and shows that early specializers do not “hedge their bets”: they invest heavily in specialized education and plan to stay in their initial field even if they turn out to be initially type-mismatched. In the opposite extreme, very late specializers - who have chosen their field on relatively poor information - know that they are likely to discover their comparative advantage is actually in the other field and accordingly spend comparatively little time specializing before

[57]We also assume that the initial distribution of priors is a Dirac mass at 1/2.

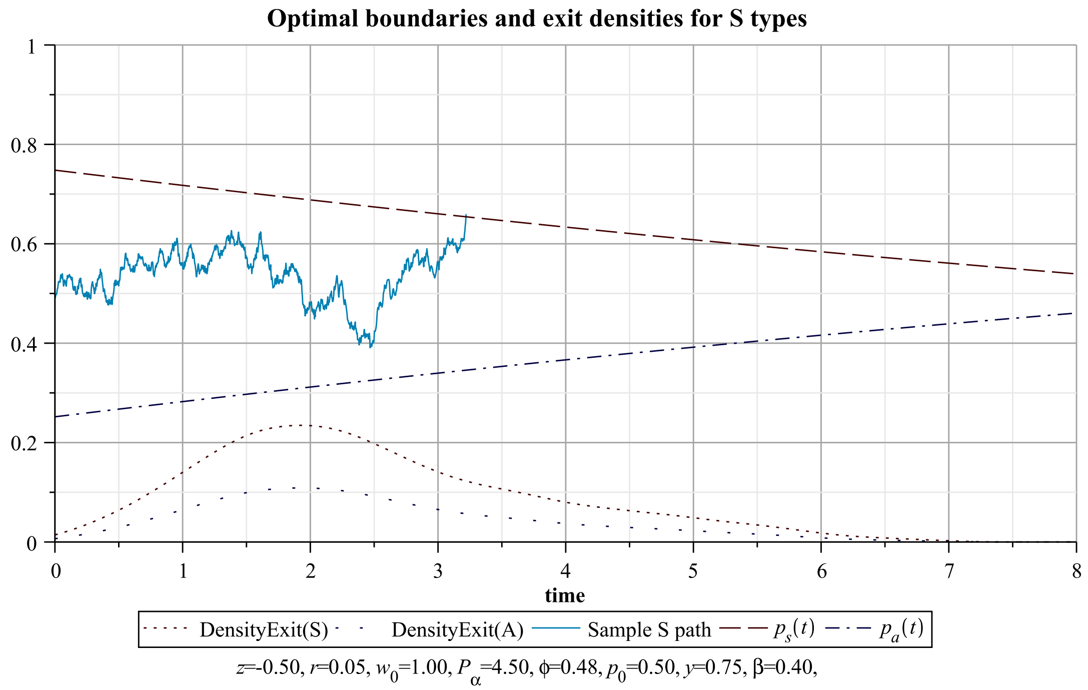


Figure 17: Optimal Boundaries and density of specialization times. *The agent starts specialized education once the belief process* p_t *escapes the interval* $(p_a(t), p_s(t))$. This particular sample path leads to specialization time $\tau_1 \approx 1.64$ and corresponds to correct specialization (in subject S).

joining the labor market (keep in mind that late specializers have accumulated a significant amount of human capital in *both* fields).

C.0.2. *Parameter changes.* We now consider how changes in individual parameter values affect the behavior of agents, *ceteris paribus*. Payoff-relevant parameters (P, w_0, y, z, β, r) do not impact the informativeness of signals; however, they influence the relative value of specialized education and type-discovery. Figure 19 illustrates how the belief boundaries change following a discrete change in the value of each of these parameters (the figures show only the upper boundary $p_s(t)$; the lower boundary $p_a(t)$ will adjust symmetrically). Since the speed at which agents reach the boundaries is not affected by these parameters, the location of the boundary itself summarizes the impact of these changes.

While the distributions in Figure 17 are generated under the assumption that all belief paths begin at $p_0 = 1/2$ (and evolve according to equation (2)), the model does allow for agents to have some information about their type at $t = 0$. This is formally represented by a distribution of time-zero beliefs p_0 that is correlated with

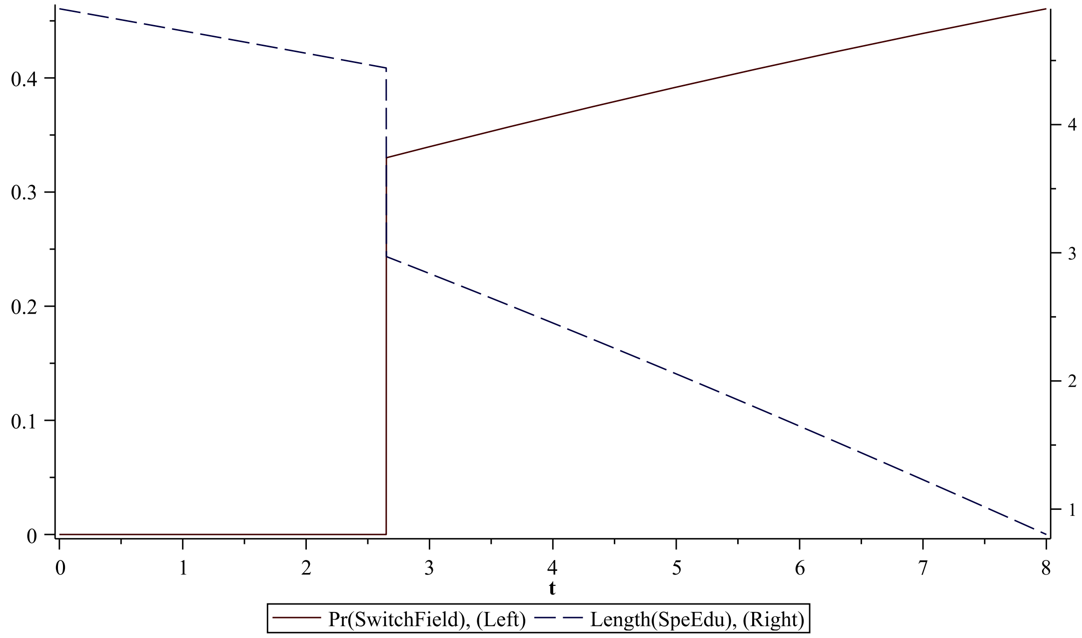
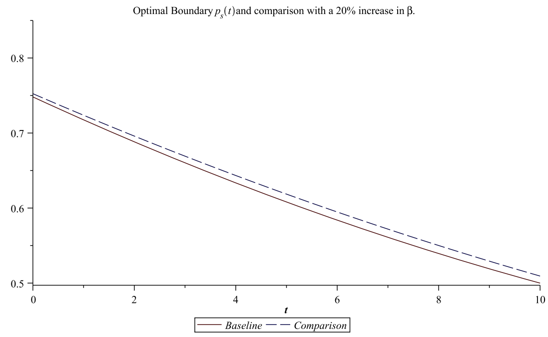


Figure 18: Probability of switching fields and length of specialized education tenure conditional on specialization time. *Early specializers ($t \leq 2.7$) remain in their field of specialization even if revealed to be of type A. They choose a longer education tenure.*

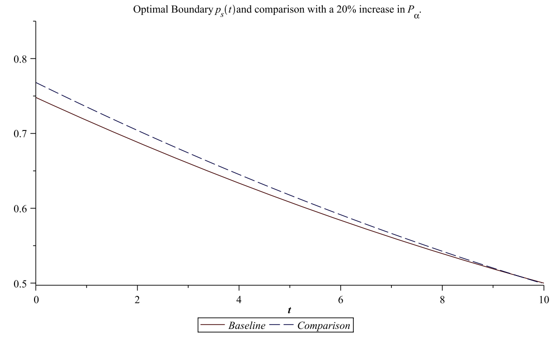
the true type. In contrast with the return parameters, the initial information set does not affect the optimal forward-looking experimentation policy, which implies that its effect is entirely driven by the density of exit times.

Informative initial beliefs are illustrated in Figure 20. Let the belief at $t = 0$ be informative in the following sense: before $t = 0$, agents receive a symmetric binary signal that agrees with their true type with probability 3/5 and is misleading otherwise.[58] Figure 20 plots the density of timings of specialization – separately for each field – for two populations of S-type agents which differ only in their prior information. The solid red line and dashed blue line give the specialization times for agents with an informative prior, while the dotted red line and dot-dashed blue line give the distribution for agents with no prior information about their type. As before, the solid and dotted lines correspond to those agents who specialize – correctly – into field S, while the dashed lines represent agents who mistakenly believe their comparative advantage to be in field A. We can see that

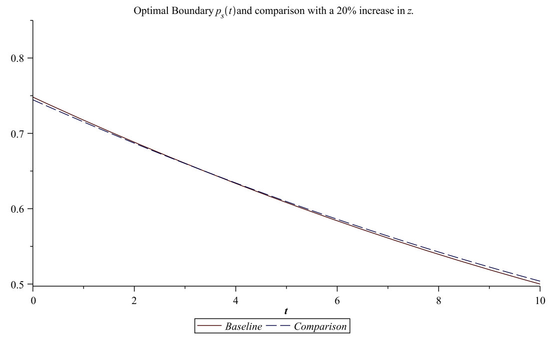
[58]Such a signal leads to type S agents holding updated belief $p_0 = 3/5$ with probability 3/5 and belief $p_0 = 2/5$ with probability 2/5, with symmetric numbers for type-A agents.



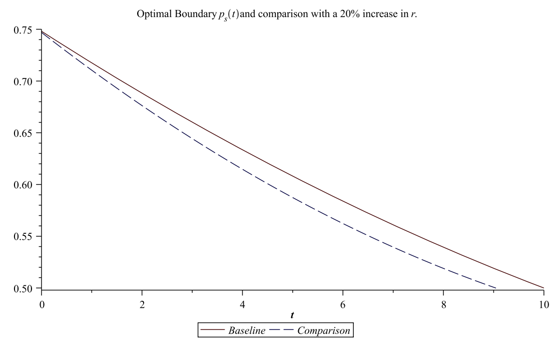
Transferability of human capita: β



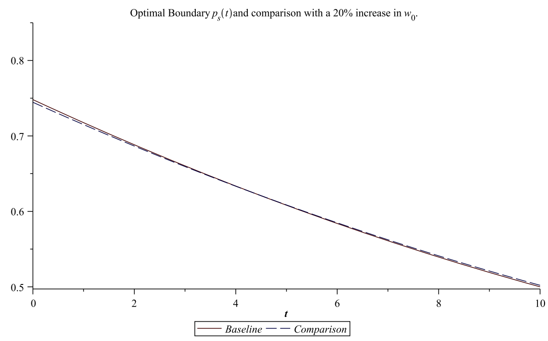
Comparative advantage premium: P



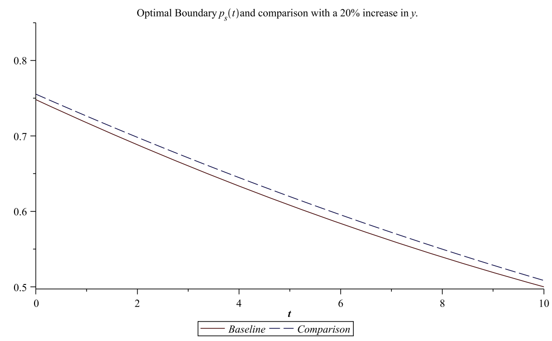
Flow payoffs during studies: z



Discount rate: r



Returns function intercept: w_0



Marginal returns to human capital: y

Figure 19: **Comparative statics**. Each sub-figure compares the optimal upper boundary $p_s(t)$ before (solid red line) and after (dashed blue line) a 20% increase of the parameter in question. For a fixed learning technology, an upwards shift of the boundary implies that agents specialize later: they hold out for more confidence before committing to either field.

improving the initial information of agents has the effect of speeding up the process in the sense of first-order stochastic dominance and of increasing the probability of correct initial match.

The precision of learning is the only parameter entering equation (2), hence the beliefs updating process. As a result, it impacts the optimal specialization decision through the choice of boundaries $p_s(t), p_a(t)$, but also the distribution of times at which boundaries are reached and the likelihood of specializing correctly. Figure

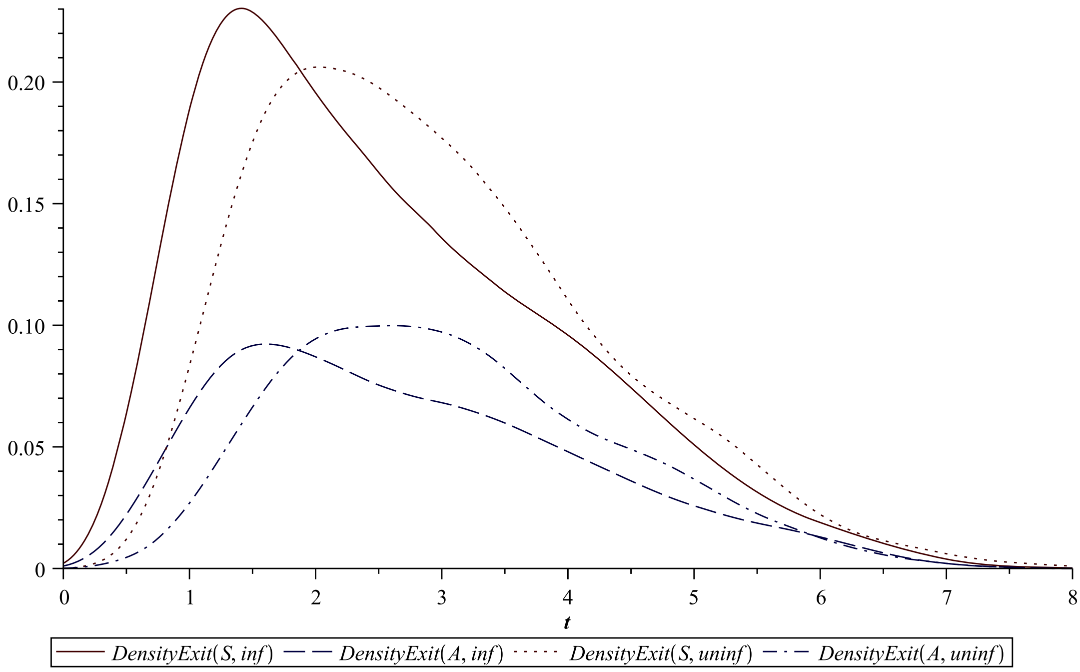
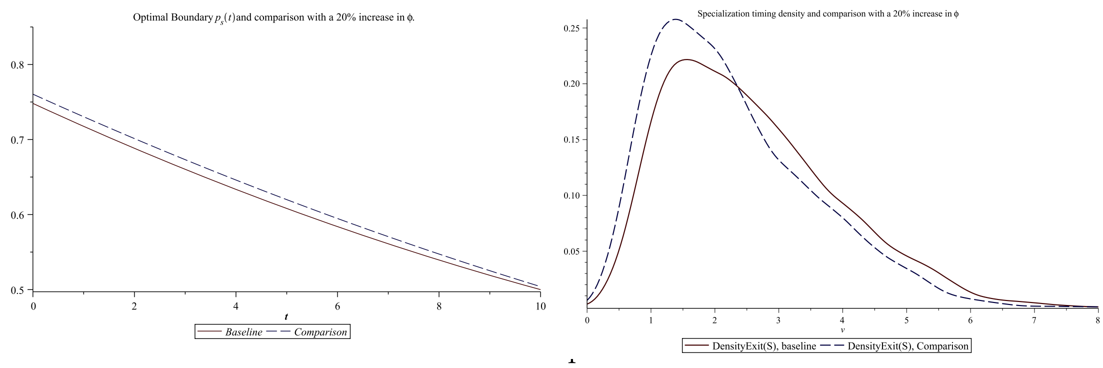


Figure 20: Specialization timing densities with informative and uninformative initial beliefs.

21 shows how the boundary and the exit density change when the precision of the signal increases by 20%. On one hand, the increased demand for experimentation (left panel) implies that agents should specialize later. On the other hand, a higher signal-to-noise ratio implies that agents reach any given target belief faster, which is the dominant effect and explains why the density of exit times puts more weight on early realizations. Notice that the two effects go in the same direction with respect to the probability of correct eventual specialization: not only the optimal boundary shifts upwards, but also agents exit faster, hence are more likely to specialize upon reaching a high confidence threshold.

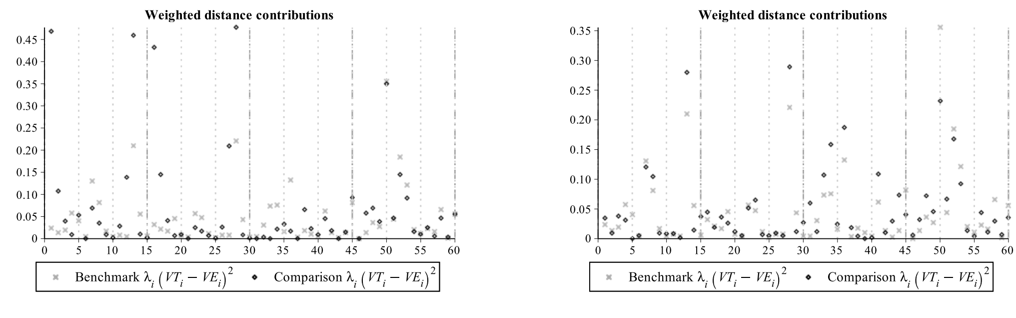


boundary $p_s(t)$ Specialization timing density (S exit)

Figure 21: Change in the precision of learning, ϕ

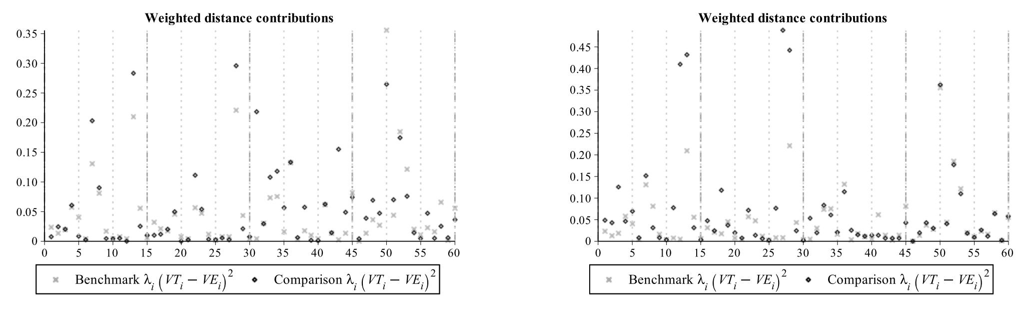
Appendix D. Identification

Figures 22 - 26 show the weighted contribution of each moment to the distance criterion, for each of the experimental conditions discussed in Section 4.6.



Low ρ High ρ

Figure 22: Contribution to distance criteria: baseline (grey) and constraint (black)



Low ρ_0 High ρ_0

Figure 23: Contribution to distance criteria: baseline (grey) and constraint (black)

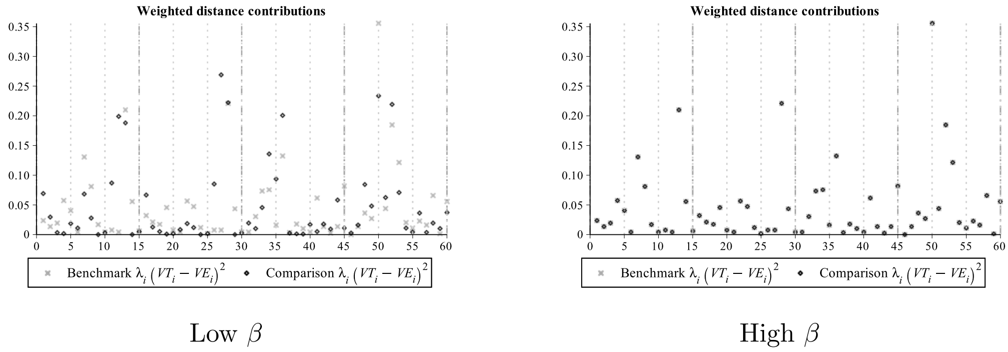


Figure 24: Contribution to distance criteria: baseline (grey) and constraint (black)

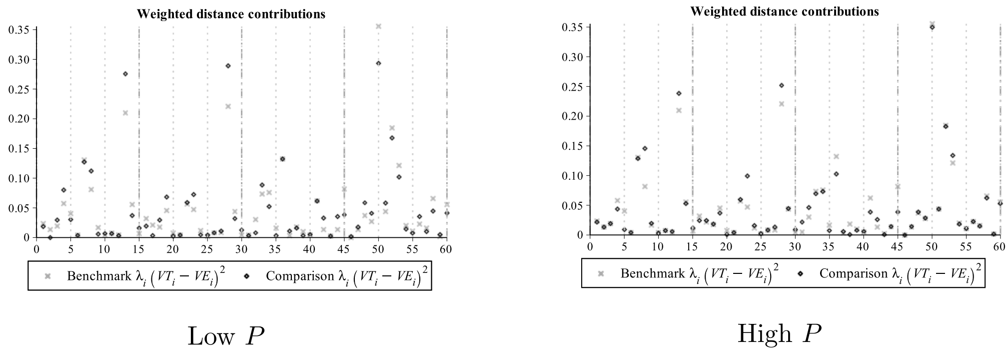


Figure 25: Contribution to distance criteria: baseline (grey) and constraint (black)

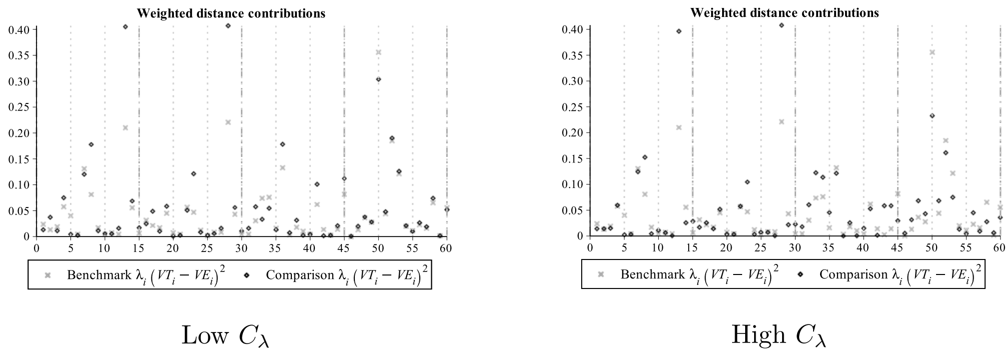


Figure 26: Contribution to distance criteria: baseline (grey) and constraint (black)